Applying Corpus Pattern Analysis to Learner Corpora:

Investigating the Pedagogical Potential of the Pattern Dictionary

of English Verbs

By

Abdullah Saad Alqarni

A Thesis Submitted to the University of Birmingham in Partial Fulfilment of the

Requirements for the Degree of Doctor of Philosophy

School of English, Drama

and American & Canadian Studies

College of Arts and Law

The University of Birmingham

December 2018

Abstract

This thesis contributes to research in learner corpora by offering a novel approach to their analysis. Specifically, it uses the concepts in the Pattern Dictionary of English Verbs (PDEV), that is, corpus pattern analysis (CPA), to describe selected learner corpora. The thesis analyses and compares the use of 16 verbs in two sections of the International Corpus of Learner English (ICLE; ICLE-Swedish and ICLE-Chinese) and in the native-speaker Louvain Corpus of Native English Essays (LOCNESS) using the descriptions of those verbs presented in PDEV.

The thesis demonstrates that the concepts in PDEV can be used successfully to describe the association between the patterns and meanings of verbs in both learner and native-speaker corpora to a degree that is not possible with other models. It argues that teachers can use PDEV to identify divergent patterns of verbs produced by their learners, and thus employ it as a tool for raising learners' awareness, in the classroom, of conventional and divergent verb patterns. In addition, verb patterns that are under- or over-represented in the learner corpora when compared with the native-speaker corpus are identified and reasons for these phenomena are offered. Overall, the thesis concludes that the approach to verb patterning articulated in corpus pattern analysis is of value to teachers and learners because of the detailed attention it pays to meaning, but that some adjustments to PDEV will need to be made for it to be maximally useful to learners.

Dedication

This work is dedicated to:

my late father

my mother for her prayers,

my lovely wife for her patience and love,

all of my family for their thoughts,

and my little cat

Acknowledgements

First and foremost, I am honoured to have been supervised by Professor Susan Hunston. Professor Susan Hunston is not just a supervisor; she is a friend and a brilliant academic coach. Her patience, advice, guidance, and sustained support throughout my doctoral journey have led me along the path towards completing this work. Now I can feel confident that I have become an independent researcher. I have always read that 'a PhD life is a lonely business'; though this opinion may generally be true, I did not feel that way under Professor Susan Hunston's supervision. There were times where I felt depressed or hopeless due to some critical life issues I went through. Professor Susan Hunston always listened to me and put me back on the right track. Indeed, this thesis would not have been completed without my supervisor, Professor Susan Hunston's support.

I also would like to extend my gratitude to my colleagues back home: Mr. Saad Al-Shamrani and Mr. Yosef Al-Wihaibi for their help and support with some administration issues concerning my scholarship. Not to forget my best friend Mr. Thamer Al-Harbi, who always kept in touch with me. Finally, I would like to thank all of my friends who kept in touch with me while I was away from home.

This thesis was copy edited for conventions of language, spelling and grammar by Editage.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

BNC   British National Corpus

CIA   contrastive interlanguage analysis

CLC   Cambridge Learner Corpus

CLCs   computer learner corpora

CPA   corpus pattern analysis

DVC   Disambiguation of Verbs by Collocation (project)

EA   error analysis

EGP   English grammar profile

GP   grammar pattern

ICLE   International Corpus of Learner English

LOCNESS   Louvain Corpus of Native English Essays

PDEV   Pattern Dictionary of English Verbs

PG   pattern grammar

SPOCA   subject, predicator, object, complement, and adverbial

UoM   unit of meaning

VACs   verb–argument constructions

List of Conventions

" " Double quotation marks: for short direct quotations from other publications (with long ones formatted as block quotes).

' ' Single quotation marks: for terminologies, senses or meanings, and emphasis on some parts from other publications.

UPPERCASE: for abbreviations, lemmas of the verbs analysed in this thesis, and verbs recorded in PDEV entries.

Italics: for examples and instances.

Chapter 1   **Introduction**

1.1   **Introduction**

This thesis is a contribution to research in learner corpora, and offers a novel approach to their analysis. Specifically, it uses the concepts in the Pattern Dictionary of English Verbs (PDEV) to describe learner corpora. This thesis has a two-fold aim: first, to see whether corpus pattern analysis (CPA) and PDEV (Hanks 2000) can be successfully used to describe learner corpora; and second, to see to what extent PDEV is useful to learners, teachers, or writers of curriculum materials. These aims will be achieved by using the concepts in PDEV to describe the use of verb patterns in both learner and native-speaker corpora and evaluating the results.

1.2   **Motivations of the research**

PDEV is an in-progress project designed to make very explicit links between pattern and meaning. It is the practical outcome of an approach to corpus analysis known as CPA. The concept of CPA chimes with current ideas about language that regard syntax and lexis as inseparable (cf. Hoey 2005; Hunston and Francis 2000; Sinclair 1991, 2004; Stubbs 2001, among others). PDEV is not the only work that has drawn on this principle; FrameNet[1] (Fillmore 2003), construction grammar (Goldberg 1995) and pattern grammar (PG) (Francis et al. 1996, 1998) are all other attempts to systematize this link. However, what motivates the choice of PDEV for this study is that PDEV is perhaps the most accessible and usable of these attempts. In addition, it is a very recently developed resource, and, to my knowledge, its data have not yet been investigated extensively.

---

[1] https://framenet.icsi.berkeley.edu/fndrupal.

Another motivation for this study is that I am a teacher of English and am always interested in the ways that learner competence and learner errors can be described. As an influential current view about language focuses on the link between pattern and meaning, and as CPA/PDEV is one of the methods by which this has been explored, I investigated learner corpora using it. My hope is that using such a concept to study learner corpora may lead to interesting conclusions, such as that CPA/PDEV can be a successful method of analysing the learner corpora or that it can be a useful tool in teaching (e.g. for learners, teachers, or even curriculum designers).

1.3 A brief introduction to PDEV

PDEV was developed by a team led by Patrick Hanks. It aims to describe thousands of verbs in English, recording for each verb the prototypical patterns and the meaning of those patterns. An example will be provided to demonstrate what it looks like (Figure 1.1).

Figure 1.1 Screenshot of PDEV entry for ENCOURAGE

Figure 1.1 shows a PDEV entry for the verb lemma ENCOURAGE, which includes useful information about the verb. It shows that ENCOURAGE has three prototypical patterns, each of them is different from the others in both form and meaning. Unlike, for example, the pattern grammar publications (e.g. Francis et al. 1996), the PDEV shows a lot of useful information about the individual verb and the individual patterns. Each pattern's description includes three elements: the pattern itself, its implicature, and an example taken from the

British National Corpus (BNC). In addition, each entry gives semantic information about elements of the pattern. For example, the subject of ENCOURAGE in all examples in Figure 1.1 is described with a semantic label: 'Human', 'Institution', or 'Eventuality'. Similarly, the object of ENCOURAGE is described as being 'Human' or 'Institution' in patterns 1 and 2 or 'Eventuality' in pattern 3. Furthermore, the PDEV entry provides grammatical information. An example is shown in Figure 1.1 where the *to*-infinitive occurs in pattern 1 of ENCOURAGE (e.g. *I encourage him <u>to work</u> harder*). Another example of grammatical information is in patterns 2 and 3, where ENCOURAGE is transitive (e.g. *The library encourages <u>the use of smartphones</u>*). Finally, information is given about the proportion of each pattern in the BNC. For example, Figure 1.1 shows that almost all the instances in the BNC are composed of pattern 1 (46.4%) and pattern 3 (47.6%), while pattern 2 accounts for only 5.6%. This allows the user to see which patterns are more frequently used and less frequently used by native-speaker writers.

As a learner of English, I find all this information useful; it reminds the user that s(he) can use abstract words such as Human or Institution or Eventuality. It also gives him/her the range of 'typical' patterns of ENCOURAGE, and how much each pattern is used by native speakers. The value of this information motivated the present research.

1.4 Previous studies

To date, there has been relatively little research undertaken using PDEV. The only exceptions I have come across are Bradbury (2016) and Rees (2018). Both these studies use the PDEV model to analyse corpora. Bradbury compares adult English, as represented by the BNC, with that of the English children in the 'BBC2013' corpus. She compares the uses of patterns of one verb, *blow*. Rees compares the patterns of verbs in a corpus compiled from research

articles in different academic disciplines. These studies both have a different focus from what I have done in this study; both these studies focus on native speakers of English. By contrast, my study relates to learners of English. In my view, it would be a great advantage to the PDEV project if it could be demonstrated that CPA/PDEV can be successfully used to describe the way the learners write, as far as verb pattern uses are concerned.

1.5 Research questions

Having set the aims of this thesis, three main research questions are posed as follows:

1. Can the concept of CPA/PDEV be used to successfully describe learner corpora?

2. How useful is PDEV to learners and teachers?

3. What are the pedagogical implications of applying PDEV to Chinese and Swedish learners?

1.6 Structure of the thesis

This thesis is divided into ten chapters. Chapter 1 introduces the aims and motivation of the study. It also gives a brief introduction to PDEV, which will be described in more detail in Chapter 2. Furthermore, it presents and differentiates the present study from those studies that use the concept of CPA/PDEV.

Chapter 2 includes several topics. The first part of the second chapter describes the origins and methodology of corpus pattern analysis using worked examples and illustrates how CPA can be used to create an inventory of typical patterns of use for a particular verb, as done in the Pattern Dictionary of English Verbs (PDEV). The second part of Chapter 2 discusses how CPA/PDEV is different from or similar to other ways of describing language. It also

introduces the field of learner corpora and reviews the key learner corpora studies that use the concept of the pattern or construction.

The third chapter introduces two concepts that are relevant to this thesis: error analysis (EA) and learner corpora (CL). It will discuss weaknesses of EA and how another new way of describing learners' language (CL) has emerged. In addition, some work done in learner corpora research will be described.

The fourth chapter introduces the methods used for corpus selection and data extraction. The size and structure of the corpora are discussed and issues around target corpora comparability are described. The reasons for the selection of the target corpora and the target verbs are justified. The extraction and interpretation of the target verbs are described step by step.

The next three chapters (5–7) each begin by presenting statistical data on patterns of the respective target verbs and then broaden the analysis to study uses of these verbs that are identified in the target corpora. Some specific questions which arise from the research findings are tackled in these chapters respectively. Chapter 5 focuses on divergent patterns of verbs identified in the Chinese sub-corpus of the International Corpus of Learner English (ICLE) data. The statistical data for each target verb are firstly presented, and then five selected target verbs are further investigated in terms of their divergence: AGREE, ALLOW, AVOID, ENCOURAGE and LEAD. Chapter 6 focuses on the notions of under- and over-representation of patterns of one target verb, ALLOW, as identified in one target corpus: the Swedish sub-corpus of the ICLE data. Chapter 7 broadens the research targets to investigate a verb the PDEV entry for which is not yet available, that is, SUGGEST. It investigates uses of SUGGEST in all four target corpora: learner corpora (ICLE-Chinese and ICLE-Swedish) and

normative ones (BNC and LOCNESS). This chapter is a good opportunity to demonstrate how CPA/PDEV can be used to identify patterns of verbs from scratch.

The findings based on the analysis of the results will be presented in Chapter 8. This will be followed by a comprehensive discussion of the pedagogical implications of this thesis, which will be presented in Chapter 9. The main research questions will be answered in Chapter 10, with a summary of the research. This chapter will also discuss the limitations of the work, suggest directions for future research, and close with final remarks.

Chapter 2   **Background and Literature Review**

2.1   **Introduction**

In this chapter, I provide the background necessary to understand the analysis and discussion in the following chapters, along with some of the previous literature. The aim of this chapter is to introduce the core concept of this thesis—corpus pattern analysis (CPA)—and the product of this concept, the Pattern Dictionary of English Verbs (PDEV). The chapter also discusses some language phenomena that underlie the concept of CPA/PDEV, and demonstrates how CPA/PDEV takes into account such language phenomena. Also, an assessment of PDEV and other dictionaries will be provided, some other approaches concerning language description that are relevant to CPA/PDEV will be discussed, and similarities and differences between them and CPA/PDEV will be shown.

Section 2.2 briefly reviews the history of the notion of linguistic pattern, acknowledging the key scholars whose work takes patterns into account. Section 2.3 gives the necessary background on CPA and PDEV, providing a detailed discussion of the approach. A comparison between PDEV and other corpus-based dictionaries will be presented in section 2.4. Further, section 2.5 will demonstrate how such language phenomena as collocation, colligation, and semantic preference underlie the approach of CPA/PDEV. This section will be followed by a discussion of how CPA/PDEV is similar to or different from other ways of describing language (section 2.6). Section 2.7 will conclude this chapter.

As this chapter will illustrate, words such as *collocation*, *colligation* and in particular *pattern* have a plethora of uses. Similarly, some phenomena, such as the co-occurrence of *agree* and *with* (e.g. 'I agree with you'), have been labelled with more than one term, increasing the possibility of confusion. To clarify matters, this thesis will use the following definitions:

- *Collocation* is the tendency for two lexical words to occur in close proximity. This may be measured statistically. Examples include *run swiftly, rectify errors, fierce storm.*

- *Colligation* is the tendency for a lexical word to occur with a grammatical class, such as a clause type, or with a specified grammatical word, such as *by*, *on*, *with*, etc. Although this can be measured statistically given sufficient information (e.g. see Gries and Stefanowitsch 2003), it is often identified by the simple observation of concordance lines. Examples include *agree to do something*, *suggestion that*, and *astonished at*.

- *Pattern* is a term used in various ways. Francis et al. (1996; 1998), for example, used it in the phrase 'grammar pattern' to refer to the phenomena which in this that thesis are called 'colligations' in this thesis. Hanks used the term 'corpus pattern' to refer to a combination of collocation and colligation. For example, in a clause such as *The wind blew the car off the road*, there is colligation between the verb, *blew*, the noun phrase, *the car*, and the prepositional phrase, *off the road*. In addition, however, there is collocation between *blew* and the physical object, *car*, and the indication of direction, *off the road*. In this thesis, 'pattern' is used in accordance with Hanks's usage unless indicated otherwise.

The background for these terms and other related concepts, such as lexical priming and semantic prosody, is given in this chapter.

2.2 The history of the notion of pattern

The notion of linguistic pattern first appeared in Hornby's (1954) *A Guide to Patterns and Usage in English*, a pedagogical grammar which takes a novel approach to describing English

for teachers and learners by classifying English sentences into 25 different patterns and providing examples of each.

A few decades later, the notion of patterns was further developed by scholars such as Sinclair (1991), as well as by the founders of pattern grammar (Hunston and Francis 1998, 2000).

2.2.1 Sinclair (1991)

In his book *Corpus Concordance Collocation*, Sinclair (1991) stresses that researchers should base their investigation of language on enormous amounts of authentic data, which by the time he was writing were storable in an electronic form. Using this method, Sinclair "sets out an agenda for a radical departure in the description of English" (cited in Hunston and Francis 2000:14). Specifically, he proposes the notions (defined below) of lexical patterning, units of meaning, the idiom principle, and other allied ideas, and demonstrates that "there is a close correlation between the different senses of a word and the structures in which it occurs" (ibid:53).

Sinclair's two principles which account for how text is patterned, and explain how meanings appear from such patterned text, are the 'open-choice principle' and the 'idiom principle' (Sinclair 1991). The idiom principle can be explained as follows:

> [E]ach word in the text is used in a common phraseology, meaning is attached to the whole phrase rather than to the individual parts of it, and the hearer or reader understands the phrase as a phrase rather than as a grammatical template with lexical items. (Hunston 2002:143)

9

To illustrate this, consider the following example taken from the BNC:

2.1 *And yet, I never, <u>to the best of my knowledge</u>, cut my toenails*. (BNC)

Here, the words in the phrase *to the best of my knowledge* together constitute one single choice and meaning; they are used to say that one believes something to be true but that one is not entirely sure. In other words, the writer or reader understands the meaning of the phrase as a unit, not as independent words. Incidentally, it is also an example where the grammar is odd and idiosyncratic—we cannot say other things like 'to the worst of my understanding'.

In the open-choice principle, "words are treated as independent items of meaning" (Sinclair 1991:175). In the same example above, for instance, items such as *and, yet, I, never, cut my toenails* each convey an independent meaning. Sinclair also refers to this principle as a 'slot-and-filler frame' (Sinclair 1991:109). This frame is like the grammatical frame in that a slot can be filled with different words; for example, *cut* could be followed by *toenails* or by *trees* or *fingers*.

These two ground-breaking principles together led to an increase in the investigation of texts to uncover a number of phenomena (e.g. Renouf and Sinclair 1991, collocation frameworks; Nattinger and DeCarrioco 1992, lexical phrases; Biber et al. 1999, lexical bundles; Stubbs 2007, n-grams; Granger and Bestgen 2014, bi-grams; González Fernández and Schmitt 2015, collocation, among many others).

Sinclair's observations raise a question as to whether such phenomena can also be observed in L2 learner texts. A number of key scholars have attempted to investigate learner language to

see whether learner language can also be idiomatic (see Pawley and Syder 1983, lexicalized sentence stems; Gläser 1998; De Cock et al. 1998; Granger 1998b; De Cock 2004, recurrent sequences of words; among others).

2.2.2 Francis (1993, 1995), Hunston and Francis (1998, 2000), and pattern grammar

The association between sense and structure (or between 'meaning' and 'pattern') observed by Sinclair (1991) was further explored in the work by Francis (1993, 1995) and Hunston and Francis (1998, 2000). These researchers observe that "certain patterns 'select' words of particular meaning" (Hunston and Francis 2000:29). For example, many verbs occurring in the pattern 'V n *to* n' (see Francis et al. 1996) can be classified under the following meaning groups: (1) *communication* (e.g. *announce*, *confess*), (2) *adding* (e.g. *connect*, *attach*) (3) *change* (e.g. *convert*, *reduce*), (4) *attracting someone* (e.g. *draw*, *recommend*) (Hunston 2002:105–106).

Based on this work, these researchers have developed an approach called pattern grammar (Francis et al. 1996), a corpus-based method based on describing the syntactic environments of lexical items. In it, each word is presented with a set of patterns, which describe particular kinds of contexts in which they are used.

Hunston and Francis (1998:45) code the complementation patterns of all the words in the *Collins Cobuild English Dictionary* (1995). For example, the convention V n *to*-inf, means Verb is followed by Noun followed by verb with *to*, as in *I wanted her to leave immediately*. This description, as the researchers mentioned, is for economy, so that the notation could fit in the *Cobuild* print dictionary and thus be accessible to learners. More importantly, this has resulted in lists of verb pattern sharing a limited set of meaning groups.

11

Hunston and Francis (2000:37) defined the patterns of a word as follows:

> The patterns of a word can be defined as all the words and structures which are regularly associated with the word and which contribute to its meaning. A pattern can be identified if a combination of words occurs relatively frequently, if it is dependent on a particular word choice, and if there is a clear meaning associated with it.

The first part of the definition above indicates that there is a correlation between lexis and grammar (or between sense and syntax) which produces a meaning that is unique to the 'pattern' of the word. The second part of the definition sets three criteria upon which such a pattern can be identified: (1) frequency of words, (i.e. the lexis-and-grammar pattern should occur frequently); (2) dependency of words (i.e. the node word should be significant to the words it appears with); and (3) forming a meaning (i.e. the close relation between lexis and grammar or sense and syntax should result in an unambiguous meaning).

For many years, the concept of pattern grammar has remained consistent, although Francis (2015) proposed some updates to the list of patterns. Other works have focused on the application of pattern grammar to lexicography and language teaching (Hunston 2002, 2004). The concept of pattern has also been used as the basis of studies on academic discourse (Charles 2006a, 2000b). More recently, Hunston addressed the theoretical concepts behind pattern grammar (Hunston 2014) and proposed that the concepts of pattern grammar and construction grammar be aligned (Hunston and Su 2018; Hunston forthcoming). This is in line with the study by Ellis et al. (2016), wherein pattern grammar was used as the input in a study on constructions in second language acquisition.

Pattern grammar is important for several reasons. First, it uses the notion of pattern, in that it does not separate between grammar and lexis, but instead treats both grammar and lexis as one part. Second, it groups verbs that appear to share particular patterns and sorts them into meaning groups, which yields useful information about semantics. Third, it uses conventions in describing patterns, which can be helpful for the user to follow.

However, pattern grammar has limitations. It only looks at words that follow the word in question for the identification of a pattern (i.e. words that follow a verb). Looking at what comes both before and after a node word (verb) would lead to more useful results for the node word, and more distinct patterns and meanings of the same node word. Second, pattern grammar identifies constructions that are generalisable (e.g. V n, V adj, V prep, V adv, etc.). For example, the word *fire* in the sentences *fires a gun* and in *fires the employee* both represent the construction V n, which does not help in distinguishing between these two meanings of *fire*. We will return to this point below (section 2.4.1). In addition, the V n construction makes no distinction between sentences like *He was a teacher* and *He wrote a book*, even though there is a grammatical difference in the two sentences: the former contains subject, verb, and complement, while the latter consists of subject, verb and object. Such a grammatical difference cannot be captured via the V n construction.

Such limitations are answered by another innovative approach, introduced in the next sentence: corpus pattern analysis (CPA), which this thesis is based on and which yields the Pattern Dictionary of English Verbs (PDEV) that this thesis employs.

2.3   Corpus pattern analysis and the Pattern Dictionary of English Verbs

Corpus pattern analysis was developed by Hanks as a method of "mapping meaning onto words in text" (Hanks 2004:87). Like other pattern-based approaches, it is based on the identification and recording of lexical patterns rather than on defining individual words. Thus, CPA agrees with the general notion of patterns that was introduced above, in which there exists a close relation between lexis and grammar or sense and syntax which results in an identifiable pattern, and this pattern constitutes a unique meaning not reducible to the meanings of the individual words. However, CPA differs from pattern grammar in certain aspects, to be discussed shortly.

CPA was originally developed to test Hanks' (2013) theory of norms and exploitations (TNE). This is an approach that examines corpus data to identify what the typical pattern or 'norms' of a word are and hence also what is abnormal or 'exploitation' usage of it (Hanks 2013). According to Hanks, there are prototypical or collocational patterns that give lexicographers a benchmark against which to explore how a particular instance of collocation is prototypically used. By 'exploitations', he then means when writers deliberately use these collocational norms in a non-typical use of a pattern. Hanks (Hanks 2013:17) explains the significance of the TNE approach, and of CPA, as follows:

> Despite practical difficulties of application of the kind just outlined, the **theory of norms and exploitations** (henceforth *TNE*) offers at least one benefit for the lexical analyst. It is this: it is a principled approach according to which exploitations can be identified as such and set on one side, leaving the distinctive patterns of normal phraseology associated with each word to stand out more clearly. If meanings (or translations, or implicatures) are associated

14

with words in context—that is, with phraseological patterns—rather than with words in isolation, some well-known linguistic problems are largely solved. For example, one of the problems confronting word-sense disambiguation programs has always been a combinatorial explosion of possibilities.

With TNE theory, Hanks aims to identify what is 'typical to say' as well as what is 'possible to say', on the basis of observation of data from the British National Corpus (BNC). By doing so, one of Hanks's goals is to come up with an inventory of verbs that contains only typical patterns: the Pattern Dictionary of English Verbs (PDEV).

Since the method for CPA of verbs differs from that for nouns (Hanks 2004) and the focus of the present study is on verbs, this section focuses on verbs and how verb patterns can be identified based on the CPA procedure.

PDEV is the direct outcome of CPA and the testing ground for Hanks' hypothesis. Since PDEV is the basis of this study, it is discussed in detail in the next section.

2.3.1   Pattern Dictionary of English Verbs (PDEV)

The PDEV project is the outcome of the CPA method; according to Hanks (2013:405), it represents "a systematic elaboration of the Sinclairian approach to lexical analysis in the field of verb meaning, which is fundamental to text meaning." The basis of PDEV is Hanks' (2013) assertion that the meanings of most verbs can be 'identified by a combination of valency and collocates' in which such a combination has a distinctive meaning (Hanks 2012:58).

PDEV employs the British National Corpus (BNC)[2], which was initiated in 1991 and completed in 1994 and represents a diverse selection of British English. It consists of 100 million words: 90% in written English and 10% in spoken English. However, only a sub-corpus of the BNC has been used for the PDEV, namely, the BNC50 (see Baisa et al. 2015). This sub-corpus consists of 50 million words of written English; it excludes the spoken corpus, which has been argued to be problematic by Hanks (2004). He suggests that language acts like reading out a paper in a lecture or reading the TV news from a teleprompter are not appropriately viewed as spoken production. The natural process of spoken language is more difficult to record and is harder for a lexicographer to interpret. The BNC50 corpus includes topics such as arts, world affairs, social sciences, belief and thought, applied science, natural science, commerce, and finance. The next section will introduce the method used to come up with the Pattern Dictionary of English Verbs.

2.3.2 The Disambiguation of Verbs by Collocation (DVC) project

The Disambiguation of Verbs by Collocation project (DVC) is a product of CPA, by which PDEV has been established. The main purpose of this project is '[to expand] the PDEV to a large scale' (El Maarouf et al. 2014b:1001). Some of the objectives[3] of the DVC are as follows:

- To discover, through CPA, the characteristic patterns of use for English verbs and link meanings to each pattern.

- To build a pattern dictionary (that is, PDEV)

- To build an ontology of semantic types

---

[2] http://www.natcorp.ox.ac.uk.
[3] These objectives were extracted from the DVC website at this link:
http://rgcl.wlv.ac.uk/research/dvc-disambiguation-of-verbs-by-collocation.

In the following section, I will illustrate the first and the second points with examples and discuss the third point afterward. The aims of the following sections are twofold: (1) to show how the CPA operates in practice and (2) to show what the PDEV infrastructure consists of. At the end of the sections, the reader should understand better what the CPA is and how the PDEV looks, which is crucial to the present study.

2.3.3 Previous studies and current trends in CPA

To date, there has been relatively little research undertaken using PDEV. The only exceptions that I have been able to find are Bradbury (2016) and Rees (2018). Both of these studies used the PDEV model to analyse corpora.

Bradbury compared adult English, as represented by the BNC, with child English in the 'BBC2013' corpus by focusing on a comparison between the uses of the patterns of one verb, *blow*. One of the key findings showed that lexical items that populate the subject slot of one pattern of *blow* were found in the BBC2013 sample but not found in the BNC. These items are *whistle* in *whistle blew*; *dragon* in *dragon blew fire*. Bradbury showed that CPA was highly effective in distinguishing adult and child uses of an individual word.

Rees (2018) examined differences in meaning and use of Academic Wordlist (AWL) across academic disciplines employing the CPA as a method. One of the key findings in this regard is that the semantic types that populate the subject slot of *assemble* in the Management sub-corpus are different from in Microbiology, where Physical Objects and Microbiological Processes *assemble* Objects, while in Management People or Organizations *assemble* Objects and Groups. Rees showed that CPA was feasible in demonstrating that "a large number of the AWL words have discipline-specific meanings, and that these meanings are conditioned by

the syntagmatic context of the AWL item" (Rees 2018:61).

CPA is conducted manually, which requires a significant amount of effort and time. One of the recent development in the field is transforming CPA into a fully automatic tool, thus a number of recent studies have assessed the various aspects of the automatic application of CPA (cf. Elmaarouf 2013; Elmaarouf et al. 2014a; Baisa et al. 2016).

Further, one of the goals of changing CPA into a fully automated process is to contribute to the development of machine translation (cf. Elmaarouf et al. 2014a; Baisa et al. 2016; Hanks 2018). Baisa et al.'s (2016) pilot study, for instance, is reportedly the first attempt to link CPA verb patterns across different languages, such as Spanish, German and Italian (ibid).[4] The authors reported that the CPA matching between Spanish and English verb patterns showed an 80% precision rate.

Hanks et al. (2018) conducted a case study of CPA for English-French translation. In their study, they focused on the English verb, *bite*, and its French equivalent, *mordre*, to "identify their features and divergences by applying CPA to corpora." (ibid: 100). They analysed a sample of 500 lines of *bite* from the BNC and 500 lines of *mordre* from the frTenTen corpus. Particularly, they studied the multi-word expressions of these two verbs "to speed up their extractions for corpora" (Hanks 2018:104). One of their results from applying CPA to both languages showed that the preferred lexical items that fill the subject slot of the English verb, *bite*, are *mosquitoes* and *bugs*, while the lexical items that fill the subject slot of the English verb, *sting*, are normally *bees*, *wasps* and *hornets*. On the other hand, the French equivalent

---

[4] For their current pilot study, only one language was used: Spansih.

translation of *bite*, that is, *mordre*, was not found to occur with *mosquitoes* or *bees*; "French speakers prefer to use *piquer* 'sting' for most of 'flying entity aggression'" (Hanks et al. 2018: 116).

Such studies seem to indicate that CPA would be a promising procedure to enhance the quality and precision of machine translations. However, more research in this area is needed to achieve this future goal.

2.3.4 Identifying verb patterns

According to Hanks (2013:92), "a pattern consists of a valency structure … together with sets of preferred collocations". The framework that he adopts for analysing patterns is 'systemic grammar'. In this framework, any text would be "seen as being made up of sets of clauses," where 'a clause typically consists of one or two' of the SPOCA clause roles: subject, predicator (or verb), object complement, and adverbial (Hanks 2013:95). Hanks refers to preferred collocations as semantic types that fill the available slot (e.g. subject slot, object slot). The semantic types are presented in a hierarchical ontology (ibid.). An example of a verb and its patterns is shown in Figure 2.1 below.

Figure 2.1 Screenshot of the entry for the verb NOTE in the PDEV database[5]

As shown in Figure 2.1, each pattern consists of three parts: Pattern, Implicature and Example. (It is important to mention here that since the PDEV project is still in progress, changes maybe be made to the notation at any time. For example, originally the words Human, Institution, Eventuality, and Visible Feature were put between double square brackets, indicating that they are semantic types. However, by 2017 the brackets had been replaced by colour.

Based on CPA, there are four patterns of the verb *note* (see Figure 2.1). The number on the left shows the pattern number (e.g. 1); the number on the opposite right indicates the relative frequency of that pattern in the selected sample for analysis (e.g. 38.0%). What appears below each pattern is the implicature.

---

[5] http://pdev.org.uk/?verb=note#browse?q=note;f=A;v=note.

2.3.5 The nature of hierarchical PDEV ontology

PDEV is an on-going project. This means that the work is not yet finished and that more details might be added later. In total, at the time of writing this thesis, PDEV ontology contains 253 semantic types. These semantic types are established hierarchically, forming the ontology. The PDEV ontology consists of five main categories: Entity, Eventuality, Group, Part and Property. Each of these categories has a number of sub-categories. For example, Entity contains five sub-categories, each of which is given a title: Abstract Entity, Energy, Physical Object, Particle and Self. Eventuality includes two sub-categories: Event and State of Affair. Group contains four sub-layers: Human Group, Vehicle Group, Animal Group and Physical Object Group. Part consists of seven sub-categories: Language Part, Music Part, Physical Object Part, Speech Act Part, Document Part, Movie Part and Recording Part. Finally, Property includes nine sub-categories: Cognitive State, Role, Visible Feature, Character Trait, Injury, Institution Role, Pace, Use and Weight.

Each of these sub-categories has a number of sub-categories, each of which contains a list of lexical items that belongs to this semantic type. For example, the sub-category, Physical Object, has three semantic types: Animate, Inanimate and Plant (Figure 2.2).

Figure 2.2, for example, illustrates the hierarchy around the semantic type Artifact. The '+' symbol before 'Artifact' shows that extended semantic types are rooted below it. Clicking on the '+' symbol before the semantic type 'Weapon', the symbol changes to '-', in which further semantic types are nested, such as 'Bomb', 'Firearm', and 'Projectile'. If a user clicks on any of the said semantic types below 'Weapon', a small box appears on the right of the ontology window which includes a lexical set or collocates which belong to the said semantic types. For example, some of the lexical items that belong to semantic type Projectile are 'shot', 'bullet', 'missile', and 'rocket'; Firearm: 'gun', 'riffle', and 'pistol'.

Figure 2.2 A snapshot of the PDEV ontology of semantic types

Some of the limitation that PDEV ontology has is that a number of semantic types do not

include lexical items. This could be due to that PDEV is not finished yet since it is an in-

progress project. Another limitation is that some semantic types in the ontology have few

lexical items. For example, not all lexical items are listed under a given semantic type. For

example, semantic type Document Part has only 2 lexical items: 'page' and 'chapter'. (This

semantic type is assumed to include a number of lexical items that belong to document part).

One of the explanations is that it could be due to the small samples analysed by the PDEV

team.

2.3.6   Valency and PDEV

Valency is an essential part of mapping meaning onto verbs in the CPA/PDEV approach.

Particularly, valency or verb arguments allow CPA to identify different patterns of a given

verb.

Valency is not a new concept; it is a common concept with European origins. Herbst et al.

(2004:xxiv) pointed out, "The basic assumption of valency theory is that the verb occupies a

central position in the sentence because the verb determines how many other elements [or arguments] have to occur in order to form a grammatical sentence." For example, some verbs, such as *ping*, are intransitive and have only one argument, that is, the subject, as in *The doorbell pinged*. On the other hand, some verbs, such as *yield*, are transitive and have two arguments, that is, the subject and object, as in *Investigation of these predictions has yielded mixed results*. Valency involves the study and specification of each verb, the number of arguments that have to occur and what these arguments are. Hanks's work is related to the concept of valency because in both cases there is a focus on what the argument is. Hanks's work gave more details about valency and the valency of each individual word.

In the CPA, different meanings of verbs are motivated by the presence of semantic types (Hanks 2013). For example, based on the CPA, there are 16 patterns of the verb *fire*. Because of space limitations, I have selected only two patterns for illustration here, patterns 1 and 5. Both of these patterns are transitive. The valency of the verb includes the subject and the direct object, in which the semantic types are found. In pattern 1, for example, the subject slot is filled with the semantic type 'Human' and the object slot with 'Firearm'. In this sense, the 'Implicature' (for our purposes, the definition) of pattern 1 is *physical action in which Human holds Firearm (e.g. gun, rifle, pistol) and causes it to discharge a Projectile (e.g. bullet, shot, missile) towards Physical Object (e.g. target)*. The words between the round parentheses constitute a lexical set of words that belong to each semantic type (see Figure 2.2).

Pattern 5, on the other hand, represents an entirely different meaning of the verb *fire*. The Subject slot is filled with the semantic type 'Human 1', and the object slot here is filled with the semantic type 'Human 2'. The 'Implicature' of the pattern denotes the sense of dismissal: 'Human 1' dismisses an employee, 'Human 2', from a job. Unlike pattern 1, the subject slot is

filled with two semantic types, 'Human' and 'Institution'. This is what Hanks (2013) refers to as 'Alternation of Semantic Type', where the semantic type can alternate within an overall pattern in relation to a set of lexical items (i.e. a set of lexical item that belongs to each semantic type that is presented in the ontology of semantic types). Hanks also maintains that such an alternation is extremely common in the subject slot. It is important to mention that this type of alternation does not change the 'Implicature' of the pattern (ibid.).

In Figure 2.3 below, which illustrates the entry for the verb *arrest*, there is an '=' symbol after 'Institution', followed by 'Police'. This is what Hanks (2013) refers to as the semantic role. According to Hanks (2013), the semantic role is different from the semantic type; for instance, in Figure 2.3, the semantic type of the subject slot is 'Human', but its semantic role is 'Policeman'. Similarly, the semantic type of the object slot is 'Human 2' while its semantic role is 'Suspect'. In other words, the semantic role further narrows down the semantic type and serves to activate the meaning of the verb *arrest* (i.e. the policeman seizes the suspect by legal authority and puts the suspect into custody). It is unusual to see an ordinary person, who has no legal authority, arrest other people in this way.

Figure 2.3 A customised screenshot of the entry for the verb ARREST in the PDEV[6]

2.3.7 Implicature and its use in PDEV

The word, 'implicature,' was first used by Grice (1968; 1975) in his work on pragmatics and maxims. He used the word to mean what an utterance implies but does not state. For instance, if a colleague asked, *How good is the student?* and one answered, *He is always punctual and*

---

[6] http://pdev.org.uk/#browse?q=arrest;f=A;v=arrest.

he dresses smartly, the implicature would be that *he is not a good student* because *dressing smartly* has nothing to do with a student's academic capabilities. Further, while *being punctual* is good, the colleague was not asking about punctuality. The statement itself did not explicitly express this implicature.

However, Hanks used this term differently, referring to the pragmatic meaning of individual words rather than the pragmatic meaning of utterances. For example, throughout the PDEV inventory, the implicatures of PDEV patterns state explicitly what a pattern indicates. For example, the implicature of the pattern, *Human or Institution agrees to-infinitive*, is *Human or Institution indicates that he, she, or it is willing to undertake something*. The implicature is anchored to the pattern in that all elements stated in the pattern are stated in the implicature description.

Therefore, although the two uses of 'implicature' have something in common they are not the same.

2.4   Assessing the relative usefulness of the PDEV

This section will compare the PDEV to other pattern-based systems, highlighting the key differences between them. Section 2.4.1 discusses the differences between PDEV and pattern grammar (PG), by Hunston and Francis (1998, 2000), while section 2.4.2 compares PDEV to several other pattern-based dictionaries. Ultimately, it will be shown that PDEV has several important advantages over the other systems.

2.4.1   PDEV v. pattern grammar

As discussed earlier in this section, PDEV and pattern grammar are both based on the premise that lexis and grammar are not separate systems and that lexicographic description should

focus on the association between meaning and pattern (Francis 1993, 1995; Hornby 1954; Hunston and Francis 1998; Sinclair 1991; Hanks 2004). However, the two systems differ in their approaches to pattern analysis. The first difference between PG and PDEV is their different purposes. PG aims to identify and group all different verbs that are used in the same way (i.e. all verbs that are used in particular patterns) (see section 2.2.2 for details), whereas PDEV seeks to identify and distinguish all verb patterns that are different from one another (see section 2.3.4 for detail).

Another difference is the treatment of the grammatical subject in the two approaches. The examples used here to illustrate the PG approach were selected from the Cobuild grammar series, which consists of two volumes, one of which is *Collins Cobuild Grammar Patterns 1: Verbs* (Francis et al. 1996). In the PG framework (Francis et al. 1996) a verb pattern includes, in most cases, the verb and the words that come after it. These words might be a noun phrase, an adjective phrase, a prepositional phrase, or a finite or non-finite clause (see section 2.2.2 for examples). Crucially, in most cases in PG, the subject does not play a role in the meaning of the pattern (Hunston and Francis 2000). The only cases when it is included in the pattern is where the subject is restricted—for example if it is always *it*, or always plural. In PDEV, on the other hand, as we have seen above, the subject is always an integral part of the pattern. In this aspect, PDEV reveals patterns that cannot be identified by the GP approach, as it provides a more detailed description without increasing the complexity of its ontology. For instance, without including the subject slot in the pattern, the PG approach would have trouble encoding and analysing the difference between the two meanings of the verb *fire* illustrated above. The different patterns and meanings of verbs can be identified by using the same ontology of semantic types as for the object slot, as shown in Figure 2.4.

Figure 2.4 A customised screenshot of the entry for the verb FIRE in the PDEV

Here, the subject in pattern 2 of the verb *fire* is filled with the semantic type 'Human' and the

object slot is filled with the semantic type 'Projectile'. In pattern 5, the subject slot is filled

with the semantic type 'Human 1' or 'Institute' and the object slot is filled with the semantic

type 'Human 2'). In this sense, PDEV is also more informative than PG, since the latter

system would not be able to capture this important difference.

The third crucial difference between the two approaches is how they record patterns with

different meanings. As was mentioned earlier in this chapter, Francis (1993, 1995) based her

observation of the association between sense and pattern on the fact that words appearing in a

particular pattern can be classified under a limited set of meaning groups (Hunston and

Francis 2000:29). That is, in GP those verbs are grouped that have similar meanings when

they are used in a particular pattern. For example, verbs used in the pattern V *to*-inf, such as

hasten, *hurry*, *hustle*, *rush*, and *scurry*, are grouped under the 'hasten' meaning.

In PDEV, by contrast, verbs are not grouped according to shared patterns. Instead, each verb

entry consists of one or more pattern, and each pattern of that verb has a unique meaning

linked to it (see Figure 2.5 below). Thus, the verbs in Figure 2.5 below cannot be divided into

meaning groups, since each of the verbs has different sets of patterns and meanings.

Figure 2.5 Verb entry in PDEV

I would argue that pattern grammar is more effective when it deals with identifying long patterns such as V *to*-inf, V n *to*-inf, V n with n, etc. than it is with short patterns, particularly transitive verbs, such as V n. PDEV, on the other hand, does better than PG when it comes to identifying patterns of transitive verbs, especially when such patterns include different subjects and objects, as shown in Figure 2.4 above. It will be beneficial if the concept that allows this advantage of PDEV is applied to *Cobuild* in the same way as was done by Francis et al. but using the PDEV ontology to group all transitive verbs that have the same semantic types in the subject and object slots. For example, all verbs whose subjects and objects are 'Human' can be grouped (e.g. 'Human' V 'Human', 'Human' V 'non-Human'), and so on. By doing so, more specific patterns and meanings would be identified than what has been recorded in Francis et al. (1996).

2.4.2 PDEV v. other dictionaries

In this section I compare PDEV to four recent dictionaries of English collocations: the

Macmillan Collocations Dictionary for Learners of English (MCD) (Rundell 2010), *Oxford*

Collocations Dictionary for Students of English (2002, 2008), the *BBI Dictionary of English*

Word Combinations (Benson et al. 2010), and the *LTP Dictionary of Selected Collocations*

(Hill and Lewis 1998). I will briefly describe each of these dictionaries in turn, and then will

evaluate their strengths and weaknesses against those of the PDEV.

The Macmillan Collocations Dictionary for Learners of English (MCD) is aimed at learners

of English and those using English in an academic or professional environment (Rundell

2010). It contains over 4,500 headwords (Coffey 2011). Its entries can be nouns, verbs, or

adjectives. For verbs MCD makes a distinction between single-word items and phrasal verbs,

and has separate entries for each. Other than those and a small group of compound nouns,

almost all entries in MCD are single words.

The collocates for each entry are grouped according to the lexico-grammatical structure

formed by the collocational phrase; for example, the lexico-grammatical groups for the noun

design are *adj n* and *n n*. The collocates are further grouped into semantic sets; for example, at

design adj n, there are five sets, labelled in the dictionary as: 'types of design', 'elegant',

'original or 'modern', 'simple and always fashionable' and 'easy to use' (Coffey 2011). These

semantic sets are somewhat similar to the semantic roles in PDEV (discussed earlier in this

chapter), but are more specific to each entry[7]. Each group of collocates has an example

demonstrating its use in a context; some entries or collocate sets also contain 'usage notes'.

---

[7] It should be noted here that MCD does not include ontology as PDEV does.

The *LTP Dictionary of Selected Collocations* (Hill and Lewis 1998) (LTP) is based on the Brown Corpus, compiled in the 1960s, and is aimed at intermediate and advanced learners of English. It consists of 50,000 collocations, which are grouped alphabetically. LTP differs crucially from other dictionaries discussed here: it is divided into two sections, one for nouns and one for adverbs[8], so the language learner must be at a relatively advanced stage to be able to use this dictionary effectively, as they need to decide in what section of the dictionary to look up the collocation. LTP has been criticised in the literature for inconsistent coverage of the data. For instance, Lea and Runcie (2002:825) argue that while it sometimes includes peculiar co-occurring words such as '*smouldering suspicion*' and '*fritter away the gains*' (neither of which is found in the BNC), it fails to include many more frequent collocations. Thus, as Lea and Runcie (2002:825) point out, the LTP authors "must have been relying on single citations (or none at all) for some of their collocations."

The *Oxford Collocations Dictionary for Students of English* (2002, 2008) (OCD) is two times as the size of MCD2, consisting of 250,000 collocations (Coffey 2011). Like some other dictionaries, it is based on the BNC corpus, with entries chosen largely based on frequency (Lea and Runcie 2002). Similarly to LTP, it takes the noun as the basic unit on which a phrase is built, and collocates are grouped accordingly based on their semantic categories. The way OCD presents its material is similar to that in LTP, discussed above, and BBI, to be discussed shortly. "Each headword is divided into a number of 'slots' according to the part of speech or function of the collocating word." (Lea and Runcie 2002:826). Some examples of entries with slots are below (from Lea and Runcie 2002):

adjective noun: *bright/harsh/intense/strong light*

quantifier noun: *a beam/ray of light*

---

[8] The adverbs section includes verbs and adjectives that collocate with adverbs.

verb noun: *cast*/*emit*/*give*/*provide*/*shed light*

noun verb: *light gleams*/*glows*/*shines*

However, as these authors point out, these slots are only 'quasi-grammatical' and are based more on the idea that the learners would want to express rather than the actual grammatical category. Thus, the dictionary allows for pre-modifying nouns to appear in the 'adjective position', or in other words, pre-modifying nouns (***tax*** *benefit*; ***takeover*** *bid*), and predicative adjectives can appear in the *noun verb* combination (e.g. *compensation is **payabl**e*).

Another prominent dictionary that is based on patterns rather than lexical items is the *BBI Combinatory Dictionary of English* (BBI). It is also partially based on the BNC, and, like other dictionaries discussed here, it offers the learner the word and its phraseological information. Every entry in BBI is divided into three parts: Definition, Exemplification and Usage Notes. BBI emphasizes the distinction between 'grammatical collocations' and 'lexical collocations', which are listed differently in the dictionary (Benson et al. 2010). Benson et al. used the term, 'grammatical collocations,' to describe the collocation between a word belonging to a lexical word class, such as *admiration*, *acceptable* and *hostility*, and a word belonging to a grammatical word class, such as *for*, *to* and *towards* (e.g. giving *admiration for*, *acceptable to*, and *hostility towards*). This collocation between lexical and grammatical words is sometimes referred to as 'colligation,' the term used in this thesis (see section 2.5.2). Grammatical collocations (e.g. *admiration for, acceptable to, hostility towards*) are listed based on the dominant word. Lexical collocations (e.g. *to put up resistance*, *to override a veto*), however, are listed differently; below the authors describe how to find them in the dictionary:

In order to find lexical collocations in the BBI, follow this step-by-step procedure: if there is a noun in the collocation, look under the noun; if there are two nouns, look under the second; if there is no noun, look under the adjective; if there is no noun or adjective, look under the verb. (Benson et al. 2010:xiv)

Unlike MCD and OCD, BBI includes information not only about phraseology but also valency (Benson et al. 2009). The two large classes of lexical pattern collocations and grammatical collocations are further divided into subgroups, for example, L1, L2, L3, G1, G2, G3[9].

Another crucial difference between BBI and the other dictionaries described here is that it is the only one that is not based on a large corpus of data. As McGee (2012:330) point out, it is "not corpus-based, but rather 'corpus-refined'". That is, it is based on the authors' native-speaker intuition, which is then backed up by examples and data from the corpus.

The *Cobuild Dictionary*, which was originally edited by Hanks and has most recently been republished in 2015, is possibly the most comprehensive English learners' dictionary to date. In its treatment of collocations it follows in many ways the other dictionaries discussed here and not PDEV. However, the senses are distinguished and defined much in the same way as Hanks does in PDEV.

---

[9] L1 = lexical collocations group 1; G1= grammatical collocations group 1, etc.

2.4.3 Summary

This section has compared PDEV to four other pattern-based dictionaries: LTP, MCD, OCD, and BBI. I have demonstrated that PDEV has many advantages over these previous systems. The entries in PDEV are more economical and more flexible at the same time. Because of its restricted ontology of semantic types, which is not used in any of the other systems (with semantic roles being subtypes of semantic types), it allows users to easily access a wide range of meanings of each verb, thus allowing space for their creativity while teaching them to adhere to the rules and patterns available. None of the other dictionaries described here offer the same combination of economy and flexibility while conveying such a vast amount of information for each entry. Another crucial difference between PDEV and the other dictionaries discussed here is that PDEV is the only one that shows the prevalence percentage for each pattern in the corpus. To the best of my knowledge, no other pattern dictionary today has this important feature.

2.5   Corpus pattern analysis and language phenomena

This section has two objectives. First, it relates CPA and PDEV to the three language phenomena underpinning concepts in the definition of words, that is, collocation, colligation, and semantic preference. Second, it shows similarities and differences between CPA/PDEV and related concepts such as 'units of meaning' (Sinclair 1996, 2004), 'lexical priming' (Hoey 2003, 2004, 2005; Hoey et al. 2007), and 'Constructions' (Goldberg 1995).

2.5.1 Collocation and meaning

'Collocation' is a term first introduced, as is well-known, by Firth (1951)[10]. He proposed that the meaning of a word can be identified by its collocates. By the passage of time, there have been statistical measures of the frequency of collocation which basically compare the expected frequency with the observed frequency of the actual collocation between the two words. According to Sinclair (1991:170), a collocation is "the occurrence of two or more words within a short space of each other in a text [which] can be important in the lexical structure of the language because of being frequently repeated." Stubbs (2001:64) redefines Sinclair's definition as follows: a collocation is "a node-collocate pair: this is a purely lexical relation … which ignores any syntactic relation between the words." For example, if the verb lemma FIRE is run in the corpus query processor (CQP) web[11], the log-likelihood test, within a span of -4 and +4, shows a list of words, in descending order of significance, that FIRE collocates with words such as *shot(s)*, *gun(s)*, *worker(s)*, *people*, appear at the top the list. The list can help the analyst to identify the distinct meanings of *FIRE*. By looking at these collocates, for instance, we can conclude that FIRE has two different meanings. Some of these collocates, such as *shot(s)* and *gun(s)*, help constitute the concrete meaning of FIRE (e.g. *The police fired shots*); other words, such as *workers, people,*, help constitute the abstract meaning of FIRE (e.g. *The company fired all workers last year*).

However, some aspects of this definition of collocation have been challenged. For instance, certain researchers warn that some important collocates may exist outside the -4:+4 window (Kennedy 1998; Altenberg and Granger 2001). Some pronouns, for example, may be found within the specified span to refer to lexical items outside of the 4:4 span, as shown below:

---

[10] It is introduced in the article titled 'Modes of Meaning'.
[11] See http://bncweb.lancs.ac.uk.

2.2 *The ballistic expert's evidence was that the defendant's **gun** was fitted with a hammer block safety mechanism, which was in proper working order, so that **it** could only be **fired** if the trigger was fully pressed back.* (BNC)

If an analyst is making an automatic list of collocates of the verb FIRE, this particular instance of *gun* will not show up in the list, although *gun* collocates with FIRE. If the purpose of obtaining a list is to find different meanings of FIRE, this issue does not matter; and indeed, if we search for collocates of FIRE in the BNC, as shown earlier, one of the significant collocates is *gun,* regardless of whether *gun* in the above instance was being counted. When the analyst's purpose is to find the lexical patterning of a particular word, for example, to identify all patterns of FIRE, the statistical measure of collocation is inadequate. In this case, the manual analysis of concordance lines is an alternative (CPA is one of these alternatives).

2.5.1.1 Collocation and CPA

As shown in the preceding section, one significant role of collocation is to identify different meanings of a particular word. Obtaining a list of collocates by using some statistical measure (e.g. log-likelihood) gives a list of words that can be interpreted by an analyst. In PDEV, some patterns' descriptions of particular verbs include collocates that are highlighted, which means that that particular collocate always accompanies that particular verb in that particular pattern. For example, the PDEV entry for the verb *ask* contains twelve patterns. Of these patterns, two different patterns stand out for including different collocates: *question*, *permission*. Each of these shows different meanings of the verb *ask* when collocating with *ask*. Some examples are shown below:

2.3 I asked a *question* about the story.

2.4 I asked *permission* to use her mobile phone.

The meanings of *ask* in these instances are different. In instance 2.3, a person puts a question to another person in order to find out about something. In instance 2.4, a person requests permission from another person to be allowed to do something.

Another language phenomenon that goes hand in hand with collocation, namely, colligation, will follow.

2.5.2   Colligation

According to Hoey (2005:43), colligation can be defined as follows:

    (a) " the grammatical company a word or word sequence keeps (or avoids keeping) either within its own group or at a higher rank;

    (b) the grammatical functions preferred or avoided by the group in which the word or word sequence participates."

An example of colligation that represents the second definition of in Hoey (2005) was proposed by Biber et al. (1999), who found that verbs such as *remember*, *reply* tend to be occur with the past tense (e.g. *I remembered that I was going in the library that day*; *I replied that this was not true*), while verbs such as *know*, *mean* are more frequent in the present tense (e.g. *I know what you want to say*; *What does this mean*?).

It is unclear whether Hoey's definition includes the co-occurrence of a word from a lexical word class, such as a noun or a verb, with a word from a grammatical word class, such as a preposition, or with another grammatical unit, such as a clause type. As much of the data

analysed in this thesis include such co-occurrences, it is important to clarify that these

phenomena are included here under the heading of 'colligation.' The work by Francis et al.

(1996) provided appropriate examples of colligation based on this definition by using the

term, 'pattern,' to refer to the colligation. For example, Francis et al. identified

verb+preposition patterns or colligations, such as **V *about* n, V *on* n,** etc. They also identified

verb+clause patterns or colligations, such as **V that, V to-inf** and **V wh**. The following are

some examples of the colligations of *decide*:

2.5 *I decided that things had to change.*

2.6 *He decided to attend his friend's birthday party.*

2.7 *I had flu and couldn't decide whether to see a doctor or not.*

Having introduced the notion of colligation, what follows is a description of how CPA/PDEV

involves colligation.

2.5.2.1 Colligation and CPA

CPA/PDEV uses colligation to uncover the meanings and patterns of the 'node' (i.e. the

verb). The uses of this phenomenon is apparent from Hanks's (2004:87) description of how

colligation plays a role in identifying verb patterns:

> Verb patterns consist not only of the basic 'argument structure' or 'valency
>
> structure' of each verb (typically with semantic values stated for each of the
>
> elements), but also of subvalency features, where relevant, such as the presence or
>
> absence of a determiner in noun phrases constituting a direct object. For example,
>
> the meaning of *take place* is entirely different from the meaning of *take his place*.
>
> The possessive determiner makes all the difference to the meaning.

In this quotation, Hanks indicates that although 'valency structure' is important in identifying patterns of a particular verb, 'subvalency' is equally important. For example, the meaning in the sentence *The meeting took place* can be identified; this is because this pattern consists of two straight arguments: *meeting* and *place*; the subject is an event, and the object is a location. However, sometimes such arguments are not sufficient to identify the meanings of verb patterns. This is because some patterns include a lower level of description. For example, the sentence *Sam took his place* involves subvalency, that is, the possessive *his*. It is not enough to look at the two valency structures *Sam* and *place*; the possessive pronoun *his* is the item that distinguishes the meaning of this pattern from that of the other one.

If we look at these examples from the perspective of colligation, we observe two separate colligation patterns. On the one hand, *took* colligates with a noun phrase consisting of a noun only (i.e. *took place*). In the other example, *took* colligates with a noun phrase consisting of a noun and a possessive determiner (i.e. *took his place*). Although the lexical collocation in each case (i.e. *took...place*) is the same, the grammatical information – the colligation – differentiates the meaning. Colligation in turn interacts with the subject-verb collocation, that is, between *meeting* and *took place* and between *Sam* and *took his place*.

Another piece of evidence for the importance of the interaction of collocation and colligation in the PDEV is the coding of the verb, AGREE. The PDEV proposes 10 patterns for the verb, AGREE. Among these 10 patterns, there are eight colligations, that is, eight kinds of grammatical units that co-occur with AGREE. These colligations are AGREE noun, AGREE that-clause, AGREE *to*-infinitive, AGREE quote, AGREE *about* noun, AGREE *on/upon* noun, AGREE *to* noun and AGREE *with* noun. In addition, one pattern combines two of these colligations: AGREE noun *with* noun (e.g. *agree terms with someone*). There are two patterns

that share the colligation, AGREE *with* noun, and these are distinguished by collocation. In one case, the subject of AGREE is a person (e.g. *Sam agreed with Jo* or *The teachers agreed with the suggestion*). In the other, the subject is an abstract entity (e.g. *These criteria agree with the others*).

2.5.3 Semantic preference

Sinclair (2004:142) defines semantic preference as "the restriction of regular co-occurrence to items which share a semantic feature …." For example, a lexical set such as *guitar*, *drum*, and *trumpet* can be grouped into the semantic feature 'Musical Instrument'. Stubbs (2001:65) redefines the definition of semantic preference as "the relation, not between individual words, but between a lemma or word-form and a set of semantically related words." He also proposes that "[m]etalinguistic labels for semantic features have to be used … such as 'animate', 'inanimate'; 'human', 'non-human …'' (Stubbs 2001:210). He observes that the adjective *large* frequently occurs with words that belong to certain semantic fields such as 'quantities' and 'sizes' (e.g. large number, large amount, etc.).

 In another study, Partington (1998) investigates the semantic preferences of the word *sheer*. The analysis shows that *sheer* often co-occurs with different lexical sets, some of which include '*volume*', '*force*', '*strong emotion*' (ibid.). In the same vein, Hoey (2005) proposes a similar term for semantic preference: 'semantic association'. He illustrates this by his typical example *thirty hour ride* (ibid.:16). He stresses that it is not the words *thirty*, *hour* and *ride* that co-occur all together; rather, it is the abstract semantic units '*number*', '*time*' and '*journey*' that are "primed" all together.

2.5.3.1 Semantic preference and CPA

Hanks (2003, 2013) proposes a similar term to semantic preference: semantic type. He states that this feature is used to identify patterns of verb lemmas (ibid.). This term is argued to be similar to the above-mentioned terms (i.e. semantic preference and semantic association) in that it groups words that share certain semantic features into abstract labels. He insightfully explains how semantic types are identified in relation to the verb by using CPA/PDEV:

> A group of words … is said to constitute a lexical set in relation to the verb…. The lexical set is united by a common semantic type. A lexical set of this kind is given a name—the name of the unifying semantic type—which is conventionally written in double square brackets with initial capital letters…. (Hanks 2013:12–13)

For instance, the CPA/PDEV shows that the lexical items *gun*, *rifle*, *pistol*, *shot*, *shotgun* often occur in the object slot of the verb lemma FIRE, which denotes the semantic type 'Firearm'.

It is apparent that Hanks' proposal of semantic types chimes with those of 'semantic preference' and 'semantic association' in that semantic types occur in verb patterns. However, what distinguishes Hanks's proposal from Sinclair's and Hoey's is that Hanks (2003; 2013) uses a hierarchical ontology (see section 2.3.5 for the description of the PDEV ontology) which contains a sequence of semantic types.

The following table illustrates how semantic preference can be used to distinguish different meanings/patterns recorded in PDEV:

Table 2.1 Semantic types and ontology

Verbs	Total number of patterns	Pattern number and patterns
FIRE	16	1. Human FIRE **Firearm** Direction 2. **Firearm** FIRE Direction
EMPTY	10	6. Human EMPTY **Firearm** Direction
BLAZE	7	5. **Firearm** BLAZE
AIM	7	1. Human 1 AIM F**irearm** *at* Human 2 or *at* Physical Object

In Table 2.1, column 1 contains different verbs; column 2 indicates the total number of patterns that each of the PDEV verbs has; and column 3 provides the pattern number where Firearm occurs. We can see in the table the occurrences of semantic type Firearm across patterns. Also, the semantic type Firearm occurs in both the subject and object slots of the verb patterns. Table 2.1, for example, shows that the verb FIRE has two patterns: in pattern 1 the object slot is filled with Firearm, while in pattern 2 Firearm occurs in the subject slot. The following are examples that illustrate this difference:

2.8 *The police fired a **gun** at the attacker.*

2.9 *The **bullet** fired into the door.*

Furthermore, the semantic types can be used to identify different meanings of verb patterns (see section 2.3.4 for explanation).

2.6 Similarities and differences between CPA/PDEV and related concepts

Having introduced CPA/PDEV and related it to the major language phenomena (e.g. collocation, colligation, semantic preference), I will discuss some concepts/methods that employ such units to study the meaning of words in the context, attempting to show similarities and differences between them and Hanks' recent method, that is, CPA/PDEV. There are some other concepts that are similar to that of CPA/PDEV. All account for the phenomena of recurring/co-occurring words. These concepts include the following:

- Units of meaning

- Lexical priming

- Constructions

2.6.1 Units of meaning and CPA/PDEV

'Units of meaning'[12], as the name suggests, is a model of meaning description, developed by Sinclair (1996, 2004). He later modified the name to 'meaning shift unit' (MSU) (Sinclair 2007), although most researchers continue to use the original term. The model of units of meaning includes four elements, discussed earlier, by which the meaning of words in the text can be described (Sinclair 2004). These elements are: a) collocation; b) colligation; and c) semantic preference.

A common example of Sinclair's concept is his famous discussion of the phrase *naked eye*. His analysis of concordance lines of *naked eye* shows that there are interactions between the four language phenomena *collocation*, *colligation*, *semantic preference*, and *semantic prosody* by which meaning of the combination is identified. In the following table (Table 2.2), I have selected some of the concordance lines presented in Sinclair (2004) and sorted them

---

[12] Defined as "extended lexico-semantic units" by Stubbs (2001:64).

into groups and classes, which show how the four language phenomena interact with each other to form a meaning.

Table 2.2 Units of meaning of *eye*

	semantic preference	**colligation**	**collocation**			
		prepositions	collocate	collocate	node	
Something is/is not	*seen* *obvious* *visible*	*with* *by* *to*	*the*	*naked*	*eye*	etc.

This table shows that the node word *eye* co-occurs with *naked*, which also collocates with *the* when they both occur with *eye*. Further, this sequence of words is also found to be colligated with some prepositions, such as *with*, *by*, *to*. However, this sequence of words is not enough to identify meaning. The lens of another meaning phenomenon in associated words is therefore needed, that is, semantic preference. This string of words *with/by/to the naked eye* is found to be followed by a group of semantic words such as *seen*, *visible*, *obvious*. This group fall into the semantic category of 'Visibility'.

 Since the underlying work of CPA involves identifying these elements, as discussed in the preceding sections, and since such elements also underlie Sinclair's work, it can be concluded that Hanks's new method of CPA/PDEV is best viewed as a development and a manifestation of Sinclair's theory of units of meaning; that is, CPA seeks to identify all canonical patterns of a given verb, taking into account collocation, colligation, and semantic preference. The product of such a method is presented in PDEV.

2.6.2 Lexical priming

Hoey (2003, 2004, 2005; Hoey et al. 2007) develops a theory called 'lexical priming'. In Hoey's conceptualization, lexical priming means that "every word is primed for use in discourse [by the individual] as a result of the cumulative effects of an individual's encounters with the word" (2005:13). In other words, the priming of words, which depends on their co-occurrence, is formed in an individual's mind as a result of the frequent exposure to such words in a given discourse, by which the individual becomes accustomed to the particular usages. In one study, for example, Hoey (2009:40) attempts to answer the question: "[i]s it possible for a child to construct a number grammar for himself or herself on the basis of typical input from what s/he encounters from being read to, etc.?" He compares the occurrence of *little* with numbers (e.g. *three little* + noun) in children and adults corpora. The analysis does not reveal that children are 'primed' to use *little* with numbers more than adults do, but does show that children encounter *little* with numbers more than adults do.

In one aspect, Hoey' theory is similar to Sinclair's 'units of meaning' proposal and Hanks's CPA/PDEV: namely, that the meaning of words can be realised by language phenomena such as collocation, colligation, and semantic association. Hoey also adds three more unit types: 'textual collocations', 'textual semantic associations', and 'textual colligations' (ibid.). However, since these three unit types are irrelevant to the current argument, they will not be dealt with any further.

CPA/PDEV seems to be similar to lexical priming in that they both employ these units of meaning, particularly collocation, colligation, and semantic preference/association, in their method of looking for meanings. What makes Hoey's hypothesis slightly different from both

Sinclair's and Hanks's, however, is that Hoey's proposal is essentially a psycholinguistic one, whereas Sinclair's and Hanks's proposals are not.

2.6.3 Constructions and CPA/PDEV

2.6.3.1 The form–meaning pairing

Goldberg (2003:219) defines linguistic constructions as "stored pairings of form and function, including morphemes, words, idioms, partially lexically filled and fully general linguistic patterns." In other words, items such as 'morphemes' (sub-word meaningful units, e.g. the comparative *-er*), 'words' (e.g. orange), 'idioms' (e.g. kick the bucket), 'partially lexically filled linguistic patterns' (e.g. verb–argument constructions) and 'fully general linguistic patterns' (e.g. negative *not*) (ibid.) are all treated as constructions that are stored in the mind and each have their own function or meaning. The major area where PDEV and construction grammar, particularly under this definition, show similarities and differences is 'partially lexically filled linguistic patterns', as exemplified for instance by 'verb–argument constructions'. Goldberg (2006) argues that verb–argument constructions have meanings in themselves regardless of the words that occur in these constructions. For example, although the verb '*sneeze*' is typically intransitive, it works in the caused-motion construction '*Sam sneezed the napkin off the table*' (Ellis et al. 2016:32).

PDEV and construction grammar have some other similarities and differences. First, like construction grammar with its focus on verb–argument constructions, PDEV seeks to identify meanings of verbs taking both lexis and grammar into account (see section 2.3) and treating them together rather than separately. However, unlike construction grammar, PDEV does not seek generalisation of particular constructions; it does not seek to identify all verbs that share a particular construction. Rather, PDEV seeks to identify all typical patterns of a verb as

45

being distinctive patterns (i.e. each pattern is different from each other pattern; see section 2.3.4). Second, construction grammar theory is based on the view that constructions are constituted as top-down structures, whereas the PDEV concept is bottom–up in nature: PDEV outcomes are based on corpus observation, and identify verb uses based on what is being observed in texts. However, it can be argued that what is observed in the texts (e.g. patterns) should already be seen as constructions, because ultimately the texts are produced by humans (writing using the linguistic knowledge stored in their minds).

I would argue, therefore, that the concept of CPA/PDEV is more detailed than that of construction grammar, as, first, CPA is rooted in a corpus-based approach in that it studies what is observed in the corpus in a collective way, and takes frequency of occurrence into account. Second, CPA/PDEV investigates every verb succinctly, in that all prototypical patterns of every individual verb are identified and every pattern is given an implicature.

2.7   Conclusion

In this chapter, I have introduced the concept of CPA/PDEV, which is the core concept underlying the work undertaken in this thesis. This was preceded by a brief history of how the notion of linguistic pattern has been defined by different scholars. I have also shown how PDEV is more adequate than other approaches such as pattern grammar. In addition, some dictionaries have been contrasted with PDEV.

The chapter has also attempted to relate CPA/PDEV to other relevant concepts in describing language. One of the key points this chapter has shown is that Hanks's method of CPA/PDEV is a development and a manifestation of Sinclair's theory of units of meaning, that is, CPA seeks to identify all canonical patterns of a given verb taking into account collocation,

colligation, and semantic preference. This conclusion was not explicitly made by Hanks. Also, the chapter has discussed how other ways of describing language such as lexical priming and construction grammar are similar or different from CPA/PDEV. Furthermore, this chapter has demonstrated how the work of Hanks on CPA/PDEV involves language phenomena such as collocation, colligation, and semantic preference. It becomes apparent that CPA/PDEV is a distinct concept of describing language that can be used to complement other concepts and approaches that treat lexis and grammar as inseparable components through which meaning can be identified.

Chapter 3   Learner Corpus Research and the Study of Learner Error

3.1   Introduction

The previous chapter introduced the approach to be employed in this thesis: corpus pattern analysis (CPA), via the Pattern Dictionary of English Verbs (PDEV). It also discussed some other concepts that are related to the CPA/PDEV approach.

In this chapter, I review two other concepts closely related to the core topic of the thesis: learner corpus research and the study of learner error. As noted in Chapter 1, this thesis means to provide a contribution to the field of learner corpus research. It also focuses on the notion of divergence, which is identified and quantified; previous work on grammatical divergence has tended to be placed within a subdiscipline of error analysis (see section 3.2). Originally, error analysis was carried out manually on small amounts of text; current approaches to learner language, however, are more likely to involve large corpora. The chapter reviews some of the very large quantity of research in this area (see section 3.3.2).

The chapter begins by introducing the notion of error analysis in more detail (section 3.2), and by discussing the procedure of EA and weaknesses of EA studies. The new era of describing learner language will then be explored in a section on learner corpora (section 3.3). Two prominent learner corpora will be introduced (section 3.3.1). A review of some of the research studies whose method of investigation is based on contrastive interlanguage analysis (CIA) follows (sections 3.3.2 and 3.3.3). Section 3.4 will conclude this chapter.

3.2   The concept of error analysis

The features of learner language can be identified by various approaches (see Spolsky 1979 for an accounting of these). In the late sixties, in particular, research investigating learner

language was mostly done by the use of the error analysis approach (Ellis and Barkhuizen 2005).

Ellis and Barkhuizen (2005) note that the first attempt to introduce error analysis as a methodological way of analysing learner' errors was proposed in Corder's (1967) seminal article 'The Significance of Learner's Errors'. Apart from the generally accepted negative view of learners' errors (i.e. that errors are signs of poor learner competence), Corder (1967) observed that three types of value can be provided by roles of learner's errors. That is, Corder suggests, the role of learner's errors can be beneficial to teachers, researchers, and the learners themselves; to researchers, errors provide more insights into how learners learn the language; to learners, errors provide a tool that can be used to cover their weak points and spur their learning; and to teachers, errors give a guide to measure learners' progress towards proficiency in the target language (ibid.:25).

In explaining the process of how EA is conducted in research studies, Ellis (1994:68–69) summarises it in four steps, as follows:

> The first step in carrying out an EA was to collect a massive, specific, or incidental sample of learner language. The sample could consist of natural language use or be elicited either clinically or experimentally. The second stage involved identifying the errors in the sample. The third stage consisted of description.… The fourth stage involves an attempt to explain the errors psycholinguistically.

Although these steps can be useful for identifying learners' errors, EA is not without limitations, as suggested by Ellis and Barkhuizen (2005:70):

> EA suffers from a number of limitations. It offers an incomplete picture of learner language because it only examines what learners do wrongly and ignores what they do correctly. Also, EA cannot account for learners' avoidance of certain L2 forms.

Also, some other scholars, such as Schachter and Celce-Murcia (1977) dedicate articles to outlining a number of weaknesses that EA studies show. Some of these weaknesses are in (1) "the analysis of errors in isolation"; (2) "the identification of points of difficulty in the target language"; (3) "the biased nature of sampling procedures"; and (4) "statements of error frequency" (ibid.). Each of these weaknesses will be dealt with in turn below.

3.2.1 The analysis of errors in isolation and statements of error frequency

As Schachter and Celce-Murcia (1977:444–445) report, the problem with this step is that the analyst only extracts errors from the corpus, and excludes looking at correct usages. These authors proceed to give an example of how considering both errors and non-errors can be of value to know more about a phenomenon under investigation. An example of this practice, as stated in their paper, is in a study by Andersen (1977). This study has been summarised by Schachter and Celce-Murcia (1977), who noted that the basic observations about the relative frequency of errors associated with definite and indefinite articles are "not very interesting facts" (ibid.: 445). On the other hand, they pointed out that Andersen's comparison of all instances of article use and the equivalent expressions in Spanish is more interesting. Andersen's conclusion that "many of the subjects were using the strategy of providing the English equivalent of the article that was required in Spanish in that context" (as cited in

Schachter and Celce-Murcia 1977: 445) could not be drawn without considering the correct instances of article use as well as the erroneous ones.

Another area of weakness is statements of error frequency. EA studies did not attempt to take frequency of errors into account. In particular, they did not compare correct and incorrect uses of the same word, or quantify errors concerning correctness. Taking both correct and incorrect uses of words into consideration, investigators can draw sounder conclusions about or a wider picture of learners' competence (see Abdulmajeed 2016 for details on the measurement of correctness).

3.2.2 The proper classification of identified errors

Another weakness of EA is the difficulty of classifying learners' errors by type and/or deriving types, as pointed out by Schachter and Celce-Murcia (1977:445). Schachter and Celce-Murcia argue that it might not always be straightforward to tell whether an identified error is deviant from the norm, and if so, what type of structure a particular error belongs to. They provide an example of an error made by a Japanese learner: *Americans are easy to get guns*. They propose two interpretations: 1) the student might have written the sentence by analogy with *Americans are easy to please*; 2) or the student might have treated the adjective *easy* as usable in the same way as *able* as in *Americans are able to get guns*.

3.2.3 The identification of points of difficulty in the target language

As a matter of concern regarding the identification of points of difficulty in the target language, Schachter and Celce-Murcia (1977:447) suggested that although EA can successfully reveal the degree of difficulty of use of a given linguistic category by showing the relation between that category and the number of occurrences of errors in it, it fails to

provide evidence for the phenomenon of avoidance. Particularly, they referred to the likelihood that 'learners avoid producing constructions which they find difficult both in terms of the actual formation of such structures and the conditions for their use' (Schachter and Celce-Murcia 1977: 447). They continued by commenting that a study by Kleinmann (1977) reported that "Arabic-speaking learners avoid the passive construction" (Schachter and Celce-Murcia 1977:447), and that EA would not be able to identify such avoidance. The authors stress that knowing what structure the learner English will avoid and why (i.e., which EA succeed to do) is as important as knowing what structure he/she will not avoid and why (i.e., which EA fails to do).

Since one of the aims of conducting EA studies is to help learners avoid errors, and this aim can be achieved by addressing errors in a syllabus, the phenomenon of avoidance tends to be ignored by EA studies, and thus such errors cannot be addressed in the syllabus.

3.2.4   The biased nature of sampling procedures

Another weakness of conducting EA is the tendency for researchers to influence sampling. Schachter and Celce-Murcia (1977:449) criticised the EA investigators for their "tendency to overlook the fact that they may be working with a very limited and biased sampling" (ibid.: 449). In addition, they point out that EA data tend not to be 'random' in some studies. The authors suggest that since sampling tends to be influenced by different researchers' perspectives (i.e. no criteria set to sampling procedures), it is not only that the interpretation of learners' errors may be problematic, but also that the statistical conclusions around phenomena may not be significant. To draw a solid conclusion which leads to fruitful implications, investigators should conduct sampling according to criteria set by them.

In addition to these weaknesses, EA studies were based on small quantities of language, especially on the language of the individual. The emergence of the new learner corpora approach and contrastive interlanguage analysis, which will be introduced in the following section, provides a new way of investigating learner language, by which such drawbacks can be avoided.

3.3   Learner corpus research

This section will introduce the field of research into computer learner corpora (CLCs) and will review the previous work that has been done in learner corpora. First, this section will discuss the development of language learner studies, and the role played by learner corpora research; then, it will review a number of studies carried out in the field of learner language research. Finally, it picks up the most recent studies that are comparable to the present study, to establish a rationale for conducting the present study.

At this point, it should be pointed out that since numerous studies have been done on learner corpora, it is beyond the scope of this thesis to review all of them in detail. Therefore, I will briefly review some studies that I believe are most worth mentioning in this section, and which are most directly relevant to the present study.

The features of learner language can be identified by various approaches (cf. Spolsky 1979 for these approaches). In the sixties, research on investigating learner language was mostly done by the use of the error analysis approach (Spolsky 1979). In addition to the weaknesses of EA, discussed earlier in this chapter, EA studies, in general, tended to be based on small quantities of language and to focus on the language of the individual, which put EA under criticism, as mentioned above. In the late eighties, a new field of learner language research

emerged, that is, learner corpora research (LCR), which commenced as a result of opportunities offered by computing power and the development of computerized corpora (see Pravec 2002 for detailed information about various corpora). One of the prominent learner corpora is the International Corpus of Learner English (ICLE). Granger (2002:7) described ICLE in the following way:

> Computer learner corpora are electronic collections of authentic FL/SL textual data assembled according to explicit design criteria for a particular SLA/FLT purpose. They are encoded in a standardised and homogeneous way and documented as to their origin and provenance.

In this quotation, Granger describes ICLE rather than define the notion of learner corpora. This description shows that learner corpora are not just a collection of texts; instead, they are collections of 'continuous texts' (i.e. not just selected sentences); authentic (i.e. real communications); collected for purpose (e.g. studying particular linguistic phenomena); annotated (e.g. POS-tagged or error-tagged); and collected according to particular criteria by compilers (e.g. age, level, timed or untimed, etc.).

Some of the elements mentioned above have on occasion been controversial. For instance, there is a dispute on the authenticity of learner corpora, that is, whether the texts collected are in fact the outcomes of genuine communications. This is because most of the learners' written texts are not written for the purpose of the corpus; they are written for some other purposes such as exams. It might be argued, though, that building any corpus should have a purpose, which means that the authentic data are not so. Another area of dispute is the proficiency level of learners contributing to a corpus. Some compilers assume that all learners in a given corpus

are at the right level of proficiency (e.g. beginners, intermediate, advanced). In Granger et al.

(2009), for instance, the proficiency level of ICLE learners is reported to be advanced.

However, when the same authors rated a randomly selected sample, they found that 19 out of

the 20 essays was were rated only as high intermediate. Such controversial issues are,

nevertheless, not unproblematic.

3.3.1   Learner corpus design

3.3.1.1 ICLE

One prominent learner corpus that has been researched extensively is the International Corpus

of Learner English (ICLE). The ICLE project was created by the Centre for English Corpus

Linguistics at the Université Catholique de Louvain, established by Sylviane Granger in early

1990s. The corpus contains argumentative essays produced in English by university students

from sixteen mother-tongue backgrounds. The corpus is both POS tagged and error tagged. It

was compiled according to a set of criteria for a specific purpose: as pointed out by Granger

(1998c), the goals of assembling the ICLE were to investigate "the interlanguage of the

foreign language learner" and "to provide an empirical resource for large-scale comparative

studies of the interlanguage of advanced EFL learners with significantly different native

language backgrounds" (Pravec 2002:83). The thread of investigation that involves ICLE is

the comparison between learners' written-linguistic features in L2 English and those of

native-speaker writers. The method that has been used for comparison is called contrastive

interlanguage analysis (CIA), which will be discussed in section 3.3.2.

3.3.1.2 The Cambridge Learner Corpus

The Cambridge Learner Corpus is an on-going collection of candidates' examination essays

compiled by the Cambridge University Press. The corpus contains over 55 million words. The

contributors to this corpus are students from all around the world, from around 143 mother-tongue backgrounds (O'Keeffe and Mark 2017). The corpus is error tagged, and every exam paper is annotated with details about the candidates, such as age, sex, nationality, and proficiency level (ibid.). The corpus can be used by researchers to describe various aspects of learner language and thus contribute to the development of new ELT materials. Some of the most recent work on the Cambridge Learner Corpus will be shown later (in section 3.3.1.2).

3.3.2 Research based on ICLE

Granger (2015:1) points out that studies investigating learner corpora have increasingly used contrastive interlanguage analysis (CIA) as a method of investigation. The core purpose of CIA is "to uncover factors of 'foreign-soundness' in learner writing" (Granger 1996:43); it "involves quantitative and qualitative comparisons between native language and learner language (L1 vs. L2) and between different varieties of interlanguage (L2 vs. L2)" (Granger 2009:18). The assumption underpinning the ICLE work is that the features of learner language will be apparent when learner output is compared with native-speaker norms. A common quantitative output is obtained by measuring the 'under-use' or 'over-use' of a feature in the L2 corpus compared with an L1 corpus. 'Under-use' is said to occur when a feature is significantly less frequent in the L2 corpus than the L1 corpus, and 'over-use' is said to occur when the feature is more frequent in L2 than in L1. Examples of each will be given below. Researchers can also identify whether features of learner interlanguage are shared by different L1 backgrounds (i.e. whether the features are caused by L1 interference or by other factors such as cross-linguistic effects).

A major advance in the study of learner language has been made by the advent of quantitative research in the field, in a move away from the use of small-scale studies as in the 1960s. With

ICLE, at last, it was possible to study learner language on a large scale and to make reliable comparisons between learners from different backgrounds—between NS and NNS as well as between different NNS groups. As pointed out by Granger (1996:43), NS/NNS comparisons can highlight a range of features of 'foreign-soundness' in learner writing and speech, such as under- and overuse of words, structures or phrases. As De Cock and Granger (2005:7) point out, "learners' overuse and underuse of words or phrases is brought to light by contrasting the frequency counts of words or phrases in a learner corpus and a comparable native-speaker corpus." These authors then identify areas in which learner language differs from NS language concerning frequency of distribution.

There are a number of studies that have been conducted investigating the several linguistic features of learner language identified in this way. The foundational scholars of ICLE have investigated ICLE data in diverse ways, such as the grammar, lexis, and discourse features of L1 learners. For example, Ringbom (1998:43) reported that learners of different L1 backgrounds overuse high-frequency verbs such as *think* and *get* in comparison to a native-speaker group (see Neff et al., 2003; Ellis et al., 2016 for a similar conclusion).

However, such findings do not necessarily show that learners producing these general high-frequent verbs are using them competently. Some studies with a more qualitative approach appear to reveal that they are not. For example, Altenberg and Granger (2001) investigate the use of English *make* by two learner groups (French and Swedish) in comparison with NSs. One of their key findings was that the learners not only "underuse delexical structures of *make*, but they also misuse them" [e.g. they produce the incorrect usage *make a step* instead of the correct usage *take a step*] (ibid.:178-179). Another key finding was that Swedish learners overuse the causative construction *make* + adjective in comparison with French

learners. The authors relate this phenomenon to a commonality between the Swedish learners'
language and English; that is, the structure *make* + adjective has an equivalent structure in the
Swedish language, whereas this is not the case with the French language (see Ellis et al.,
2016, for a similar conclusion, as well as Chapters 5 and 8 in this thesis). A similar study was
conducted by Gilquin (2010) in an attempt to find an answer to whether learners overuse or
underuse causative constructions as compared to native speakers. The author compared the
frequencies of ten causative constructions identified in ICLE learners and in a particular
version of BNC (BNC-10W). One of her striking findings was that the construction *make* +
noun + infinitive (without *to*) (e.g. *This makes me feel bad*) was heavily used by the ICLE
learners (403.93 per million words), in comparison with the native writers (50.77 per million
words).

In a more recent development of learner corpus research from a methodological perspective, a
new technique proposed by Stefanowitsch and Gries (2003) and Gries (2004), that is,
collostructional analysis, has been applied to learner corpora. This notion focuses on the
association between constructions and words in these constructions (e.g. the tendency
of *remember* to occur in imperative constructions (ibid.). Examining native and non-native
selection in verb constructions using collostructional analysis, Gilquin (2012) analysed how
ICLE learners, relative to native speakers (BNCw), use particular verbs in association with
causative constructions.[13] Particularly, she focused on two causative constructions: [X *MAKE*
Y V_{inf}] and [X *MAKE* Y V_{pp}]. The data used for comparison were the ICLE and BNCw.
Gilquin discovered that "learners' causative constructions are often unidiomatic in that they

---

[13] "The analysis was performed by means of *Coll.analysis 3* (Gries 2004) which, on the
basis of a list of the verbs occurring in the construction(s) under investigation, computes the observed
and expected frequencies and, through a binomial test, the (log-transformed) probability of a particular
observed frequency given the expected frequency. The "distinctiveness value" resulting from this
calculation indicates which verbs are significantly associated with a given construction, that is, are
distinctive collexemes of this construction" (Gilquin, 2012:6).

contain verbs which are unlikely to occur in native English with that particular construction" (Gilquin 2012: 1). In the [X $MAKE$ Y V_{inf}] construction, for example, native speakers use relational verbs, such as *seem* and *appear*, in the non-finite slot, while learners use relational verbs that are unlikely to be used by native speakers in a similar context (e.g. *become* as in *make their dreams become true* [Gilquin 2012: 14]). In addition, the analysis of [X *MAKE* Y V_{pp}] revealed that while "there is a greater variety of collexemes occurring in the non-finite verb slot of [X *MAKE* Y V_{pp}]" produced by learner English (consisting of approximately 16 verbs), "the construction [produced by native English] is highly restricted in terms of lexical preferences" (consisting of only four verbs: *known*, *felt*, *recognized* and *swallowed*) (Gilquin 2012: 15-16).

Regarding nouns, several research studies have reported that the over-use of 'vague' and 'general' nouns is one of the observed features of learner language. In their research on the automatic profiling of learner texts, Granger and Rayson (1998) compare the frequency of nouns identified in the ICLE French sub-corpus and a corpus of native speakers known as LOCNESS (see Chapter 4 for more details on LOCNESS). They found that the learners rely much more on general or all-purpose nouns (e.g. *thing*, *people*, *phenomenon*, *problem*) than native speakers do (ibid.:128). This phenomenon has also been observed in a study conducted by Cobb (2003). He found that advanced learners overuse general words such as *thing*, *problem*, *change*, and *strong* (ibid.:402).

The phenomenon of the use of general words by learners can be interpreted using Hasselgren's (1994:237) concept of the "teddy bear principle". This concept suggests that learners rely much more on "words they feel safe with" (ibid.:256). She drew this conclusion based on her study's findings that 'core words' or general words such as *very (much)*, *a lot*

(of), and *extreme(ly)* are overused by advanced Norwegian learners in comparison to native speakers (ibid.:255).

Recently, as a step to reappraise the field of learner corpus research, learner corpus research community published a state-of-the art account of learner corpus research. Granger et al. (2015) produced a seminal book namely *The Cambridge Handbook of Learner Corpus Research*, offering a comprehensive evaluation of the existing research by emphasising "some of the major strengths of the research conducted to date … [as well as the] shortcomings that need to be addressed and gaps that need to be filled" (ibid.: 5). One of the shortcomings of LCR that this handbook pointed out is that "there are still many studies which remain rather descriptive, focusing primarily on learner errors and contrastive analysis, often without the theoretical frameworks which would enable rigorous interpretation or explanation of the data" (ibid.: 330). The authors, therefore, called for more focused research that brings the relationship between SLA and LCR closer.

In an attempt to address this need, the International Journal of Learner Corpus Research (2018) has designated a special issue for the study of one of the hot topics of SLA, namely, *tense and aspect in second language acquisition and learner corpus research*, since "the acquisition of tense and aspect has been identified as one of the biggest obstacles for language learners striving for the emulation of target-like patterns" (Fuchs and Werner 2018: 145). The special issue includes various studies on this topic, which serve "to reduce the still somewhat marginalized status of corpus-based studies of learner production in studies of SLA and move it closer to the core of established methodological approaches" (ibid.: 146).

This movement within LCR towards SLA indicates that LCR has become an interdisciplinary field; it contributes, by cooperating with the SLA field, to the understanding of how other interlanguage features are acquired. In addition, it seems that LCR welcomes new proposed method to investigate learner interlanguage feature, as seen earlier in Gilquin's (2012) use of collostructional analysis (see Gilquin 2015 for contrastive collostructional analysis: causative constructions in English and French).

3.3.3   The English Grammar Profile (O'Keeffe and Mark 2017)

The English Grammar Profile is a sub-project of the English Profile, which intends to describe learners' grammar competence (O'Keeffe and Mark 2017:457). As pointed out by O'Keeffe and Mark (2017:404):

> In summary, through the EGP, it is aimed to arrive at a corpus-based description of what learners can do with grammar at each level of the CEFR based on what they have written in Cambridge exams. The present study, therefore, represents a large-scale attempt to look at learner language at all levels across the CEFR.

O'Keeffe and Mark (2017) aim to come up with a descriptive inventory of what particular grammar categories are used by learners at each of the six CEFR (Common European

Framework of Reference for Languages)[14] levels (O'Keeffe and Mark 2017:457). They established a six-step process to come up with the final product, a detailed inventory of these categories. This process involves steps assessing the following aspects: 'frequency of use' (i.e. there are sufficient occurrences of grammatical items to be worth investigating), 'rate of correct uses' (i.e. rates of correct uses at each level), 'range of users' (i.e. number of learners using the given grammatical item correctly), 'spread of first language families' (i.e. the extent to which correct uses of particular grammatical items are seen in learners from given L1 backgrounds) , 'spread of contexts of use' (i.e. the extent to which correct uses of a given form are found in different registers), and 'avoiding the effect of a task' (i.e. whether a correct usage of a given form is repeatedly used due to a task rubric) (see O'Keeffe and Mark 2017 for more details on how this process works). A number of grammatical categories are set to be investigated, such as adjective, adverbs, tenses, verbs, etc.

One of O'Keeffe and Mark's key findings is that there is a recognisable improvement between A2 and C1 in use of the pattern 'pronoun + linking verb (+ adverb) + adjective + (*that*)-clause', as shown in Table 3.1.

---

[14] "The Common European Framework of Reference for Languages (CEFR; Council of Europe, 2001) is intended to provide "a common basis for the elaboration of language syllabuses, curriculum guidelines, examinations, textbooks, etc. across Europe" (p. 1). It comprises (a) a descriptive scheme for analyzing what is involved in language use and language learning and (b) a definition of communicative proficiency at six levels arranged in three bands—A1 and A2 (Basic User), B1 and B2 (Independent User), C1 and C2 (Proficient User). The CEFR is not language-specific. It describes, for example, the communicative functions that learners should be able to perform at different proficiency levels, but does not specify how those functions might be realized in, say, French or German. It thus assumes that any communicative task requires a comparable level of proficiency from language to language." (Little 2007:para.1).

Table 3.1 Examples of 'pronoun + linking verb (+ adverb) + adjective + (that)-clause'

across levels

Level 'pronoun + linking verb (+ adverb) + adjective + (that)-clause'
A2 I am sure we will find something to do. (A2, Norwegian, 2003)
B2 It seems obvious that this oil comes from the gas station. (B2, French, 2008)
C1 It is highly unlikely that the goods can vanish from your warehouse without being noticed. (C1, Russian, 2008)

Through examples like these, O'Keeffe and Mark explain that the learners become gradually knowledgeable about the range of words that can be used in each slot, and how these words are patterned. Such a finding is useful in showing how forms and their uses are progressively used by learners across the six levels of CEFR.

A description of what learners can do (as opposed to what they cannot) is important because it reveals the status of their use of what is correct rather than what is incorrect. Besides, the goal of the EGP is to describe the development of correct uses of grammatical categories. This focus on 'correctness' is to be welcomed; ideally, though, it will still be complemented by a study of error. Combining the two will give a more complete indication of learner achievement and will be useful for teaching.

3.4   Conclusion

In this chapter, I have reviewed two concepts relevant to the topic of this thesis. First, I have introduced the concept of focusing on identifying learners' errors: the error analysis approach. The procedure of EA has been briefly presented, followed by some main weaknesses and criticism underpinning EA studies. Second, the new present era of learner language research

has been introduced: that of learner corpora. I have shown what a learner corpus involves, adopting Granger's (2002) description of learner corpora. I have also discussed areas where defining criteria for good learner corpora or those suited to a given purpose can be controversial. Next, I presented two prominent examples of learner corpora: ICLE and CLC. Each of these corpora has different criteria and purposes. After that, the method that underlies most ICLE studies was described: contrastive interlanguage analysis (CIA). A number of learner corpora studies were reviewed from a CIA perspective. Among the main principles that these ICLE studies employed is the principle of over and under-use. The outcomes of such studies showed that they were able to identify learner interlanguage features, by which pedagogical implications can be formulated. Finally, another, and probably the most recent, work on learner corpora attempting to describe learner language was presented: the English Grammar Profile (EGP). This work focuses more on describing 'what learners can do' with grammatical categories across different levels of proficiency, using the levels of the Common European Framework of Reference for Languages (CEFR).

This thesis attempts to introduce and test Hanks' innovative approach to describing language—corpus pattern analysis—and to test its outcomes using the Pattern Dictionary of English Verbs (PDEV). The next chapter will set out the study's methodological framework.

Chapter 4   **Methodology**

4.1   Introduction

This chapter will first introduce the data used in this thesis. Then it will provide justifications

for the corpora selected for the present study. The tool used to extract and analyse target verbs

will be presented, followed by the procedure for selection and extraction of target verbs. Next

a worked example to illustrate the method of identifying and interpreting patterns of the target

verbs will be provided. Finally, a pilot survey of patterns of the target verbs, as identified in

the target corpora, is presented.

4.2   Data

Three corpora were used for the present study:

- Two sub-corpora of the International Corpus of Learner English (ICLE): ICLE-
  Chinese and ICLE-Swedish

- The Louvain Corpus of Native English Essays (LOCNESS)

- The British National Corpus

4.2.1   The ICLE

The corpus study presented here draws on data from the International Corpus of Learner

English (Granger 1998), launched in 1990 by Sylviane Granger. ICLE contains written essays

(mostly argumentative essays) produced by EFL learners from 16 different mother-tongue

backgrounds: Bulgarian, Chinese, Czech, Dutch, Finnish, French, German, Italian, Japanese,

Norwegian, Polish, Russian, Spanish, Swedish, Turkish, and Tswana. The corpus includes

6085 essays totalling 3,753,030 words; each sub-corpus contains approximately 200,000

words, except for the Chinese sub-corpus, which contains 490,617 words. Some essays were

produced under exam conditions, with no access to reference tools such as grammars and

dictionaries, while others were written as homework, with access to reference tools (Granger

et al. 2009). The learners are university students of English (usually in their third or fourth year) who have learnt English as a foreign language in a classroom. Their proficiency level is reported to be higher intermediate to advanced. The ICLE essays have an average length of 617 words. The essays cover various topics, such as 'advantages and disadvantages of banning smoking in restaurants'; 'advantages and disadvantages of using credit cards'; 'value of theoretical university degrees'; 'poverty'; 'effect of technology on our imagination'; 'feminists'; and 'money'. The average age of the learners is 20 years old.

4.3   Justification for selection of ICLE and other normative corpora

According to Granger (1993, 1998), some very strict design criteria were set to collect the ICLE data. This was in order to achieve the main goals of compilation of ICLE: to uncover the interlanguage of EFL learners, and to study the 'foreign-soundingness' phenomena in EFL learners' essays, which is usually apparent by comparison with target native-speaker language. Furthermore, ICLE data have been extensively researched, and many studies doing so reported significant findings concerning linguistic features used by foreign language learners. Given that, I chose ICLE as the corpus for the present study.

It should be noted that there are 2 versions of the ICLE data: ICLEv1 and ICLEv2. The main differences between the two versions are that, unlike ICLEv1, ICLEv2 comes in a CD-ROM with a built-in concordance, and that data in ICLEv2 but not ICLEv1 are POS tagged. The present study used ICLEv1 for several reasons. First, ICLEv2 is not free and it is quite expensive to purchase. Second, since the present study follows the CPA/PDEV method of identifying verbs, in which every concordance line should be investigated manually, it was felt that the POS-tagged version was not helpful and that ICLEv1 was a more suitable option for the present study.

As mentioned earlier, the ICLE data includes 16 different mother-tongue backgrounds, each of which is presented in a sub-corpus. It was beyond the scope of this thesis to investigate all ICLE data, which would have been prohibitive in terms of time requirements. I therefore restricted the investigation to two sub-corpora: ICLE-Chinese and ICLE-Swedish.

ICLE-Chinese was selected because it has the highest number of word tokens (490,617 words) and the highest number of essays (982 essays) in relation to the other ICLE sub-corpora (each of which includes around 200,00 words).

ICLE-Swedish was selected because it was anticipated that Swedish learners would exhibit a higher level of proficiency than Chinese learners, both because English shares several similarities with Swedish and because English is obligatory in Swedish schools and it is introduced at an early learning stage (Altenberg 2009).
Each of these two sub-corpora will be presented below.

4.3.1   ICLE-Chinese

As mentioned, the Chinese sub-corpus comprises 982 essays with a total word count of 490,617. The essays come from students at two universities: Hong Kong Polytechnic University and the University of Portsmouth. All the texts are argumentative essays, with an average length of 500 words. The average age of the writers is 20 years. In all, 64% of the learners who contributed data to ICLE-Chinese are female. The proficiency level for all of them is reported to be advanced; according to Granger et al. (2009), however, a quick inspection of some essay samples revealed that the sub-corpus covers two proficiency levels: higher intermediate and advanced. This was established by rating a randomly selected sample

of 20 essays and rating them on the basis of CEFR category criteria (ibid.). Of the 20 essays, 19 were rated higher intermediate (B2).

4.3.2   ICLE-Swedish

The ICLE-Swedish sub-corpus contains 355 essays with a total of 200,033 words. Of the 355 essays, 302 (85%) are argumentative essays, while 53 are literary essays. The writers of the essays were students from three universities: the University of Lund, the University of Växjö and the University of Göteborg. The average length of the essays is 564 words, and the average age of the writers is 27; in all, 77% of them are female. Their proficiency level was reported to be advanced; the same procedure of essay level verification conducted with the Chinese sub-corpus above was followed, and the results showed that all 20 essay samples were rated advanced (Granger et al. 2009).

4.3.3   The Louvain Corpus of Native English Essays (LOCNESS)

LOCNESS was collected by the Louvain team in Belgium from 1991 to 1995 (Granger et al. 2009). LOCNESS comprises two sub-corpora, respectively consisting of argumentative and literary essays written by British and American native speakers with ages ranging from 16 to 23. The British texts contain 266 argumentative and literary essays with a total of 168,400 words, produced by A-level and university students; the American component comprises 196 argumentative and literary essays with a total of 155,904 words. The average length of the essays in both sub-corpora is 500 words. The components of the two corpora are free essays, timed essays, and examination papers. Reference tools were used in some free essays and timed essays, whereas no reference tools were used in the examination papers. The topics of the A-level essays include boxing, transport, the parliamentary system, and others. The British university students write about topics such as the French intellectual tradition, French

higher education, and others. The American argumentative essays cover a wide range of topics, such as abortion, feminism, recycling, the drinking age, gender equality, ethics, suicide, and others.

LOCNESS was used as a native-speaker comparison corpus in the present study for two reasons: first, to see how similar/different the uses of target verbs produced by the learners are from the comparable target language norm; second, to see if the native-speaker students produce deviant patterns of target verbs. In addition, since the major advantage of LOCNESS in relation to ICLE is the comparability of text types (Granger et al. 2009), in that both corpora contain argumentative essays, comparative investigation should yield reliable results: "[t]he reason [LOCNESS] has been used so extensively is that it contains argumentative essays and is therefore arguably a more reliable basis for quantitative comparisons with L2 corpora like the ICLE than more general corpora" (Granger 2015:17) (cf. Ringbom 1998, Lorenz 1998, Virtanen 1988, Aarts and Granger 1998, Aijmer 2002, Lin 2002, among others, for comparative studies).

A number of researchers assert the need for a comparable targeted corpus when it comes to comparison of learners with native-speaker writers. For instance, Leech (2001:333) suggests that "we also need *targeted* corpora – corpora targeted to represent as closely as possible the learner's future communicative needs". A similar view is stated by Kaszubski (1998:25) (cited in Chen 2013:419): "it may be psycholinguistically more appropriate to compare EFL learner corpora not with ideal 'expert performances' in the target language but (more realistically) with the attainable performances of native learners of a comparable, preferably slightly lower, age and experience."

Table 4.1 Comparison between the target corpora (ICLE and LOCNESS)

Parameter	ICLE		LOCNESS	Comparability*
	Chinese	Swedish	Argumentative and literary essays	+++
Essay type	Argumentative essays	Argumentative essays and literary essays		
Number of essays	982	355	322	++
Size	490,617 words	200,033 words	324,304 words	+
Length of each essay (tokens)	500	564	=/+ 500	+++
Average age of students	20	27	17–22	++
Level of study	University level (Third to fourth year)	University level (Third to fourth year)	University level (Third to fourth year) and A-levels	+++
Topics	Banning smoking in restaurants; credit cards; abortion; money; university degrees; television; feminists; technology, among many others	Immigrants; banning smoking in restaurants; credit cards; abortion; money; university degrees; television; feminists; technology	A single Europe: a loss of sovereignty for Britain, aspects of studying ethnic American literature, aspects of social psychology, water pollution, nuclear power, gender roles, violence, sex, drugs, Parliament, freedom of religion, the French intellectual tradition	++
Compilers	Professionals in computer learner corpora		Professionals in computer learner corpora	+++
Time of completion	1998	1988	1998	+++

*+ = low; ++ = between average and high; +++ = high

As seen in Table 4.1, there are aspects where comparability between the ICLE sub-corpora data and the LOCNESS are high, between average and high, or low. High comparability (+++) was found for essay types (e.g. argumentative), length of each essay, level of study, compilers, and time of completion. In addition, there are aspects where the target corpora show average comparability (++), such as number of essays, average age of students, and

topics. On the other hand, there are aspects where comparability between the three corpora is low (+), such as size of corpus. It can be concluded that the corpora are sufficiently comparable for me to base a thesis on comparing them.

4.3.4   The British National Corpus (BNC)

The British National Corpus (BNC) was built in the period between 1991 and 1994. The BNC represents a wide range of sources of British English from the 20th century. It comprises 100 million words. It contains two different components: spoken and written texts produced by native speakers. The written part of the BNC (90%) includes various types of texts, such as academic books, school and university essays, extracts from newspapers, specialist periodicals and journals, fiction, published and unpublished letters and memoranda, etc. The spoken part (10%) contains transcriptions of informal conversations from different contexts, such as formal government meetings, business meetings, radio shows, etc. The corpus is POS tagged using CLAWS software. BNC is synchronic, which means that no new components were added by the time of completion of the corpus.

The BNC was used for the present study for several reasons. First, since CPA/PDEV is based on the BNC, I chose to be consistent by using the same corpus. Second, the BNC was used to search for divergent patterns of target verbs identified in the target corpora of learners (ICLE-Chinese and ICLE-Swedish) or native speakers (LOCNESS).

At this point, it should be acknowledged that the BNC is not in any way comparable with either ICLE or LOCNESS, for two reasons. First, the BNC has many genres/registers, while the genre in ICLE and LOCNESS is mainly academic argumentative essays. Second, the BNC does not actually represent current English. These two points of divergence may appear

to affect the comparison in this study between the ICLE/LOCNESS data and the BNC. However, since the present thesis focuses on the meanings and patterns of verbs, it is unlikely that this incomparability would have a major effect on the study, because language does not change so quickly. Another reason is that BNC was used in this thesis as a check when I was unsure of the acceptability of a given pattern in British English. (That is, I was not *comparing* its data as such with those of the other two corpora, so issues of comparability are less important). In contrast, ICLE and LOCNESS are being directly compared, so comparability between them is important.

4.4   The tools

To analyse the target corpora, AntConc Tools version 3.4.4 (Anthony 2005) was used. This software package has several features for lexical analysis, three of which are especially useful for the present study: 'Concordance', 'Concordance Hits', and 'File View'.

The *Concordance* tool shows the items selected for detailed analysis, in a KWIC (key word in context) format (see Figure 4.1). The selected item, which appears in the centre of each concordance line, is known as the node word. Such a format enables the linguist to draw conclusions about the usage of the target words in question. Hunston (2002:38ff), for instance, provides several types of observations that can be made using concordance lines, one of which is *observing meaning and patterns* (i.e. the association between the meaning of a word and the surrounding words).

Figure 4.1 Screenshot of the AntConc Concordance tool

Concordance Hits is another useful tool for the purposes of this study. It shows the raw frequency of concordance lines for the target word under study. This enables the linguist to decide whether the frequency of a given word is sufficient for investigation. In addition, this information also allows the linguist to compare frequency across corpora.

The third concordance tool, *File View*, is equally important. *File View* allows the linguist to see where the target word under study comes from, as further information about the target word may be needed (see Figure 4.2). For example, there are instances when the KWIC index of the target word is not long enough to identify the meaning of the target word. The File View tool allows access to the wider text in which the target word occurs (e.g. an essay), so that any ambiguity about the usage of the target word can be resolved.

Figure 4.2 Screenshot of the AntConc File View tool

4.5   The procedure and criteria for selection of target verbs

This study was originally intended to study a broad range of verbs, but once it was underway

it became clear that that task would be so great that within the scope of this thesis it would

only be possible to cover a relatively small number of verbs. The verbs to be investigated

were selected according to a number of criteria. First, they had to be verbs that had already

been analysed for the PDEV project. Second, every selected verb from the PDEV entry

needed to have a minimum number of patterns, at least 3, and a maximum number of patterns,

up to 21. Second, the verbs selected had to occur relatively frequently in each of the target

corpora (ICLE-Chinese, ICLE-Swedish, and LOCNESS).

Having set these criteria, the next step was to look for verbs that fulfilled them. This step was

not straightforward, for several reasons. Since there are thousands of English verbs recorded

in the PDEV, it would have been time-consuming to go over every verb and then search for

them in the target corpora. Therefore, I decided to select only those verbs that occur with four

patterns: V *that*, V n *that*, V *to*-infinitive, V n *to*-infinitive. This is because such patterns are very common in English, and verbs that have these patterns tend to also have a range of other patterns which may represent problems for learners.

To start the process of selection, I referred to *Grammar Patterns 1: Verbs* (Francis et al. 1996)[15] to get complete lists of the verbs that have the patterns mentioned above. Tables 4.2 and 4.3 below show the verbs and their raw frequencies in the target corpora. In particular, Table 4.2 presents list of verbs that belong to V *that*-clause or V n *that*-clause patterns, taken from Francis et al. (1996), whereas Table 4.3 presents verbs that belong to V *to*-inf or V n *to*-inf, also taken from Francis et al.

Table 4.2 Verbs and their raw frequencies in the three target corpora (V *that*-clause/ V n *that*-clause)

No.	ICLE-CHINESE		LOCNESS		ICLE-SWEDISH	
	V	Raw frequency	V	Raw frequency	V	Raw frequency
1.	Accept	79	ACCEPT	180	ACCEPT	97
2.	ADAPT	19	ACQUIRE	19	ADAPT	21
3.	ADOPT	54	ADAPT	12	AFFORD	35
4.	ADVISE	21	ADVISE	9	ALLOW	74
5.	ADVOCATE	20	ADVOCATE	80	APPEAR	28
6.	AFFECT	606	AFFECT	92	ARRANGE	5
7.	AFFIRMS	2	AFFORD	39	ASK	63

---

[15] This book presents all patterns of English verbs and links these patterns to meanings (Francis et al. 1996). For example, it groups all verbs that are followed by a *that*-clause whose meaning belongs to the 'say' group (e.g. *agree*, *allow*, *accept*, *advise*, etc.), and separately groups all verbs that are followed by a *to*-infinitive and belong to the 'begin' meaning group (e.g. *begin*, *continue*, *commence*, etc.), and so on.

8.	AFFORD	189	AGREE	80	BOUND	9
9.	AGREE	209	AIM	69	BRAINWASH	5
10.	AIM	51	ALLOW	270	CHOOSE	80
11.	ALLOCATE	12	APPEAR	85	CLAIM	52
12.	ALLOCATE	12	ARGUE	164	CONSTRUCT	4
13.	ALLOW	203	ARRANGE	6	CONTINUEW	47
14.	APPEAL	21	ASK	113	DEVOTE	7
15.	APPEAR	52	ASPIRE	3	ENCOURAGE	34
16.	APPROVE	14	ASSUME	41	FAIL	27
17.	ARRANGE	2	BOUND	15	FORCE	94
18.	ASK	82	CALL	123	FORGET	64
19.	AVOID	104	CHOOSE	125	HESITATE	4
20.	BOTHER	2	CONTINUE	176	JUSTIFY	4
21.	BUILD	193	COST	111	LEAD	118
22.	CHOOSE	252	DEEM	17	LOVE	116
23.	CONDUCT	164	DENY	53	MANAGE	22
24.	CONSTRUCT	217	ENCOURAGE	77	NEED	323
25.	CONTINUE	47	ENGAGE	13	PLAN	34
26.	DENY	35	ENLIST	3	PROPOSE	4
27.	EDUCATE	74	FAIL	53	SAY (SAID)	305
28.	ENCOURAGE	181	FORCE	166	SEE	334
29.	EXPLAIN	61	FORGET	30	TELL	87
30.	FAIL	20	HATE	19		
31.	FORCE	201	LEAD	281		
32.	FORGET	46	LOSE	196		
33.	IMPORT	836	MANAGE	27		
34.	LEAD	482	NEED	438		
35.	LOSE	384	NEGLECT	14		
36.	MAINTAIN	61	PLAN	70		
37.	MANAGE	363	SACRIFICE	30		

38.	NEED	948	SAY	508
39.	NEGLECT	27	SEE	331
40.	OPEN	83	TELL	148
41.	PAVE	1	WORK	168
42.	PLAN	164		
43.	PROPOSE	65		
44.	SAY	458		
45.	URGE	37		
46.	Warn	22		
47.	Watch	507		
48.	Work	772		

Table 4.3 Verbs and their raw frequencies in the three target corpora (V *to*-inf/ V n *to*-inf)

No.	ICLE-CHINESE		LOCNESS		ICLE-SWEDISH	
	V	Raw frequency	V	Raw Frequency	V	Raw frequency
1	ACCEPT	79	ACCEPT	180	ACCEPT	97
2	ADD	35	ACKNOWLEDGE	14	ADD	25
3	ADMIT	40	ADMIT	40	ADMIT	20
4	ADVISE	21	ADVISE	9	AGREE	43
5	ADVOCATE	20	ADVOCATE	80	ANSWER	95
6	AFFECT	606	AGREE	80	APPEAR	28
7	AFFIRM	2	ALLEGE	3	ARGUE	30
8	AGREE	209	ANNOUNCE	11	ASSUME	12
9	AID	8	APPEAR	85	CLAIM	52
10	AMAZE	8	APPRECIATE	15	DENY	21
11	ANTICIPATE	1	ARGUE	164	EXPLAIN	42
12	APPEAR	52	ASK	113	FORGET	64

13	APPROVE	14	ASSUME	41	LEAD	118
14	ARGUE	336	AVOW	2	MAINTAIN	18
15	ARISE	18	CALCULATE	9	POINT	94
16	ASSERT	4	CLAIM	179	SAY	305
17	ASSUME	3	COMPLAIN	7	SEE	334
18	AVOID	104	DEBATE	88	STRESS	22
19	CLAIM	173	DECREE	4	TELL	87
20	COMPLAIN	13	DENY	53		
21	DEEM	6	DISTINGUISH	9		
22	DENY	35	EXCLAIM	2		
23	ENCOURAGE	181	EXPLAIN	73		
24	EXPLAIN	61	FOLLOW	124		
25	FACE	201	FORGET	30		
26	FORGET	46	MAINTAIN	40		
27	MAINTAIN	61	NOTE	29		
28	NEED	948	OBJECT	43		
29	NOTE	27	POINT	228		
30	POINT	684	PROPOSE	20		
31	PROPOSE	65	SAY	508		
32	SAY	458	SEE	638		
33	SEE	304	STRESS	26		
34	STRESS	51	SUBMIT	9		
35	TELL	87	TELL	148		
36	URGE	37	WARN	11		
37	WARN	22				
38	WATCH	507				

As seen from these tables, the raw frequencies of verbs are not the same across the three target corpora. For example, ACCEPT occurs 79 times in ICLE-Chinese, 180 times in LOCNESS, and 97 times in the ICLE-Swedish sub-corpus. In addition, some verbs are

infrequent in one target corpus while they frequently occur in the other. For example, MAINTAIN occurs 61 times in ICLE-Chinese and 40 times in LOCNESS, but only 18 times in the ICLE-Swedish sub-corpus.

This inconsistency led me to add more criteria for the selection of target verbs: first, since ICLE-Chinese was the primary learner sub-corpus for the present study, verbs in Tables 4.2 and 4.3 were eliminated according to their raw frequency; that is, verbs whose raw frequency was less than 50 or more than 200 in ICLE-Chinese were eliminated. These numbers are admittedly arbitrary but were expected to provide data that would be manageable in scope.

Given that, the target verbs studied are as follows:

AGREE, ALLOW, AVOID, ASK, AFFORD, ADVISE, APPEAR, ADMIT, CLAIM,

DENY, ENCOURAGE, MAINTAIN, PROPOSE, PLAN, TELL, URGE

These verb lemmas are the ones that fulfilled the above-mentioned criterion of minimum and maximum raw frequency in the set.

At this point, it should be mentioned that since each of the following three chapters of results and findings has different aims and a different focus, not all these target verbs are discussed in all three chapters. In addition, some chapters focus on divergent usages of the verb lemmas (Chapter 5), while others focus on under/over-representation of such verbs (Chapter 6).

4.6   Identifying patterns

Hanks (2004:88) summarises the CPA method of identifying verb patterns and their implicatures in a corpus as follows:

No attempt is made in CPA to identify the meaning of a verb … directly, as a word in isolation. Instead, meanings are associated with prototypical [patterns]. Concordance lines are grouped into semantically motivated syntagmatic patterns. Associating a 'meaning' with each pattern is a secondary step, carried out in close coordination with the assignment of concordance lines to patterns. The identification of a syntagmatic pattern is not an automatic procedure: it calls for a great deal of lexicographic art.

Patterns of the target verbs for this study were identified following a consistent methodology based on the process described above. This was translated into concrete steps for the present study as below.

1. Each target corpus was uploaded into AntConc version 3.4.4 (Anthony 2005) to generate concordance lines for the target verbs. The target corpora were stored as .txt files (i.e. as plain text).

2. Once the text files were uploaded, the next step was searching for every target verb lemma. Since the ICLEv1 comprises plain text, the search for target verbs was done by querying the word-forms (e.g. *agree*, *agrees*, *agreed*, *agreeing*). AntConc 3.4.4 allowed me to search for all word-forms at a time. This was done by inserting the '|' symbol between each form of each target verb (e.g. agree|agrees|agreed|agreeing). However, this way of querying also required additional cleaning of data. For example, when PLAN was searched for in the target corpora, the raw frequency was too high. By looking at the concordance lines of PLAN, I found that a quite number of instances of PLAN were nouns rather than verbs. Thus, this issue was easy to deal with by simply eliminating

such instances.

3.   Once the concordance lines of the target verb were generated in KWIC format, the lines were copied from AntConc and pasted to an Excel worksheet.

4.   As the concordance lines for every target verb were copied to the Excel worksheet, each concordance line was studied manually and annotated. This was done by matching the patterns of the target verb identified in the corpora to those in the PDEV. Those instances that did not match the patterns in the PDEV were put aside for further investigation. There were two types of unmatched instances: divergent instances and unclassified instances. The divergent instances are patterns that are recognisable and produced by several learners. The unclassified patterns are patterns that are one-off errors. For example, the instance *The teacher agreed the students to use my mobile phone in the classroom* is considered divergent. This is because this pattern is recognisable, in that it belongs to the verb *allow*: *allow* can be followed by a noun followed by a *to*-infinitive (e.g. *I allowed him to use my mobile phone*). In contrast, an instance such as *The car is allowed to drive* is labelled as unclassified because it would be difficult to describe that in pattern terms and because it only occurred once.

5.   Having annotated the concordance lines of the target verb, non-divergent instances that shared the same pattern and meaning were grouped together and counted. Divergent instances that were produced by several learners and shared the same pattern and meaning were grouped into sets. Every set of divergent patterns was searched for in the BNC to make sure it was not a pattern that had not found its way into PDEV merely due to the small size of the sample analysed by the lexicographers. To illustrate this process, a worked

81

example has been provided below.

For example, in the PDEV entry for *encourage*, 3 patterns are recorded from the annotation of 250 instances in the BNC50 as shown in Table 4.4 below.

Table 4.4 PDEV entry for ENCOURAGE[16]

Pattern number	Pattern and implicature	% BNC50 sample based on 250 lines
1	**Pattern**: Human 1 or Institution 1 or Eventuality encourages Human 2 or Institution 2 *to*-infinitive **Implicature**: Human 1 or Institution 1 or Eventuality has the effect of causing Human 2 or Institution 2 to want to do something **Example**: The chemists encouraged people to have fun in the sun[17]	46.4%
2	**Pattern**: Human 1 or Institution 1 or Eventuality encourages Human 2 or Institution 2 **Implicature**: Human 1 or Institution 1 or Eventuality has the effect of causing Human 2 or Institution 2 to feel more confident or positive **Example:** The good trainees were thus encouraged	5.6%
3	**Pattern**: Human or Institution or Eventuality 1 encourages Eventuality 2 **Implicature**: Human Institution or Eventuality 1 has the effect of causing Eventuality 2 to be more likely **Example**: CAP reform must cut costs and prices and encourage structural rationalisation	47.6%

Column 1 indicates the pattern number in the PDEV entry for ENCOURAGE. Column 2 contains a description of the pattern, that is, the form of the pattern and its implicature. Column 3 indicates how common this pattern was in the sample of data analysed, shown as a percentage of the total sample size of 250 concordance lines.

One hundred and seventy-four concordance lines of ENCOURAGE were manually annotated in the ICLE-Chinese corpus and checked against the patterns recorded in the PDEV entry for

---

[16] Adapted from original PDEV entry http://pdev.org.uk.
[17] For simplicity, I replaced the original example with the current one for an easy read.

ENCOURAGE, respectively. Of 174 lines, 7 instances produced by the Chinese learners did not match any of the three patterns in Table 4.4 above. These 7 instances instantiate a pattern: Human encourages Human 2 do something, as shown in Figure 4.3, below. This divergent pattern was searched for in the BNC. The result revealed that such a pattern does not exist in the BNC, which confirmed that it is divergent. Finally, an adapted PDEV entry was built for every divergent set of patterns of the target verbs identified in the target corpora, as shown in the present chapter.

Although this chapter and each of the following three chapters of results and findings has its own aim (Chapters 5, 6, and 7), the above-mentioned steps were followed throughout the analysis presented in all three chapters (i.e. using the CPA/PDEV method to interpret instances).

Figure 4.3 Concordance lines of ENCOURAGE followed by a noun group and infinitive (without *to*) in ICLE-Chinese

1	enlighten the grade school child,	encouraging	them <u>participate</u> with the environmenta
2	there was the need for attracting and	encouraging	people with talent <u>come</u> to Hong Kong.
3	of Mainland Professional Scheme to	encourage	professionals from Mainland China <u>work</u> in Hong Kong
4	tend to offer high credit limits to	encourage	spending students <u>consider</u> loweri
5	soccer betting. Actually, gambling	encourages	the youth <u>want</u> to get adventages withou
6	abortion a women and the abortion	encourages	the teenagers <u>have</u> a more liberal attit
7	for opposing it is that it would	encourage	youth <u>involved</u> in gambling. In fact, footba

4.7   Pilot study

A pilot study was carried out on the 16 target words. The aim of the pilot study was to get an overview of the use of the target verbs in all the corpora in order to identify issues that might

be investigated further. Each word was searched for in all three target corpora, ICLE-Chinese, ICLE-Swedish and LOCNESS, by following the steps in section 4.5. Having identified and annotated the concordance lines of every target word in every target corpus, the frequency of the outcomes of every target word was recorded in a table, presented as Table 4.5 below:

Table 4.5 Survey of the target verbs

No.	Verb	Chinese		LOCNESS		Swedish		NO. OF patterns in PDEV
		No. of matched instances with PDEV	No. of divergent instances	No. of matched instances with PDEV	No. of divergent instances	No. of matched instances with PDEV	No. of divergent instances	
1	ADMIT	40	0	40	0	20	2	13
2	ADVISE	11	6	9	0	1	1	9
3	AFFORD	189	14	39	0	35	0	5
4	AGREE	157	51	80	0	42	0	10
5	ALLOW	164	39	270	0	73	1	8
6	APPEAR	52	3	85	0	28	0	12
7	ASK	66	16	113	0	55	2	12
8	AVOID	71	33	37	0	31	2	3
9	CLAIM	173	0	179	0	52	0	6
10	DENY	35	0	53	0	21	0	9
11	ENCOURAGE	158	23	77	0	34	0	3
12	MAINTAIN	61	0	40	0	18	0	5
13	PLAN	50	5	64	0	15	0	4
14	PROPOSE	63	2	22	0	3	0	8
15	TELL	55	32	104	0	74	1	21
16	URGE	30	7	3	0	3	0	6
Total		1375	231	1162	0	505	9	

Column 2 includes the target verbs that have been completed in the PDEV; columns 3, 5 and 7 indicate the numbers of non-divergent instances identified in the respective target corpora (e.g. instances that matched the patterns in the PDEV entry); columns 4, 6 and 8 contain the numbers of divergent instances identified in the target corpora (e.g. instances that did not match the patterns in the PDEV entry); column 9 includes the number of patterns of each target verb recorded in the PDEV entry.

As seen in Table 4.5, the number of divergent instances is not the same across the target corpora. More specifically, the Chinese learners appear to be producing more divergent instances than in the other two corpora. The Swedish learners, on the other hand, are close to the native-speaker writers, in that they produce very few divergent instances. Given this difference, it was decided that focusing on the divergent instances identified in the ICLE-Chinese sub-corpus would reveal some interesting results. Thus, Chapter 5 is devoted to divergent patterns of the target verbs identified in the ICLE-Chinese sub-corpus.

Another important decision relating to the non-divergent instances was made, as follows. Table 4.5 shows that the numbers of divergent instances of the target verbs in ICLE-Swedish is almost zero. This suggests that studying divergent uses in the Swedish corpus would be unproductive, as there are very few such uses. Instead it was decided to study the relative frequency of each pattern of selected verbs, focusing on under- and over-use in comparison with LOCNESS. This is done in Chapter 6.

Finally, since all 16 target verbs were taken from the PDEV entries, it was felt that it would be interesting if I could add a chapter studying a verb that has not been completed yet in the PDEV. Chapter 7 was thus devoted to the study of uses of SUGGEST in the learner corpora

and the native-speaker ones.

4.8 Conclusion

This chapter has introduced the data for investigation, from ICLE, LOCNESS, and the BNC. The justification for the selection of sub-corpora has been provided. Then, the tool used to investigate data has been introduced and illustrated with some figures. The several phases of the process of selection of target verbs have been discussed, as has the method of identifying and interpreting the uses of target verbs; this latter method has also been illustrated with a worked example. Finally, a pilot study of the target verbs identified in the target corpora has been presented, and decisions concerning the aims of the results chapters made on its basis have been provided. We will now move on to these chapters presenting the results and findings: chapters 5, 6, and 7.

Chapter 5   **Divergent Uses of Verbs in ICLE-Chinese**

5.1   Introduction

The aim of this chapter is to demonstrate that PDEV can be used to successfully identify divergent usages of target verbs identified in the Chinese section of the International Corpus of Learner English (ICLE-Chinese). Sixteen target verbs were chosen for this study, as described in Chapter 4: ADMIT, ADVISE, AFFORD, AGREE, ALLOW, APPEAR, ASK, AVOID, CLAIM, DENY, ENCOURAGE, MAINTAIN, PLAN, PROPOSE, TELL, and URGE. These were used to establish overall differences in frequency between the two corpora (section 5.5). Of these 16 verbs, 5 were then selected for further investigation: AVOID (section 5.5.3), ALLOW (section 5.5.4), ENCOURAGE (section 5.5.5) AGREE (5.5.6), and LEAD (5.5.7). The investigation focused on patterns of these target verbs identified in ICLE-Chinese that do not match the ones in the Pattern Dictionary of English Verbs.

The chapter starts by introducing its research questions. Then, it discusses the issue of terminology and the distinction between 'innovation' and 'error'. Next, the issue of 'systematicity' and 'randomness' is discussed, and the method of identifying and interpreting verb patterns is illustrated. After that, the results and findings are presented and discussed. Finally, a conclusion is provided.

5.2   Research questions

The research questions for this chapter are set out as follows:

1. Can PDEV be used to successfully identify divergent patterns of target verbs identified in the ICLE-Chinese corpus?

2. What do these divergent patterns tell us about learners' competence and global English?

5.3   Terminology: 'Error' vs. 'innovation'

In this section I discuss the difference between an error and an innovation, and I introduce the novel term 'divergence' that will be used throughout this chapter. I have chosen this term as it indicates difference from the native-speaker norm but does not stigmatize that difference as an error.

In general, unconventional usages by language learners or low-prestige speakers are interpreted as errors, whereas if a high-prestige speaker uses the same language, this is an innovation. For example, the word 'likely' which used to be only an adjective (e.g. *This is still the likely explanation*) has fairly recently started to be used as an adverb (e.g. *She will likely address this issue tomorrow*). If only learners did this, it would be seen as an error, whereas because it is widely adopted by native speakers it is an innovation and an instance of language change. However, there is an area of dispute here, because many people would argue that learners are also permitted to innovate. And a distinction is sometimes made in this regard between EFL and ESL learners (see Edwards and Lange 2016; Schneider and Gilquin 2016). Edwards and Lange (2016:252–253) link the notion of innovation to 'structural nativisation'. Structural nativisation is defined by Schneider (2007: 5–6) as "the emergence of locally characteristic linguistic patterns", in that different uses of linguistic patterns are only observable in specific ESL countries. This distinction emerges from Kachru's (1985) Three Circles model, which, however, is no longer believed to be an accurate model of world English. In this model, the 'Inner Circle' countries include the UK, the US, and those former settler colonies where English is the dominant first language; the 'Outer Circle' includes non-settler postcolonial societies (e.g. India, Singapore) where English is not the dominant first language; and the 'Expanding Circle' refers to countries where English has traditionally been taught and used only for purposes of trade and international communication, such as China,

Japan, Sweden, and many others (Kachru 1985). Similarly, Quirk et al. (1972:3–4) presents the three circles as 'native' (ENL) 'second' (ESL), and 'foreign' (EFL) circles. This in turn shows that EFL learners are not allowed to be creative (i.e. deviations are still regarded as errors).

Several recent corpus-based studies reveal, however, that EFL and ESL learners share some similarities in producing non-native like usages in the domain of lexis/lexico-grammar (Edwards and Laporte 2015; Gilquin 2011; Gilquin and Granger 2011; Götz and Schilk 2011; Laporte 2012; Nesselhauf 2009, cited in Schneider and Gilquin 2016), which suggests that there is an increasing agreement that learners can be permitted to innovate and that learner innovation should be treated rather like native-speaker innovation. A recent study by Schneider and Gilquin (2016) aims to present a method, the collocation ratio, to detect innovations and describe differences and similarities between EFL and ESL in terms of some linguistic patterns (i.e. over/or –under representation of usages such as verb + preposition and adjective + preposition). They used collocation ratio to analyse EFL data from ICLE and some ESL sub-corpora from the International Corpus of English (ICE; Nelson et al. 2002). Their analysis revealed that some non-standard patterns are shared by EFL and ESL learners, and they conclude that the distinction between the EFL and ESL is no longer valid. A similar comment is made by Deshors et al. (2016:7-8), who say that such a conclusion increases "the credibility of EFL learners in terms of their own ability to be creative in their L2".

Following Schneider and Gilquin (2016), even though the corpus being investigated in the present study is produced by EFL learners, I will treat patterns in the Chinese corpus as 'divergences' rather than errors. By 'divergence' I mean a pattern that is recognisable and has unambiguously 'correct' usages with some given verb(s), that is, in which a 'wrong' verb has

been used with a pattern that belongs to a 'right' verb. For example, some learners produced

divergent instances of ALLOW such as *I allowed her check my mobile phone*, a pattern that is

unambiguously legitimate with the non-divergent verb *let*, as in *I let her check my mobile*

phone.

Having discussed this issue of terminology, the accompanying distinction between

'systematicity' and 'randomness' and its relevance will be discussed in the following section.

5.4   Identifying and confirming divergence in ICLE-Chinese

The chapter uses PDEV as a benchmark to identify divergent uses of target verbs in the ICLE-

Chinese corpus.

I have looked at the concordance lines and identified usages that are not found in PDEV. I

have then checked those usages in the *Grammar Patterns* book and the BNC. If a usage is not

found in any of these sources, I regard it as a divergence. These steps will be shown in section

5.4.2. The next section will introduce the two resources used to double-check that the

divergent patterns identified in the ICLE-Chinese are not just missing patterns from the

PDEV entry, since PDEV is still in progress.

5.4.1   References used along with PDEV

The Cobuild pattern grammar series consists of two volumes, one of which is *Grammar*

Patterns 1: Verbs by Francis et al. (1996). It is based on the Bank of English corpus. The aim

of the volume is to present all the verb patterns and all the verbs with each pattern, and to sort

the verbs into meaning groups (Francis et al. 1996). These patterns are given codings. For

instance, V n *to*-inf means 'verb followed by a noun group and *to*-infinitive (e.g. someone

asks someone to do something)' and V *to*-inf means 'verb followed by a *to*-infinitive (e.g. someone *agrees to do* something), and so on. There are a number of reasons for the adoption of Francis et al. (1996) as a double-check reference for divergent patterns. First, PDEV only records prototypical patterns. There is always a possibility that some pattern actually does exist but is quite rare; in principle such a pattern would be recorded in Francis et al. (1996). Second, PDEV is based on a relatively small corpus, whereas Francis et al. (1996) is based on a large corpus. Thus, consulting Francis et al. (1996) enables me to double-check if the divergent pattern was not merely omitted from PDEV. In particular, I will use this volume to double-check the grammatical structure of identified divergent uses of a target verb and what comes after it (e.g. suggest *that*-clause, agree *with*, allow n *to*-infinitive, etc.).

Another reference used to double-check whether the divergent pattern does exist is the British National Corpus (BNC). As the sub-corpus of the BNC used for building the PDEV was the written part, my search will exclude the spoken part. The advantage of using the BNC (CQL version) to check the identified patterns is that the BNC shows the full text around where instances occur. This bears out the classification of the instance as divergent. In addition, the BNC (CQL version) allows me to count numbers of instances. This feature allows me to decide whether a particular use of a verb identified in the learners' corpus, is frequent enough to add to PDEV entry for a verb under question.

The process of interpreting the patterns with a worked example will be illustrated in the following section.

5.4.2 Interpreting the divergent patterns

This chapter applies the CPA method described by Hanks. Once the concordance lines of a particular verb have been annotated in the ICLE-Chinese corpus and sorted into patterns that have the similar syntactic structure and implicature, following the CPA technique described above, the next step is to check these patterns against the patterns recorded in the PDEV. For example, in the PDEV entry for ENCOURAGE, 3 patterns were recorded from the annotation of 250 instances in the BNC50 as shown in Table 5.1 below:

Table 5.1 ICLE-Chinese PDEV entry for ENCOURAGE

Pattern number	Pattern and implicature	% BNC50 sample based on 250 lines
1	**Pattern**: Human 1 or Institution 1 or Eventuality encourages Human 2 or Institution 2 *to*-infinitive **Implicature**: Human 1 or Institution 1 or Eventuality has the effect of causing Human 2 or Institution 2 to want to do something **Example**: The chemists encouraged people to have fun in the sun	46.4%
2	**Pattern**: Human 1 or Institution 1 or Eventuality encourages Human 2 or Institution 2 **Implicature**: Human 1 or Institution 1 or Eventuality has the effect of causing Human 2 or Institution 2 to feel more confident or positive **Example:** The good trainees were thus encouraged	5.6%
3	**Pattern**: Human or Institution or Eventuality 1 encourages Eventuality 2 **Implicature**: Human Institution or Eventuality 1 has the effect of causing Eventuality 2 to be more likely **Example**: CAP reform must cut costs and prices and encourage structural rationalisation	47.6%

Column 1 indicates the pattern number in the PDEV entry for ENCOURAGE. Column 2 contains a description of the pattern, that is, the pattern and its implicature or definition. Column 3 indicates how common this pattern was in the sample of data analysed, shown as a percentage of the total sample size of 250 concordance lines. In all, 174 concordance lines of

encourage were manually annotated in the ICLE-Chinese corpus and checked against the

patterns recorded in the PDEV entry for ENCOURAGE. Of 174 lines, 16 instances did not

match any of the three patterns in Table 5.1 above; of those 16 instances, 7 followed the

pattern *Human encourages Human infinitive* without *to*, as shown in Figure 5.1 below:

Figure 5.1 Concordance lines of ENCOURAGE followed by noun group and infinitive (without *to*) in ICLE-Chinese

1	the grade school child,	encouraging	them participate with the environmenta
2	there was the need for attracting and	encouraging	people with talent come to Hong Kong.
3	the Admission of Mainland to	encourage	professionals from Mainland China work in Kong
4	tend to offer high credit limits to	encourage	spending students consider loweri
5	Actually, gambling	encourages	the youth want to get adventages withou
6	of abortion a women and the abortion	encourages	the teenagers have a more liberal attit
7	reasons for opposing it is that it would	encourage	youth involved in gambling. In fact, footba

These instances are divergent instances that were found in the learners' corpus, and did not

conform to any of the patterns of ENCOURAGE in the PDEV entry. I then double-checked

Grammar Patterns to see if the grammatical structure in this set of instances is previously

attested. After that, I searched for this pattern (*Human Encourages Human infinitive*) in the

BNC data, as a final check, to see if any instances that instantiate this pattern are found there.

In this case, since no instances of such a pattern were found in either *Grammar Patterns* or

the BNC data, it was confirmed that such instances are divergent.

The reader must be reminded that PDEV is an in-progress project, slated for further updates.

This makes the backup sources and confirmation/verification process indispensable, for two

reasons: first, to check if such instances are just missing from the PDEV as it stands; second,

to make sure that such instances are (systematic) divergences and not (from the perspective of the whole data, even if motivated by the status of individual learners, random) errors, thus putting them aside for further investigation. If the results are inconsistent, I will refer to a native-speaker rater.

Since the verb pattern in Figure 5.1 (i.e. ENCOURAGE followed by a noun group and infinitive verb without *to*) was not found in either *Grammar Patterns* or BNC, this pattern was confirmed to be divergent (i.e. non-standard). In order to find out what the learners meant to convey by using this pattern, a further step was taken: checking the implicatures of all patterns in the PDEV entry for ENCOURAGE, as shown in Table 5.1 above. An examination of the instances in Figure 5.1 suggests that they are closest in meaning to pattern 1 in Table 5.1: *Human 1 or Institution 1 or Eventuality has the effect of causing Human 2 or Institution 2 to want to do something*. This in turn suggests that pattern 1 is the one that the learners are aiming for.

To see to what extent the instances produced by ICLE-Chinese in Figure 5.1 are random or systematic, three criteria for being systematic are set. First, the number of identified instances should not be less than five. Second, all the identified instances should be produced by more than one learner. Third, all the instances should have the same implicature. Since the set of instances together satisfied the three criteria, but not all seven instances separately, it can be regarded as a case of systematic error, that is, a divergent pattern.

On the other hand, there was an instance in which a pattern was classified as a random error. This is presented in Figure 5.2 below:

Figure 5.2 Concordance lines for the verb *encourage* followed by noun group and *that*-clause in ICLE-Chinese corpus

1	result. It's a good example to encourage the other criminals <u>that</u> lives could rewrite

As can be seen, here, the verb *encourage* is followed by a noun group and *that*-clause. According to Frances et al. (1996), ENCOURAGE is not found in any of the meaning groups that have this pattern: V n *that*. To validate this assertion, ENCOURAGE was run in a BNC query, and the result showed no instances with this pattern. In addition, this pattern was not recorded in the PDEV entry for ENCOURAGE, as shown in Table 5.1. A further investigation of the PDEV description revealed that a more appropriate verb in this case would be *tell*. The implicature of *tell*, as described in the PDEV, is as follows: *Human 1 informs Human 2 that [clause]*. Both *tell* and *encourage* thus have similar meanings, but different patterns; in this case, *encourage* is being used with the pattern associated with *tell*. Therefore, since this deviation does not fulfil the criteria for being an innovation (only used by one learner, not acceptable by other Chinese learners), this pattern was regarded as an error.

Having discussed how to identify and interpret patterns, a summary of the main findings, together with answers to the research questions, is provided in the next section.

5.5   A survey of the target verbs

As a reminder, the research questions for this chapter were set out as follows:

1. Can PDEV be used to successfully identify divergent patterns of target verbs identified in the ICLE-Chinese corpus?

2. What do these divergent patterns tell us about learners' competence and global English?

One of the important purposes of this thesis is to see how useful CPA/PDEV is to find deviations in verb patterns in learners' writing in ICLE-Chinese and to determine to what extent PDEV helps in interpreting the implicatures of the deviated verb patterns as far as the ICLE-Chinese corpus is concerned. To answer this question, 16 verbs were selected for investigation. The verbs to be investigated were selected according to a number of criteria: first, they had to be verbs that had already been analysed and completed in their PDEV entry; second, the verbs selected had to occur relatively frequently in the target corpora; third, it was important to select verbs that had no less than three patterns in PDEV.

It must be noted that the range of patterns set in the third criterion was for comparison purposes. In a later chapter (Chapter7), a comparison of one selected verb between Chinese learners and native speakers will be carried out.

Originally, this section was intended to include the Louvain Corpus of Native English Essays (LOCNESS) for comparison with the ICLE-Chinese. The reason behind including LOCNESS was the expectation that native speakers also make mistakes in their writing but that they will differ in type and pattern from non-native speakers' mistakes in systematic ways. However, the analysis of the 16 target verbs in LOCNESS revealed that no major mistakes were identified that were worth presenting in this section. Thus, the exclusion of the LOCNESS data was settled on.

Based on the set criteria for the verbs under study, what follows is a survey of the data for the

16 target verbs: ENCOURAGE, TELL, ALLOW, ASK, AGREE, CLAIM, DENY, AFFORD,

MAINTAIN, ADMIT, APPEAR, PLAN, ADVISE, PROPOSE, URGE, and AVOID. Two

assumptions lie behind the selection of these verbs. First, as these verbs are concerned with

forms of communication and as the target corpora (ICLE-Chinese and LOCNESS) consist of

a collection of mainly argumentative essays, such verbs are expected to offer interesting scope

for comparison with LOCNESS, which might yield some important implications for learners.

Second, as far as the deviations are concerned, the investigation of these verbs might reveal

some syntactic structures that are different from the norms for the respective verbs. The 16

verbs are presented in Table 5.2 as follows:

Table 5.2 The frequencies of the 16 verbs in the ICLE-Chinese corpus

No.	Verbs		Chinese		
		No. of concordance lines	No. of matched instances to the PDEV	No. of divergent instances	No. of patterns in the PDEV
1	ADMIT	40	40	0	13
2	ADVISE	17	11	6	9
3	LEAD	425	381	44	12
4	AGREE	208	157	51	10
5	ALLOW	203	164	39	8
6	APPEAR	55	52	3	12
7	ASK	80	66	16	12
8	AVOID	104	71	33	3
9	CLAIM	173	173	0	6
10	DENY	35	35	0	9
11	ENCOURAGE	81	158	23	3
12	MAINTAIN	61	61	0	5
13	PLAN	55	50	5	4
14	PROPOSE	65	63	2	8

15	TELL	87	55	32	21
16	URGE	37	30	7	6
Total			1,375		231

This table is sorted by decreasing number of deviations (incorrect patterns). The second column contains the target verbs. The third column lists the raw frequencies of the correct patterns (i.e. the patterns that correctly matched the ones in the PDEV for that verb). The fourth column presents the raw frequencies of the deviations (i.e. the patterns that did not match the ones in the PDEV); the fifth column shows the raw frequency as a percentage. It is beyond the scope of this thesis to discuss all of these 16 verbs in an exhaustive way. Thus, I set the following criteria to eliminate some verbs in Table 5.2:

Criterion one: if the number of PDEV patterns of any verb in the table is less than 12, the verb should be eliminated.

Criterion two: if the number of divergent instances of any verb in the table is less than 5, the verb should be eliminated.

Criterion three: if the number of concordance lines identified in the ICLE-Chinese for any verb is less than 81, the verb should be eliminated.

The application of these criteria to the verbs in Table 5.2 results in only five remaining verbs, as shown below.

Table 5.3 The frequencies of the five verbs identified in the ICLE-Chinese corpus

No.	Verbs	Raw frequency	Non-divergent usages	Divergent ('incorrect') usages
1	AGREE	209	157 (75.12%)	40 (19.14%)
2	ALLOW	203	164 (80.79%)	35 (17.24%)

3	AVOID	104	71 (68.27%)	33 (31.73%)
4	ENCOURAGE	181	158 (87.29%)	9 (4.97%)
5	LEAD	483	381 (78.88%)	44 (9.11%)
Total		1,180	931 (78.90%)	161 (13.65%)

Table 5.3 shows the frequencies of five verbs identified in the ICLE-Chinese corpus.
This table is again sorted by decreasing percentage of divergent usages of the five verbs
identified in the ICLE-Chinese corpus. The second column contains the target verbs under
study. The third column lists the raw frequencies of the verbs; the fourth column shows the
raw frequencies and percentage rates of non-divergent patterns of every verb (i.e. the patterns
that correctly matched the ones in the PDEV); the fifth column presents the raw frequencies
and percentage rates of the verb usages that were divergent from those in the PDEV). The
reader should remember here that the term 'divergent' is intended to include any verb patterns
identified in the ICLE-Chinese corpus that are divergent from those in the PDEV and contain
one of the five target verbs in Table 5.3. For validity purposes, all of the divergent verb
patterns will also be searched for in the BNC to see whether they do occur and simply are not
included in PDEV.

As can be seen, the proportion of divergent usages in Table 5.3 is not the same across all the
verbs. AVOID (31.73%)*,* AGREE (19.14%), and ALLOW (17.24%) each fall into the highest
category for the divergent usages (above 10%). LEAD (9.11%) and ENCOURAGE (4.97%)
then come at the end of the table, showing the lowest proportion of divergent usages (below
10%).

While the percentage is a useful tool for comparison, the raw frequency is equally important. For instance, although LEAD comes at the end of Table 5.3, with a percentage of 9.11% of the divergent usages, its raw frequency (44) shows the highest frequency of divergent usages compared with the ones for the other target verbs in Table 5.3. This is because LEAD is the most common of the verbs in the data. This suggests that researchers should take both the percentage and the raw frequency into account and not limit themselves to one of them.

Although quantitative studies "can be interesting starting points for further quantitative analyses, they do not usually in themselves contribute much to language learner analysis …" (Nesselhauf 2004:136), and "more qualitative analyses [should be] carried out." The present study, therefore, will extend the analysis of the divergent usages of the target verbs, as presented in Table 5.3, to qualitative analysis, in order to see what the divergences are in ICLE-Chinese as far as the verbs under study are concerned.

The following section will present analyses of the divergent usages of the verbs in Table 5.3. It starts with the verbs that show the highest proportion of divergence and concludes with the verbs that show the lowest proportion of divergent usages.

5.5.1   Qualitative analyses

In the previous section, Table 5.3 showed that there were in total 161 concordance lines classified as presenting divergent usages of the five target verbs identified in the ICLE-Chinese corpus. Qualitative investigation was carried out—specifically, I looked at each instance and derived a classification of divergence types. The results suggest three categories: first, divergence types that occur frequently (more than four occurrences); second, divergence types that occur infrequently (fewer than four occurrences); third, unclassified divergences.

This re-raises the notion of 'systematicity' and 'randomness' introduced in the previous chapter.

The following section will be divided into two parts: (1) the findings on the systematic divergent usages of the target verbs under study; (2) the findings on the random divergent usages of the target verbs. The unclassified divergences will not be dealt with, as they either did not make any sense or were produced only once by one learner.

The section that follows will begin by presenting the systematic divergent usages of the target verbs; then, the random divergences of the same target verbs will be presented.

5.5.2   Systematic divergences

Before presentation of the findings on the systematic divergences identified in the ICLE-Chinese, it is important to re-highlight the criteria set out for the systematic divergent verb patterns, as presented in section 4.5.1. To qualify for systematicity, three criteria should be met in a given set of divergent verb usages. First, the number of instances of the divergent verb pattern should not be less than five[18]. Second, every set of no less than five annotated lines should be produced by two learners or more [19]. Third, every set of all instances together should constitute the same pattern and implicature.

What follows is the findings on the systematic divergent usages of the five target verbs identified in the ICLE-Chinese corpus—AVOID, ALLOW, ENCOURAGE, AGREE, and LEAD.

---

[18] This is an arbitrary number that many studies employ.
[19] To find out whether the divergence is used by more than one Chinese learner, I manually checked the text files where each instance occurred.

5.5.3 AVOID

The total raw frequency of AVOID in the ICLE-Chinese corpus is 104. Thirty-three

(31.273%) instances were identified as divergent usages of AVOID. Of these 33 instances, 29

were regarded as systematic, as they fulfilled the criteria of systematicity.

Table 5.4 below shows four systematic divergent patterns of AVOID identified in the ICLE-

Chinese corpus but not found in the PDEV or the BNC. This table has been configured

following PDEV conventions.

**Table 5.4 Divergent patterns of AVOID identified in the ICLE-Chinese corpus but not
found in the PDEV or the BNC**

Pattern number	Pattern and implicature	Raw Frequency ICLE-CHINESE	Raw Frequency BNC
1	**Pattern**: Human 1 or Eventuality avoids Human 2 infinitive	5	0
	Implicature: Human 1 or Eventuality 1 prevents Human 2 from doing activity		
	Example: we can avoid so many person lose their lives It can avoid students get money from parents		
2	**Pattern**: Human 1 or Eventuality avoids Human 2 *to*-infinitive (*to* followed by verb)	8	0
	Implicature: Human 1 or Eventuality prevents Human 2 from doing activity		
	Example: It can avoid the people to depend on the benefit		
3	**Pattern**: Human avoids *to*-infinitive	6	0
	Implicature: Human takes action to stop himself, herself, or others doing activity		
	Example: They avoid to get into serious problems		
4	**Pattern**: Eventuality 1 avoids Eventuality 2	11	5+

Implicature: Eventuality 1 prevents Eventuality 2 from
occurring

Example: Recycling can avoid much pollution

In the table above, column 1 shows the pattern number; the first line of column 2 shows the pattern, the second line presents the implicature, and the third line gives an example of the pattern in use; column 3 shows the raw frequency of the pattern in ICLE-Chinese; and column 4 presents the raw frequency of the pattern in the BNC, if any.

What follows is four sets of divergent patterns of *avoid* identified in the ICLE-Chinese corpus. The first set (Figure 5.3) comprises five instances of the first divergent usage by Chinese learners.

Figure 5.3 Set of divergent usages of AVOID identified in ICLE-Chinese (pattern 1)

1	if monitoring is in place to	avoid	**employers** *use* it to import
2	their children tell to them. **It** can	avoid	**student** *get* money from parents
3	murdered, and so on. However, **we** can	avoid	so many **person** *lose* their lives.
4	of their financial affairs and **it** may	avoid	**students** *have* serious debt in
5	As the economic downturn, **it** may not	avoid	**people** include students *use* credit card

The divergent pattern of *avoid* that all these instances represent in Figure 5.3 is *Human or Eventuality avoids Human infinitive*. This divergent usage was not found in the BNC. This confirms that the Chinese learners' usage diverges from that of native speakers.

At this point, as this set of divergent instances has been verified, the next step is to find out what this set of instances means. An investigation of these instances revealed that their implicature is *Human 1 or Eventuality 1 prevents Human 2 from doing activity*.

A second set of divergent usages identified for AVOID in ICLE-Chinese is presented in Figure 5.4 below.

Figure 5.4 Set of divergent usages of AVOID identified in ICLE-Chinese (pattern 2)

1	problems. Inspections are needed to	avoid	**teenagers** *to get* wrong messages.
2	rightly. There are many ways of	avoiding	**children** *to watch* TV and letting them
3	monitoring for the **owners**. In	avoiding	**customers** *to smoke* in the restaurants
4	**using credit card** which has been argued may	avoid	**them** *to hold* a lot of money. However
5	finance. indicates that **student** can	avoid	**them** *to get* serious debt problems in
6	health seriously. **Banning smoking** can	avoid	**you** *to* be a second-hand smokers or
7	retirement's fund will be more. **It** can	avoid	the **people** *to depend* on the benefit
8	with the credit cards. **This** can	avoid	**students** *to bring* much more money

Eight divergent usages of AVOID are presented in Figure 5.4. As can be seen, the set of all instances constitutes the pattern *Human 1 or Eventuality avoids Human 2* to-infinitive (*to* followed by verb). The implicature of all instances together is construed as *Human 1 or Eventuality prevents Human 2 from doing activity*.

The third set of divergent usages of AVOID identified in ICLE-Chinese corpus is shown in Figure 5.5 below.

Figure 5.5 Set of divergent usages of AVOID identified in ICLE-Chinese (pattern 3)

1	maximining to use of resources to	avoid	*to waste* of time, the
2	post with similar work nature to	avoid	*to contribute* the MPF.
3	**Students** using credit cards can	avoid	*to bring* so large amounts
4	and bars better. This can cause **us**	avoiding	*to breath* secondhand smoke
5	of eighteen or nineteen. **They**	avoid	*to get* into serious problems
6	found that most **smokers** would	avoid	*to eat* out if smoking was banned

The six annotated lines altogether constitute the pattern *Human avoids* to-*infinitive*. The implicature of all instances together is construed as *Human takes action to stop himself, herself, or others doing activity.*

The last set of divergent usages of AVOID is presented in Figure 5.6 below:

Figure 5.6 Set of divergent usages of AVOID identified in ICLE-Chinese (pattern 4).

1	debt can evenly distributed **which** can	avoid	great debt burden. By contrast
2	the benefits of economy, **recycling** can	avoid	much pollution. Also, I would
3	industrial use. Then the urban **area** can	avoid	over-crowding. Hong Kong
4	**choosing** the sex of the baby can	avoid	some hereditary diseases.
5	of to burn it or landfill it, **this** can	avoid	the harmful by-products.
6	space. **Recycling** of Plastic products	avoid	the above problems so that it
7	globally and in some cases, **it**	avoids	the troubles of carrying too
8	to saving money, **recycling** also	avoid	the production of pollutants.
9	and can also ensure your safety	avoid	the high unemployment rate
10	also provide an advantage that to	avoid	the problems of toxic by-products

11   of saving our lives from hell, **it** can avoid any inherited diseases such as

What makes this set divergent from the one in the PDEV is the lexical items that fill the

subject slots. The verb pattern is *Eventuality 1 avoids Eventuality 2*, and thus the semantic

type that fills the subject slots of AVOID is *Eventuality*, whereas the semantic type that fills

the subject slot of pattern 1 AVOID in the PDEV is *Human* or *Institution*, as in *He avoided so*

many problems. In addition, the implicature of pattern 1 of AVOID identified in the PDEV is

Human or Institution takes action to prevent Eventuality from occurring. However, if this

implicature were to be anchored to the set of instances in Figure 5.6, it would sound peculiar,

as the lexical items that fill the subject slots lack one key characteristic of human, that is,

taking action. No instances of this divergent pattern were found in the BNC data.

5.5.4 ALLOW

As seen in Table 5.4, the total raw frequency of ALLOW identified in ICLE-Chinese is 203.

Thirty-five (17.24%) instances were tagged as divergent usages of *allow*; of the 35 instances,

12 were regarded as systematic.

Table 5.5 below presents two systematic divergent patterns of ALLOW identified in ICLE-

Chinese.

Table 5.5 Divergent verb patterns of ALLOW found in ICLE-Chinese corpus but not in

PDEV

Pattern number	Pattern and implicature	Raw frequency	RF BNC
1	**Pattern**: Human 1 or Eventuality allows Human 2 infinitive (without *to*)	6	**6**
	Implicature: Human 1 or Eventuality gives Human 2 permission to do activity		

	Example: it is a place that allows customers chat with each other			
2	**Pattern**: Building = restaurant is (not) allowed *to-*infinitive = smoke	6	0	
	Implicature: Human is not given permission to do activity = smoke.			
	Example: Many high-class restaurants in Hong Kong is not allowed to smoke			

Six divergent instances of ALLOW were tagged as pattern 1, as presented in Figure 5.7 below:

Figure 5.7 Set of divergent patterns of ALLOW identified in ICLE-Chinese but not found in PDEV

1	clothes **shops**, convenient stores,	allow	**customers** *use* a credit card
2	the internet. It is a **place** that	allows	**customers** *chat* with each
3	traditional **TV game** which only	allow	**Teenage** *meet* other friends
4	general function of cyber cafe and	allow	**people** *play* online game.
5	the law in TV or radio and did not	allow	these **companies** *become* the sport's
6	, the **restaurants** was really not	allowing	**them** *smoke*, they spent little

As shown, the divergent pattern of ALLOW is *Human 1 or Eventuality allows Human 2 infinitive* (verb without *to*). The implicature of the set of all instances together is interpreted as *Human 1 or Eventuality gives Human 2 permission to do activity*. Surprisingly, an investigation of the BNC revealed that this pattern occurred six times there, as shown in Figure 5.8 below:

Figure 5.8 Occurrences of ALLOW in the BNC

1	canvas' on to which gliffs can be 'painted' will	allow	**developers** create things like two-dimensional
2	ICL will also introduce Access Manager 200, which	allows	**users** secure single point of login

3	assess the effects of changing those structures by	allowing	**them** <u>make</u> such changes directly through
4	Our data do not	allow	**us** <u>determine</u> whether HBIG is beneficial
5	At the same time, he determined never to	allow	**anyone** <u>develop</u> power independent of the
6	I can't see him	allowing	**me** <u>take</u> the children so far away from

All instances constitute the same pattern and implicature as that in the BNC data. However, consulting a native-speaker rater on whether these two instances could be correct uses of ALLOW confirmed that these instances are incorrect. This raises the question of the consistency of the BNC data, in that, apparently not all instances occurring in the BNC are always correct; there are cases where uses might be incorrect, and a further step, such as referring to a human native-speaker rater, thus needs to be taken to verify them.

Another set of divergent usages of ALLOW identified in ICLE-Chinese is presented in Figure 5.9 below:

Figure 5.9 Set of divergent usages of ALLOW identified in ICLE-Chinese

1	Many high-class **restaurants** in Hong Kong is not	allowed	to **smoke**. This can attract more
2	**restaurants** from smoking leaving the rest of them to be	allowed	to **smoke** so that people can choose.
3	dining out if the **restaurant** is not	allowed	to **smoke**., stated that The
4	If all **restaurants** are not	allowed	to **give up smoking** in restaurant
5	casual. If the **restaurants** is not	allowed	to **smoke**, they may rather

As can be seen, although all five instances are grammatically accurate (passive voice), they sound odd. Here, the learners treated a *restaurant* as a *human*, which is wrong. The process of smoking (i.e. sucking the smoke from a cigarette or pipe into one's mouth and blowing it out again) is limited to human beings. The implicature of this pattern is thus *Building =*

restaurant is not given permission to do activity = smoke. No such pattern was found in the

BNC.

5.5.5   ENCOURAGE

The total raw frequency of ENCOURAGE identified in ICLE-Chinese is 181. Of the 181

instances, 9 (4.97%) instances were identified as divergent usages of *encourage*. Of the 9

instances, 7 were considered as systematic, as they fulfilled the criteria of systematicity set

out by the researcher. Table 5.6 below presents the one systematic divergent pattern of

encourage identified in the ICLE-Chinese corpus.

Table 5.6 Divergent verb pattern of ENCOURAGE found in the ICLE-Chinese corpus

but not in the PDEV

Pattern number	Pattern and implicature	Raw frequency	RF BNC
1	**Pattern**: Human or Eventuality encourages Human infinitive	7	0
	Implicature: Human 1 or Eventuality has the effect of causing Human 2 to want to do something		
	Example: Actually, gambling encourages the youth want to get advantages		

Seven divergent instances of ENCOURAGE were tagged as pattern 1 of ENCOURAGE in

ICLE-Chinese, as shown in Figure 5.10 below.

Figure 5.10 Divergent patterns of ENCOURAGE identified in ICLE-Chinese but not

found in PDEV

1	enlighten the grade school child,	encouraging	them **participate** with the environmenta
2	there was the need for attracting and	encouraging	people with talent **come** to Hong Kong.

3	of Mainland Professional Scheme to	encourage	professionals from Mainland China **work** in Kong
4	tend to offer high credit limits to	encourage	spending students **consider** loweri
5	soccer betting. Actually, **gambling**	encourages	the youth **want** to get adventages withou
6	of abortion a women and the **abortion**	encourages	the teenagers **have** a more liberal attit
7	reasons for opposing it is that **it** would	encourage	youth **involved** in gambling. In fact, footba

The divergent pattern of ENCOURAGE here is *Human or Eventuality encourages Human infinitive* (verb without *to*). The implicature of the set of all these instances together can be construed as *Human 1 or Eventuality has the effect of causing Human 2 to want to do something.* No instances of this pattern were found in the BNC.

5.5.6 AGREE

The total raw frequency of AGREE in the ICLE-Chinese is 209. Of those 209 instances, 40 (19.14%) instances were identified as divergent usages of AGREE. Of the 40 instances, 26 were regarded as systematic divergences.

Table 5.7 below presents three systematic divergent patterns of AGREE identified in the ICLE-Chinese corpus.

Table 5.7 Divergent verb patterns for AGREE found in ICLE-Chinese but not in PDEV

Pattern number	Pattern and implicature	Raw frequency	RF BNC
1	**Pattern:** Human agrees *to*-infinitive	12	0
	Implicature: Human indicates that he or she has a favourable attitude towards a suggested proposition that is proposed by another human.		
	Example: In conclusion, I agree to construct a second railway link.		
2	**Pattern**: Human 1 agrees Human 2 *to*-infinitive (to	5	0

followed by verb).

Implicature: Human 1 indicates that he or she has a favourable attitude towards an activity that is carried out by Human 2.

Example: For my opinion, I agree students to own credit cards.

3	**Pattern**: Human agrees Eventuality = abortion.	9	0

Implicature: Human indicates that he or she has a favorable attitude towards Eventuality = abortion.

Example: In my opinion, I don't agree abortion in most of the cases.

Twelve instances of divergent usage of AGREE were identified in ICLE-Chinese (Figure 5.11).

Figure 5.11 Divergent usage of AGREE identified in ICLE-Chinese but not found in PDEV (pattern 1).

1	and cons in each aspect, **I** don't	agree	**to construct** a second railway in this
2	an economic benefits. In conclusion, **I**	agree	**to construct** a second railway link to
3	would be much more better. Anyway, **I**	agree	**to construct** a second railway link.
4	mentally feeling. In my opinion, **I** am not	agree	**to build** a second railway in Hong Kong
5	its advantages and disadvantage. For me, **I**	agree	**to ban** smoking in restaurants as I realize
6	Hong Kong. However, in my view, **I** totally	agree	**to use** the recycling as a method of
7	may increase. In my opinion, **I** do not	agree	**to ban** smoking in restaurants. It is
8	in restaurants. From my point of view, **I**	agree	to ban smoking. Smoking is really a bad
9	play and bet on them. So, **I**	agree	**to legalize** soccer betting in Hong Kong.
10	will complaint the government. **I** don't	agree	**to ban** smoking in restaurants and it is not
11	affect others' health. However, **I** do not	agree	**to ban** smoking in bars because smoking
12	smoke-free establishments. In my opinion, **I**	agree	**to do** a favour by banning smoking in

The set of all such instances together constitutes the pattern *Human agrees* to-*infinitive*. Interestingly, an examination of the instances revealed that this divergent usage is grammatically the same as pattern 4 in the original PDEV entry for AGREE (*Human agrees* to-*infinitive*). However, the implicature of all instances in Figure 5.11 together is different from that of pattern 4 in the PDEV entry for AGREE. The implicature of pattern 4 of AGREE recorded in the PDEV entry is *Human indicates that he, she, or it is willing to undertake V*, whereas the implicature of all instances in Figure 5.11 above is *Human indicates that he or she has a favourable attitude towards a suggested activity carried out by another human.* That is, the former is construed to indicate that Human himself or herself agrees to undertake the activity, whereas the latter does not mean that Human is undertaking any activity.

The following figure presents the set representing divergent pattern 2 of AGREE identified in ICLE-Chinese:

Figure 5.12 Divergent pattern of AGREE identified in ICLE-Chinese but not found in PDEV (pattern 2)

1	teenagers under 16. My conclusion is **I**	agree	**people** <u>to</u> *get* abortion because
2	to their part-time jobs. For my opinion, **I**	agree	**students** <u>to</u> *own* and use credit cards
3	that of the smokers. **I**, therefore, would	agree	the **government** <u>to</u> *ban* smoking in
4	development and their future. No **one** will	agree	their **offspring** <u>to</u> *spend* a large
5	adverse effect on people. In my opinion, **I**	agree	the cyber cafe for **people** <u>to</u> *relax*

The divergent pattern of AGREE in all instances in Figure 5.12 is *Human 1 agrees Human 2* to-*infinitive*. The implicature would be that Human 1 indicates that he or she has a favorable attitude towards an activity intended by Human 2. No such pattern was found in the BNC.

Another divergent usage of AGREE identified in ICLE-Chinese but not found in the PDEV is shown in Figure 5.13 below:

Figure 5.13 Divergent usage of AGREE identified in ICLE-Chinese but not found in PDEV (pattern 3)

1	most of Hong Kong **women** are not	agree	**abortion**, however, because of
2	**woman** in Hong Kong almost does not	agree	**abortion**. But in fact, abortion is
3	of the middle class **women** in Cosmo	agree	**abortion**. They believe that they
4	the pros and cons of abortion. **People**	agree	**abortion** because they think it is
5	. On the other hands, **others** do not	agree	**abortion** since it may harm the
6	do not agree at that. Those **people** who	agree	**abortion** are called pre-choice, it
7	to , most middle class **women** in Cosmo	agree	**abortion** as they feel abortion is
8	and right of living. In my opinion, **I** do not	agree	**abortion** in most of the cases. I think
9	of abortion is rise. In my opinion, **I** do not	agree	**abortion** because I feel that pregnancies

Figure 5.13 includes the annotated divergent instances of pattern 3 in Table 5.7. The set of instances shows that the pattern of AGREE here is *Human agrees Eventuality = abortion*. The BNC was searched for AGREE followed by a noun group. There were a large number of instances where *agree* is followed by words such as *terms*, *conditions*, *plans*, *policies*, etc.; however, in instances where AGREE is followed by the word *abortion*, only one instance of *abortion* was used with *agree with*, namely, *I agree with abortion*. Similarly, an investigation of LOCNESS revealed one occurrence of *agree with* followed by *abortion*: *...though I do not totally agree with abortion....* The data therefore suggest that all instances in Figure 5.13 taken together diverge from the usage of native speakers.

5.5.7 LEAD

As a reminder, the total raw frequency of LEAD as identified in ICLE-Chinese is 483. Of the 483 tokens, 44 (9.11%) instances were regarded as systematic divergent usages of LEAD.

Table 5.8 below shows four systematic divergent patterns of LEAD identified in the ICLE-Chinese corpus but not found in the PDEV.

Table 5.8 Divergent verb patterns for LEAD found in ICLE-Chinese but not in PDEV

Pattern number	Pattern and implicature	Raw frequency ICL-Chinese	Raw Frequency BNC
1	**Pattern**: Eventuality (Modality) 1 leads Eventuality 2.	19	0
	Implicature: Eventuality 1 is the cause of Eventuality 2.		
	Example: It will lead a serious problem to the local university.		
2	**Pattern**: Eventuality 1 leads Eventuality 2 adjective.	5	0
	Implicature: Eventuality 1 makes Eventuality 2 to be more likely [adjective].		
	Example: It leads the teenagers' life more colourful.		
3	**Pattern**: Eventuality leads Human or Institution infinitive (without *to*)	9	0
	Implicature: Eventuality makes Human do activity.		
	Example: The enough skilled labour force can lead Hong Kong become a global financial.		
4	**Pattern**: Eventuality leads *to*-infinitive.	11	4
	Implicature: Eventuality is the cause of activity.		
	Example: Importing professionals, however, may lead to worsen the economic situation in Hong Kong.		

Eighteen divergent instances were tagged as pattern 1 of LEAD in ICLE-Chinese, as shown in Figure 5.14 below:

Figure 5.14 Divergent usage of LEAD identified in ICLE-Chinese but not found in PDEV

1	, but some people think that it must	lead	a **bad effect** on the society. This
2	is not without disadvantages. It will	lead	a **serious problem** to the local university
3	propose to ban smoking in restaurants. It	leads	a lot of **debates** in this topic.
4	football gambling. It is inefficient. It will	lead	a **gambling** atmosphere which is **bad** to
5	policy. The most important point, it may	lead	a many **corruption** between the gambling
6	pay taxes, the whole royal family is	leading	a **parasitic** life, doing nothing but spend
7	of abortion is uterus perforation, it may	leads	**death**. In Hong Kong, the number of
8	was concluded that smoke-free bars will	lead	improvement of **respiratory** health. At the
9	, the power in the market is low	leading	**low profit**. Then businessmen are unwilling
10	this on the wrong way, it will	lead	many **problem**. In the future use computer
11	a lot of money. Second, it may	lead	many job **losses**, according to the research
12	passive smoking. Passive smoking can	lead	many kinds of respiratory **diseases** and heart
13	many people strongly disagree. This railway	leads	**serious** environment **impact** to Long Valley.
14	commonly believe that the debt problem will	lead	some physical **problem** and even the
15	decreasing the revenue of catering sector and	leading	the job **losses**. Evaluating the evidence le
16	student fall into debt. Getting in debt	lead	the other **problem**. Students need to find
17	tobacco and harmful chemicals. These	lead	the higher chance of getting lung **cancer**
18	Hong Kong. The disadvantage is that they	lead	the increase of **unemployment** rate.

As shown, eighteen divergent instances of LEAD were identified in ICLE-Chinese but not found in PDEV under the pattern *Eventuality 1 leads to Eventuality 2*. The implicature of the pattern would be *Eventuality 1 is the cause of Eventuality 2,* as shown in Table 5.8. An examination of the lexical items that fill the object slot *Eventuality 2* in this set of all instances together shows that they connote a bad event (e.g. a problem, corruption, death). This suggests that what the learners probably intended to use was the verb *cause* instead of *lead*. In

addition, several corpus studies revealed that most uses of *cause* are linked to a negative/bad meaning.

Two non-divergent usages can thus be suggested as alternatives to the divergent instances produced by the learners in Figure 5.14. In them, the verb *lead* (the node word) in all the instances presented above can be substituted by the verb *cause*, or *lead* followed by preposition *to* (*lead to*).

Another divergent pattern of LEAD identified in ICLE-Chinese but not found in the PDEV is *Eventuality 1 lead Eventuality 2 adjective,* as shown in Figure 5.15 below:

Figure 5.15 Divergent usage of LEAD identified in ICLE-Chinese corpus but not found in PDEV (pattern 2)

1	information in cyber cafes says. It	leads	a teenager's life more **colourful**
2	cafes can really improve our life and	lead	it more **colourful** and **meaningful** way.
3	problem of lacking skilful workers and	lead	the company more **competitive**. It also
4	during the economic downturn, it may	lead	the situation in Hong Kong **worse** and
5	Import of specialists from the mainland may	lead	the jobless rate **higher**. The next

This set of instances is linked to pattern 2 in Table 5.8. The set of all instances constitutes the pattern *Eventuality 1 leads Eventuality 2 adjective*. The implicature of the pattern would be *Eventuality 1 makes Eventuality 2 more likely adjective*. The search of the BNC showed no occurrence of such a pattern. Non-divergent alternatives can be suggested for the instances produced by the learners in Figure 5.15. These alternatives are shown below:

Figure 5.16 Non-divergent alternatives of instances of LEAD produced by Chinese

learners

1	in cyber cafes says. It	makes	a teenager's life more **colourful**
2	cafes can really improve our life and	make	it more **colourful** and **meaningful** way.
3	problem of lacking skilful workers and	make	the company more **competitive**. It also
4	during the economic downturn, it may	make	the situation in Hong Kong **worse** and
5	Import of specialists from the mainland may	make	the jobless rate **higher**. The next

Figure 5.16 shows alternatives in which the verb *lead* has been replaced by *make*.

The third divergent pattern of LEAD identified in ICLE-Chinese is presented in Figure 5.17 below:

Figure 5.17 Divergent usage of LEAD identified in ICLE-Chinese but not found in PDEV (pattern 3)

1	economic downturn, such a banning will	lead	fewer smokers **dine** out. This will
2	, The enough skilled labour force can	lead	Hong Kong **become** a global financial.
3	import of professionals from Mainland can	lead	local professionals **adapt** this challenging
4	could not. Evaluating the pros and cons	leads	me **have** a conclusion that the operation
5	. Interest rates are, however, very high. This	leads	students **get into** debt. To pay the
6	to protect the right of employees which	leads	the Hong Kong employers **fire** the local
7	their payment ability. All these elements	lead	the students **feel** some pressures and nervous.
8	card is useful, however, credit card can	lead	the students **become** materizatism. According
9	than that of cash. The credit cards	lead	the students **go out** with little cash.

This set of nine instances of LEAD constitutes the pattern *Eventuality leads Human infinitive* (without *to*), as shown in Table 5.8. Here, the implicature of the set of all instances together would be *Eventuality makes Human do activity*. The learners, I argue, replaced the node word *lead* with *make*.

The last set constituting a divergent pattern of LEAD identified in ICLE-Chinese but not found in PDEV is presented in Figure 5.18 below:

Figure 5.18 Divergent usage of LEAD identified in ICLE-Chinese but not found in PDEV

1	need to spend. As a result, it	leads	**to rise up** the tax payment of
2	heard objection to importing professionals	leads	**to increase** the unemployment rate. According
3	the Hong Kong's local workers. It	leads	**to increase** the unemployment rate and
4	the job. Although using credit card may	lead	**to affect** one's study, indicates that
5	Smoke- free establishments would then	lead	**to increase** the economy in Boston. In
6	If smoking is prohibited, this may	lead	**to reduce** the business of catering industry.
7	.Importing professionals, however, may	lead	**to worsen** the economic situation in Hong
8	Professional Scheme does not necessary	lead	**to increase** the unemployment rate in Hong
9	On the other hand, It can	lead	**to hurt** the local experts' prospects and
10	problem of the decline of population that	lead	**to weaken** strength, reduce the economy

This set of ten instances of LEAD is tagged as pattern 4 in Table 5.8. This pattern constitutes *Eventuality leads* to-*infinitive*. The implicature of the set of all instances together would be *Eventuality causes activity*. Surprisingly, an investigation of the BNC showed two instances of pattern 4, as shown in figure 5.19 below.

Figure 5.19 Usage of LEAD identified in the BNC but not found in the PDEV

| 1 | **some traces** might be found which might | lead | **to discover** some of Mr. Stone's accomplices. |
| 2 | **Implementation** of the plan and evaluation … which | lead | **to overcome** problems and then check. |

However, consulting a native-speaker rater whether these two instances are correct uses of LEAD confirmed that these instances are incorrect. This raises the question of consistency of

the BNC, in that not all instances occurring in the BNC are necessarily correct; there are cases where some instances might be incorrect uses, and a further step should be taken, that is, referring to a human native-speaker rater.

5.6 Discussion

5.6.1 Can PDEV identify divergent patterns of target verbs in ICLE-Chinese?

The research data reported in this chapter suggest that this question can be answered in the affirmative. The method of matching non-divergent patterns of target verbs identified in the learner corpus to those in the PDEV has successfully filtered divergent usages of the target verbs found in the learner corpus. These divergent uses are sorted into syntagmatic pattern groups, each of which comprises a distinctive meaning and pattern. In addition, the use of references such as *Grammar Patterns* and the BNC was also useful in making sure that all unconventional usages of the target verbs were divergent.

5.6.2 What do divergent patterns tell us about learners' competence and global English?

The investigation of all 16 target verbs identified in the ICLE-Chinese corpus (see Table 5.2), shows that the Chinese learners produce non-divergent (correct) patterns most of the time, and that when they produce divergent (incorrect) patterns, they still produce them in sensible ways, that is, they still produce patterns that are recognisable. Furthermore, Chinese learners often adopt verb patterns by analogy with another verb; for instance, one of the sets of divergent instances of ALLOW produced by the learners, comprises the pattern *Human or Eventuality allows Human infinitive* (see Table 5.5 and Figure 5.8), as in for example *It is a place that allows customers chat with each other*. This pattern belongs to the verb *let*, as in *It*

120

is a place that lets customers chat with each other. What is involved here is simply that the

learners divergently substitute *allow* for *let.*

Second, the numbers of divergent patterns of the target verbs, as produced by the learners, are

not the same across the verbs (see Table 5.4–5.9). A summary of the numbers of sets of

divergent patterns for each verb is shown in Table 5.9 below:

Table 5.9 Number of sets of divergent patterns identified in ICLE-Chinese

No.	Verbs	No. of sets of divergent patterns
1	AGREE	3
2	ALLOW	2
3	AVOID	4
4	ENCOURAGE	1
5	LEAD	4

The variation in these numbers may suggest the degree of difficulty associated with each

verb. For instance, AVOID and LEAD show the highest number of sets of divergent patterns

in ICLE-Chinese, four each, while only one set of divergent data for ENCOURAGE was

produced by the Chinese learners. This might indicate that Chinese learners may struggle with

producing correct patterns for AVOID and LEAD more than they do with ENCOURAGE, for

instance.

Another possible explanation for these variations might be due to features of English

influenced by learners' first language, features of global English, or language change in native

English. The last of these possibilities means that divergent patterns might represent usages

that could potentially be used in the future by all or most English users. This might add to the neutral deviations in some linguistic patterns between ESL or EFL and native speakers that we have seen several studies report exist in section 5.3, which might add to (but not account for all of) L1-influenced or global-English-influenced divergence; although Kachru's model is no longer accepted, it still at least establishes that there are varieties of English usage that exist between non-native and native norms. With all these in mind, the emergence of all the divergent patterns of verbs identified in ICLE-Chinese in this study could potentially be explained in relation to such factors.

5.7   Conclusion

The major aim of this chapter was to discover whether the concepts in PDEV could be used to successfully identify divergent patterns of target verbs identified in ICLE-Chinese. Some terminological and conceptual issues around constructs such as error, innovation, and random and systematic error have been discussed. In addition, the dichotomy between ESL and EFL has been discussed. Next, the method of applying the concepts in PDEV to the ICLE-Chinese corpus has been illustrated, and the data considered. Since the aim of this thesis is to demonstrate how CPA/PDEV can be used to successfully describe learner corpora, this chapter has mainly focused on establishing the ability of PDEV to identify divergent patterns of certain target verbs in ICLE-Chinese.

Some of the key findings this chapter has revealed are that Chinese learners produce non-divergent patterns (correct patterns) most of the time and that when they produce non-divergent patterns (incorrect patterns), they produce them in a sensible way, in that they use patterns that are recognisable and largely interpretable. It has been speculated that the reason the learners produced such divergent patterns could be related to the features of global

English and/or L1 interference and/or the effects of language change. Another key finding was that the proportions of divergent patterns differ across the target verbs. Possible explanations could include the degree of difficulty that each of the target verbs may have, that is, the data may reflect and reveal their difficulty for Chinese learners.

Finally, PDEV entries for divergent patterns of every target verb, as identified in ICLE-Chinese, have been established following PDEV conventions. These entries should be useful for both learners and teachers, for example for use in classroom exercises. However, this chapter is already too lengthy to include the pedagogical implications, which will instead be considered in detail in Chapter 9.

Chapter 6   **Focus on the ICLE-Swedish Corpus (Analysis of ALLOW)**

6.1   Introduction

The previous chapter focused on identifying the divergent verb patterns found in the ICLE-Chinese corpus, that is, incorrect verb patterns of some target verbs that were not matched to those in PDEV. This was done by employing the CPA method.

This chapter focuses on conducting a comparison between ICLE-Swedish and LOCNESS on verb patterns of the verb ALLOW, because ALLOW is the most noticeably different of the verbs investigated. The reason for selecting the Swedish sub-corpus for this study is that it encourages a discussion of the concepts of over- and under-representation rather than the concept of 'divergence'. The quantitative analysis of the 16 verbs in Chapter 5 showed that for the most part errors in ICLE-Swedish are simple typographical errors, suggesting that a study of divergence would be unproductive. The aim of this chapter is thus to complement the study of divergence in the Chinese data in Chapter 5 with a study of non-divergent difference between ICLE-Swedish and LOCNESS.

The chapter starts by presenting the overall findings on the target verbs in the Swedish data. Some justifications are provided for the selection of the specific verb ALLOW for investigation. Next, a statistical presentation of the verb patterns of ALLOW as identified in the two target corpora is done. This is followed by a detailed investigation of the concordance lines of each verb pattern of ALLOW. The chapter concludes with a discussion section in which the research questions are answered. Given that, the research questions for this chapter are as follows:

Question 1: What are the similarities and disparities between the ICLE-Swedish writers and the LOCNESS writers in terms of the verb patterns of ALLOW?

Question 2: To what extent do the Swedish learners (ICLE-Swedish) and the native-speaker writers (LOCNESS) respectively make full use of the range of the patterns of ALLOW?

Question 3: Has the PDEV entry for ALLOW allowed the researcher to successfully identify the learners' usages of this verb with sufficient specificity, and how does that compare with other methods of coding?

6.2   Overall target verbs

This section re-presents the findings on the target verbs discussed in Chapter 4. The quantitative analysis of the selected 16 verbs showed quite minor errors in ICLE-Swedish that were not worthy of mention, most of them related to typos. Therefore, for this chapter, I chose instead to look at the under- and over-representation of verb patterns in learners' usage of ALLOW in comparison with the native data in LOCNESS. In addition, provided that Swedish English is seen to be better (more correct) than Chinese English on the whole, the aim of this comparison will be to see how individual verbs are used by the ICLE-Swedish learners and how the learner English in ICLE-Swedish is similar to or different from that of native speakers of English in LOCNESS.

As a reminder, the target verbs to be investigated in this thesis were selected according two criteria: first, they had to be the verbs that already had a complete PDEV entry; second, the verbs selected had to occur relatively frequently in both target corpora (learner and native-speaker; that is, the verbs selected differed for the Swedish and the Chinese data). Table 6.1 below presents the target verbs.

Table 6.1 Verbs (lemmas) identified in both corpora

No.	Verb	ICLE-Swedish		LOCNESS		Over-/under-represented	Log	Ratio
		RF	NL	RF	NL			
1	ADMIT	20	9.99	40	12.4	-	0.63	2.41
2	ADVISE	3	1.49	9	2.79	-	0.95	1.3
3	AFFORD	35	17.49	39	12.09	+	2.49	5.4
4	AGREE	42	20.99	80	24.8	-	0.78	3.81
5	**ALLOW**	**74**	**36.99**	**270**	**83.73**	**-**	**44.51**	**46.74**
6	APPEAR	28	13.99	85	26.35	-	9.28	12.36
7	ASK	64	31.99	113	35.04	-	0.34	3.05
8	AVOID	33	16.49	37	11.47	+	2.27	5.02
9	CLAIM	52	25.99	90	27.91	-	0.17	1.92
10	DENY	21	10.49	53	16.43	-	3.2	5.94
11	ENCOURAGE	34	16.99	77	23.87	-	2.84	6.88
12	MAINTAIN	18	8.99	40	12.4	-	1.33	3.41
13	PLAN	16	7.99	18	5.58	+	1.08	2.41
14	PROPOSE	40	31.99	30	9.25	+	10.31	2.16
15	TELL	61	43.49	148	45.89	-	0.16	2.4
16	URGE	10	4.99	15	4.62	+	0.04	0.11

Boldface = subject to further analysis.

As can be seen, the quantitative investigation of the selected target verbs (lemmas) revealed that the normalized frequencies of most of these verbs in ICLE-Swedish are less than those in LOCNESS. Only 5 verbs out of the 16 were found to be statistically significantly over-

represented in ICLE-Swedish relative to the LOCNESS, as shown in Table 6.1 (Significance was identified using the on-line log-likelihood calculator provided by Lancaster University. Verbs whose log-likelihood was less than 5.92 were eliminated.)

It is beyond the scope of this thesis to investigate all the target verbs in Table 6.1. Elimination of some verbs was an ideal alternative. Ultimately, the investigation was restricted to one verb, that is, ALLOW, for a number of reasons. Unlike the other verbs, ALLOW contains patterns that are neither too many nor too few; in fact, it presents eight patterns, as recorded in the PDEV entry. Second, as shown in Table 6.1, the raw frequencies of ALLOW are the highest in both ICLE-Swedish and LOCNESS. Thus, ALLOW was seen as a good verb to investigate.

The following section presents and discusses the basic findings on ALLOW in ICLE-Swedish.

6.2.1 ALLOW

As indicated by the log-likelihood data in Table 6.1 (see section 6.2), ALLOW is under-represented in ICLE-Swedish (36.99%) relative to LOCNESS (83.73%), with an LL value of 44.51, which is significant at 99.99% (p <0.0001). To see how similar or different the use of ALLOW was across both corpora, the patterns found in both corpora were matched to those of the PDEV entry for ALLOW. Table 6.2 below presents the PDEV entry for ALLOW[20].

---

[20] Adopted from the original PDEV at http://pdev.org.uk.

Table 6.2 PDEV entry for *allow*

Pattern no.	Pattern and implicature	% BNC sample (250 lines)
1	**Pattern**: Human 1 or Institution 1 or Eventuality allows Human 2 or Institution 2 or Self *to*-infinitive **Implicature**: Human 1 or Institution 1 or Eventuality gives Human 2 or Institution 2 or Self the opportunity, time, or permission *to*-infinitive **Example**: the government allows certain individuals affected by the wrong to the public interest to bring action to redress the wrong	54.8%
2	**Pattern**: Human 1 or Institution 1 or Eventuality allows Institution 2 or Self or Human 2 Privilege **Implicature**: Human 1 or Institution 1 or Eventuality gives Human 2 or Institution 2 or Self permission or the opportunity to have Privilege **Example**: Another great spiritual advantage was that the people's occupation allowed them the opportunity to read good books or to engage in godly conversation	4.4%
3	**Pattern**: Human or Eventuality or Institution allows Physical Object *to*-infinitive **Implicature**: Human or Institution or Eventuality causes or provides the opportunity for Physical Object to be V-ED **Example**: The banking service allows funds to be deposited in numerous foreign currencies	7.6%
4	**Pattern**: Human or Eventuality 1 or Institution allows Eventuality 2 **Implicature**: Human or Institution or Eventuality 1 gives the opportunity, time, or permission for Eventuality 2 to occur **Example**: there are a number of people who, though fully Christian, have never allowed justification by faith to be an experiential reality in their life	17.2%
5	**Pattern**: Human 1 or Institution 1 allows Human 2 or Animal Direction **Implicature**: Human 1 or Institution 1 gives permission for Human 2 or Animal to move in Direction **Example**: A written request for a brief photocall from the media, who were not allowed up the 100-yard driveway to the hotel, was conveyed by a garda to Mr Keenan	2.0%
6	**Pattern**: Human allows Resource (for Activity) **Implicature**: When planning, Human ensures that a sufficient amount of Resource is made available to complete Activity **Example**: allow plenty of time in your total schedule for the physical typing and reproduction of your report	2.8%
7	**Pattern**: Human or Eventuality 1 allows for Eventuality 2 **Implicature**: Human *or* Eventuality 1 makes provision for Eventuality 2 **Example**: Human rights organizations opposed the emergency powers, which allowed for indefinite detention without trial	8.0%
8	**Pattern**: Human allows that-clause	1.6%

	Implicature: Formal. Human admits the truth of that-clause **Example**: the Levelers did allow that in practice not all men would be enfranchised by their proposals.	

Table 6.2 shows eight patterns of ALLOW. Column 1 indicates the pattern number in the PDEV entry for the verb ALLOW. Column 2 contains a description of the pattern (i.e. pattern, implicature, and an example). Column 3 indicates the proportion rate of each pattern in the BNC, as provided in the original PDEV entry.

Table 6.3, below, shows the findings on the verb patterns of ALLOW found in ICLE-Swedish and in LOCNESS according to the PDEV entry convention.

Table 6.3 Log-likelihood values of verb patterns of ALLOW

	ICLE-Swedish		LOCNESS			
	RF	NL	RF	NL	Over-/under-representation	Log-likelihood
pattern 1	52	25.99	149	46.20	-	**13.85**
pattern 2	0	0	14	4.34	-	**13.51**
pattern 3	0	0	7	2.17	-	**6.76**
pattern 4	3	1.49	66	20.46	-	**44.79**
pattern 5	0	0	0	0		
pattern 6	0	0	2	0.62	-	1.93
pattern 7	2	0.99	15	4.65	-	**6**
pattern 8	0	0	0	0		

Terms: RF = raw frequency; NL = Normalized frequency, per 100,000.

Log-likelihood critical values of 3.84 or higher are significant at the level of $p < 0.05$; those above 6.63 are significant at $p < 0.01$; those above 10.83 are significant at $p < 0.001$. If a p value is < 0.05, one can be 95% positive that the differences are not due to random chance; if

a p value is <0.01, one can be 99% positive that the differences are not random; if a p value is < 0.001, one can be 99.9% positive that the differences are not random; and if a p value is < 0.0001, one can be 99.99% positive that the differences are not due to random chance.

Of the 8 patterns recorded in the original PDEV entry for ALLOW, 6 patterns were found in at least one of the two corpora, as shown in Table 6.3. The two patterns that were not found are patterns 5 and 8. Since these two patterns were not used by either the learners or the native-speaker students, there is a possibility that it was genre influence that made them absent in both corpora. On the other hand, there was a significant difference between the Swedish learners and the native-speaker students in each of the other patterns. What is striking about the numbers in this table is that all the verb patterns of ALLOW apart from pattern 6 were statistically significantly under-represented in the ICLE-Swedish corpus in comparison with the LOCNESS. Table 6.4, below, re-presents the patterns in decreasing log-likelihood values.

Table 6.4 Log-likelihood values of verb patterns of ALLOW

	ICLE-Swedish		LOCNESS			
	RF.	NL.	RF.	NL.	Over-/under- representation	Log-likelihood
Pattern 4	3	1.49	66	20.46	-	**44.79**
Pattern 1	52	25.99	149	46.20	-	**13.85**
Pattern 2	0	0	14	4.34	-	**13.51**
Pattern 3	0	0	7	2.17	-	**6.76**
Pattern 7	2	0.99	15	4.65	-	**6**

Terms: RF = raw frequency; NL = Normalized frequency, per 100,000.

Table 6.4 is a reproduction of Table 6.3 above but with some data eliminated. As can be seen, pattern 6 has been eliminated, as the log-likelihood value was below 3.84, that is, there was no significant difference. In addition, patterns 5 and 8 have been excluded, due to their absence in both corpora. Detailed investigations of each pattern in Table 6.4 are presented in the following sections.

6.2.2 Pattern 4

As indicated in Table 6.4, pattern 4 is under-represented in ICLE-Swedish (1.49%) relative to the native-speaker students in LOCNESS (20.46%), with an LL value of 44.79, which is significant at 99.99% (p <0.0001). The description of pattern 4 in the PDEV is *Human or Institution or Eventuality 1 gives the opportunity, time, or permission for Eventuality 2 to occur.*

Two important observations about this pattern should be made. One is that this pattern lacks the grammatical component of the *to*-infinitive. The second is that in this pattern, ALLOW is followed directly by Eventuality.

To see how this pattern was used by the Swedish learners and the native speakers, a detailed investigation of the concordance lines of pattern 4 in the two corpora was carried out. Only 3 occurrences of this pattern were found in the ICLE-Swedish, whereas the native-speaker students used it 66 times. According to the PDEV description of pattern 4 (see section 6.2.1, Table 6.2), pattern 4 is the only one that is immediately followed by the semantic type *Eventuality 2*. The following example has been taken from LOCNESS:

131

6.1 *The power granted under article twelve to dissolve parliament would, according to Michel Debr, only allow **a short exchange with the public***. (LOCNESS)

Here the pattern is *Eventuality 1 allows Eventuality 2*. A detailed investigation of the 66 concordance lines of pattern 4 in the LOCNESS revealed that *Eventuality 2* could be a phrase of several words, whereas in the ICLE-Swedish *Eventuality* 2 was limited to one word, as shown in bold in the following quote: …*by telling the government not to allow **logging, mining or hunting***.

6.2.3   Pattern 1

Table 6.4 shows that pattern 1 is under-represented in the ICLE-Swedish (25.99%) in comparison with the native-speaker students (46.20%), with an LL value of 13.85, which is significant at 99.9% (p <0.001).

According to the PDEV entry for *allow*, the description of pattern 1 is *Human 1 or Institution 1 or Eventuality allows Human 2 or Institution 2 or Self* to-*infinitive*; the implicature of pattern 1 is *Human 1 or Institution 1 or Eventuality gives Human 2 or Institution 2 or Self the opportunity, time, or permission* to-*infinitive*. It should be highlighted here that the convention to-*infinitive* provided in pattern 1 means a verb preceded by *to* such as *to study*, *to eat*, etc.

The investigation of the instances in LOCNESS revealed that there were many instances where the subject slots were mainly filled with the semantic type Eventuality. By contrast, Human and/or Institution were found to be much more frequent than the Eventuality in the subject slots in ICLE-Swedish. The following table demonstrates the proportions of the three semantic types that fill the subject slots of pattern 1 in the two corpora.

Table 6.5 Proportions of the three semantic types identified in ICLE-Swedish

Corpora	ICLE-Swedish			LOCNESS		
	RF.	%	NL.	RF.	%	NL.
Human/Institution	48	92.30	23.99	82	55.78	25.42
Eventuality	**4**	7.69	1.99	**65**	44.21	20.15

RF = raw frequency; % = proportion; NL = normalised frequency

The following are some uses of pattern 1 where *Eventuality* populates the subject slot:

6.2 ***Lowering the drinking age*** *would allow college society to get a grip on alcohol.*

(LOCNESS)

6.3 ***Changing technology*** *has allowed factories to use robots for production.*

(LOCNESS)

The subject slots in these two examples involve a linguistic phenomenon called 'nominalization'; this means that processes or attributes that would congruently be construed by verbs or adjectives are instead construed by nouns (Halliday 1993). In example 6.2, *lowering the drinking age*, *lowering* is a noun that is derived from the verb *lower*; similarly, in example 6.3, *changing technology*, *changing* is an adjective that is based on the verb *change*. If we were to rephrase examples 6.2 and 6.3 in order to see how nominalization works, the rephrased examples would be as follows:

6.4 *If the government were to lower the drinking age, college society would be able to get a grip on alcohol.*

6.5 *Technology has changed; therefore, factories use robots for productions.*

In example 6.6, the noun *lowering* becomes the verb *(were to) lower;* similarly, in example 6.7, the adjective *changing* becomes the verb *changed*. Employing words such as *lower* and *change* as verbs is simpler than having them as nouns or adjectives because they reflect what is happening more.

Another example that involves nominalization, taken from LOCNESS, is as follows:

> 6.6 *Allowing alcohol consumption at age eighteen would change the way America viewed alcohol use as a society.* (LOCNESS)

This example can be paraphrased to avoid nominalization, as follows:

> 6.7 *If the US government were to allow alcohol consumption at age eighteen, the way America viewed alcohol use as a society would be changed*.

In this example, the noun *allowing* becomes the verb *allows*.

Compared with the LOCNESS writers, the Swedish learners are more likely to use non-nominalised phrases. An example taken from ICLE-Swedish is provided below:

> 6.8 *If the government allows immigrants to keep their traditions and costumes, these immigrants sooner or later will be assimilated into the society as the children of the 1960's and 70's immigrants have been* (ICLE-Swedish)

This example shows that the subject slot is filled with the noun phrase *the government* followed by the verb *allow*. If we were to paraphrase this example to make the subject slot involve nominalization, it would look as follows:

> 6.9 *Allowing immigrants to keep their traditions and costumes will cause these*
>
> *immigrants to be assimilated into the society.*

The verb *allow* becomes the noun *allowing*.

I would argue that there are cases where the use of nominalization would be better than the use of non-nominalization, particularly cases in which the subject slot includes Human or Institusion (e.g. *if we/the government lower[s] the drinking age…*; *if the government allows immigrants….* If the speaker or writer intends to reflect what is actually happening, s(he) would likely be avoiding the use of nominalization, which involves expression of agency (i.e. the person who does the action). Example 6.7, for instance, *if the government lowers the drinking age…* includes a lot of information, that is, there are people involved in the process of lowering the drinking age and those people are the ones who work in the government; in addition, people are allowed to drink at a particular age and other people can change that age and make it higher or lower. All of this information is summarised in the nominalized phrase *lowering the drinking age,* as shown in in example 1.

Another phenomenon observed is that of metonymy, defined in this dissertation as "the substitution of a word referring to an attribute for" [a thing that is meant one thing stands in another thing] (*Collins English Dictionary*). This is similar to what Hanks (2013:17) refers to as 'semantic type alternation'. Hanks (2013:177) noted:

135

There is regular alternation in different kinds of contexts between lexical sets that
have different semantic types. For example, in the context of verbs denoting cognitive
actions (think, say, propose, negotiate, decide, implement, etc.), there is regular
alternation in the subject slot between Human and Human Institution.

To illustrate this idea, two made-up examples are provided below:

6.10 ***The Vice-Chancellor*** *allowed more international students to enter the*
university this year.

6.11 ***The university*** *allowed more international students to enter the university this*
year.

Here, the two semantic types alternate with no effect on the meaning of the two sentences:
Human = the Chancellor; Institution = the university.

Moreover, '[[Human Institution]] regularly alternates with [[Social Location]]: 'The
administration [[Human Institution]]' (ibid.:177). The following examples were made up to
illustrate:

6.12 ***The administration*** *passed a new law on gun control.*

6.13 ***New York*** *passed a new law on gun control.*

The two semantic types in the examples can be alternated with no effect on the meaning of
the sentences: Human Institution = the administration; Social location = New York. This is
the phenomenon of metonymy.

Going back to Table 6.5, it should be noted that there are a small number of instances in which a Human subject is used by Swedish learners although an Institution subject is felt to be more appropriate for use than a Human subject. The following are some instances taken from the ICLE-Swedish corpus:

6.14 *In the production-processes e.g. **we** allow our factories to let out a lot of pollution and we are increasing the damage on the ozon-layer by our use of Freon.* (ICLE-Swedish)

6.15 *There are those who say that **we** should not allow any immigration, and there are those who claim that immigration is good for Sweden.* (ICLE-Swedish)

6.16 *How could the **leaders** of a country allow this to happen? How could they conceal the truth to the people, or did they know.* (ICLE-Swedish)

In these examples, the subject slots of *allow* are filled with Humans. In examples 6.14 and 6.15, for instance, the learners use the pronoun *we*. The meaning of example 6.14 is that Swedish factories are allowed to let out a lot of pollution. The intended referent of the subject pronoun *we* is not ordinary people; rather, it is the Swedish government, which has the authority to allow or not allow factories to let out a lot of pollution. Similarly, the subject pronoun *we* in example 6.15 again refers to the government, which has the authority to allow or not allow immigrants to cross the border into Sweden. The same idea applies in example 6.16, in that the subject *leaders* alternates with the *government*, which allows chemical substances to be buried in the soil.

The learners who produced these examples refer to the government by using a Human subject (*we* and *leaders*), but it would be more accurate if an Institution subject, *the government*, was used to replace these Human words.

6.2.4   Pattern 2

As shown in Table 6.4, pattern 2 is under-represented in ICLE-Swedish (0%) relative to the native-speaker students (4.34%), with an LL value of 13.51, which is significant at 99.9% (p <0.001).

The description of pattern 2 in the PDEV is *Human 1 or Institution 1 or Eventuality allows Institution 2 or Self or Human 2 Privilege*; the implicature of pattern 2, according to the PDEV, is *Human 1 or Institution 1 or Eventuality gives Human 2 or Institution 2 Self permission or the opportunity to have Privilege.*

The important part of this pattern is the semantic type that fills object 2 of ALLOW (e.g. *allow people **access***). This is because the PDEV ontology includes a list of lexical items that belong to the 'Privilege' type, such as *opportunity*, *chance*, *authority*, and others; these words can be looked at if someone wants to see what Privilege is.

This pattern was not found at all in ICLE-Swedish, whereas there were 14 occurrences of this pattern found in LOCNESS. Interestingly, of these 14 instances, 8 (57.14%) instances were used in the passive voice and only 6 (42.85%) were used in the active voice. The following are some of the pattern 2 usages:

6.17 *I completely agree that before **a professor should be** allowed **domain over a***

classroom *at USC, he should be able to speak English well enough to teach his*

specific discipline to his students. (LOCNESS)

6.18 *Lowering the drinking age would **allow** eighteen-year old access to alcohol.*

(LOCNESS)

Although the difference is not significant, the higher rate of passive use by native speakers

might suggest that the native-speaker students preferred the use of the passive with pattern 2.

6.2.5 Pattern 3

Table 6.4 shows that pattern 3 is under-represented in ICLE-Swedish (0%) relative to the

native-speaker students (2.17%), with an LL value of 6.76, which is significant at 99% (p

<0.01).

According to the PDEV entry for ALLOW, the description of pattern 3 is *Human or*

Eventuality or Institution allows Physical Object to-*infinitive*; and its implicature is *Human or*

Institution or Eventuality causes or provides the opportunity for Physical Object to be V-ed[21]

(i.e., a passive meaning). To illustrate this, some examples are provided below from the

analysed samples in PDEV and LOCNESS:

6.19 *The banking service allows funds to be deposited in numerous foreign*

currencies[22] (PDEV)

---

[21] *To be V-ed* in PDEV means that verb is in its past participle form (e.g. *to be taken*, *to be delivered*, etc.)

[22] http://pdev.org.uk/conc.php?verb=allow&patnum=3&expl=both&ssize=.

6.20 *trying fertility treatment and some eggs are frozen should they be allowed to be implanted in the mother years later* (LOCNESS)

6.2.6 Pattern 7

Table 6.4 indicates that pattern 7 is under-represented in ICLE-Swedish (0.99%) relative to the native-speaker students (46.20%), with an LL value of 6, which is significant at 99% (p <0.01).

The description of pattern 7 in the PDEV is *Human or Eventuality 1 allows for Eventuality 2*; and its implicature is *Human or Eventuality 1 makes provision for Eventuality 2*. Some examples of this pattern are presented below:

6.21 *the constitution did allow for an extension of the role of the president granting him new powers* (LOCNESS)

6.22 *Time, unfortunately does not allow for a more lengthy and thorough discussion of the conflicts* (ICLE-Swedish)

6.3 Discussion

Based on the findings that are presented above, the three research questions for this chapter will be answered in the following sub-sections.

6.3.1 Question 1: What are the similarities and disparities between the ICLE-Swedish writers and the LOCNESS writers in terms of the verb patterns of ALLOW?

There are several points that can be drawn from the comparison between ICLE-Swedish and LOCNESS data. First, in terms of similarities, Table 6.3 (see section 6.2.1 for the table) shows that there exists a degree of similarity between the Swedish learners and the native

speakers in the use of verb patterns of ALLOW. Both groups of writers avoid the use of

patterns 5 and 8; pattern 8 is particularly old-fashioned—the common alternative verbs for

pattern 8 would be *admit* or *acknowledge*. The absence of these patterns in both corpora might

suggest genre differences between the two target corpora and the BNC; that is, the texts in

ICLE and LOCNESS are mainly argumentative essays, while the BNC includes various genre

types (e.g. fiction, novels, newspapers, among others). Second, Table 6.3 also indicates a

degree of disparity between corpora. Almost all the verb patterns of ALLOW are under-

represented in ICLE-Swedish because the verb ALLOW per se is less used by the Swedish

learners; thus, it is not surprising that all the verb patterns are under-represented. In that

context, the investigation of the concordance lines revealed a difference between the two

corpora in terms of the semantic types that populate the subject slots across all the patterns;

that is, the learners overuse the semantic type Human while they underuse Eventuality (see

section 6.2.4 for details).

6.3.2 Question 2: To what extent do the Swedish learners (ICLE-Swedish) and the native-speaker writers (LOCNESS) make full use of the range of the patterns of ALLOW?

The PDEV not only allowed me to explore how patterns of ALLOW were used in ICLE-

Swedish and LOCNESS, but also to see to what extent both groups of writers make full use of

the range of the patterns of ALLOW.

Table 6.6 below presents the verb patterns of ALLOW found in either of the two corpora:

Table 6.6 Verb patterns of ALLOW found in either ICLE-Swedish or LOCNESS

	ICLE-Swedish	LOCNESS
	Raw frequency	Raw frequency
pattern 1	52	149
pattern 2	0	14
pattern 3	0	7
pattern 4	3	66
pattern 5	0	0
pattern 6	0	2
pattern 7	2	15
pattern 8	0	0

As a reminder, column 1 indicates the full patterns that are recorded in the PDEV entry for the verb ALLOW; column 2 shows the frequency of each pattern in ICL-Swedish; and column 3 presents the raw frequency of each pattern identified in LOCNESS. Of the 8 patterns recorded in the PDEV, only 3 patterns are found in ICLE-Swedish: patterns 1, 4, and 2. By contrast, 6 of these patterns are used by the LOCNESS writers: patterns 1, 2, 3, 4, 6, and 7. Given that, it can be clearly seen that the Swedish learners used a narrower range of patterns of ALLOW (3 patterns) than that of the native writers (6 patterns). This is not surprising, as several studies have reported that one feature of learner language is the use of a narrower range of senses of multi-meaning verbs (cf. Altenberg and Granger 2001; Ringbom 1998). This study shows that such a phenomenon is not restricted to multi-meaning verbs, since *allow* is not a multi-meaning verb.

6.3.3 Question 3: Does the PDEV entry for ALLOW let the researcher identify learners' usage of the verb with sufficient specificity, and how does it compare to other methods of coding?

The answer to this question is divided into three subtopics: semantic type; implicature; and range of verb patterns.

6.3.3.1.1 Semantic types

As a reminder, semantic types are subparts of a hierarchical ontology. Each semantic type includes sets of lexical items that refer to the same kind of concept (Hanks 2013). This feature is significant for determining the disparity between ICLE-Swedish and LOCNESS. For example, the semantic types that populate the subject slot of pattern 1 of ALLOW in the original PDEV are Human, Institution, and Eventuality (see section 6.2.1 and Table 6.2). The comparison between ICLE-Swedish and LOCNESS shows that the Swedish learners underuse Eventuality (1.99 per 100,000 words), while it is used quite frequently by the native-speaker writers in LOCNESS (20.15 per 100,000 words). Other methods of coding are unlikely to catch such a difference.

Table 6.7 below shows the difference between two coding systems. The method developed by Francis et al. (1996) (see section 2.4.1) does not for instance take the subject slot into account. In fact, it instead focuses on words that come after verbs[23] (ibid.).

---

[23] In some cases, the subject slot is treated as part of the verb pattern. However, the subject is limited to introductory *it* and general *it* (ibid.). for example, '***It emerged that he had a violent criminal record***…' (introductory *it*). (Francis et al. 1996:519).'***It was coming on to rain*** *when finally Mac's lorry arrived'*. (general *it*) (Francis et al. 1996:551).

Table 6.7 The coding systems of the PDEV and *Grammar Patterns*

	Pattern 1
PDEV	<u>Human 1 or Institution 1 or Eventuality</u> <u>allows</u> **Human 2 or Institution 2 or Self** *to*-infinitive
GP1*	<u>V</u> *n* **to-inf**

*GP1= *Grammar Patterns 1: Verbs* (Francis et al. 1996)

The double-underlined words indicate the subject slot in the pattern; the single-underlined word represents the verb ALLOW; the words in bold represent the object; and the words in italics represent the *to*-clause followed by the verb.

Table 6.7 shows several differences. As can be seen, the subject slot is not included in Francis et al.'s (1996) convention, while it is present in the PDEV. Another difference is that of the object slot; in Francis et al. (1996) the object slot is indicated broadly by *n* (noun group), while in the PDEV the same slot is more specific to such types as Human, Institution, or Self.

At this point, it should be warned that the code V in Francis et al. (1996) is not restricted to the verb ALLOW; the V represents any verb that qualifies for this pattern as far as the Francis et al. pattern method is concerned.

6.3.3.1.2 Implicature

Implicature is equally important as semantic types to distinguish between different patterns. If we take patterns 1 and 3 of ALLOW as examples, as both of them have the same structure, as indicated in the PDEV (see section 6.2.1 and Table 6.2 for details). The following are the implicatures of these patterns:

1. Pattern 1: *Human 1 or Institution 1 or Eventuality allows Human 2 or Institution 2 or Self* to-*infinitive*.

   Implicature: *Human 1 or Institution 1 or Eventuality **gives** Human 2 or Institution 2 or Self the opportunity, time, or permission to/inf. [verb].* (PDEV)

2. Pattern 3: *Human or Eventuality or Institution allows Physical Object* to-*infinitive.*

   Implicature: *Human or Institution or Eventuality **causes** the opportunity for Physical Object to be V-ed.* (PDEV)

The important part in the implicatures that distinguishes patterns 1 and 3 from each other is the occurrence of verb *give* in pattern 1's implicature as against *cause* in pattern 3 implicature. In addition, the information presented in the implicature is anchored to that in the pattern description, as shown above.

To my knowledge, no other coding methods can provide such a concise implicature as the one used here. If we again take Francis et al. (1996) as an example, it does not present every verb individually; in fact, it groups those verbs that share the same pattern and meaning and labels them with one of the verbs that belongs to the same group. For example, the following verbs that share the pattern *V n* to-*inf.* are labelled under the 'help' group: "*aid*, *allow*, *assist*, *authorize*, *clear*, *(usu passive) empower*, *enable*, *entitle*, *equip*, *fit*, *free*, *help*, *license*, *permit*, and *qualify*" (Francis et al. 1996:294). A general description is then applied to this group, such as 'these verbs are concerned with allowing, enabling, helping, or qualifying someone to do something' (ibid.:294). By contrast, the PDEV considers each of these grouped verbs separately by designating it with a unique pattern and implicature, in the same way as in the PDEV entry for ALLOW (see section 5.2.1.2 and Table 2).

6.3.3.1.3   The range of verb patterns

Hanks (2013) states that the main goal of the PDEV project is to record a set of 'prototypical' patterns or 'norms' of verbs as observed in the BNC using the CPA method. This allows researchers to use the PDEV as a reference, checking the verb patterns recorded in their target corpus as against the ones recorded in the PDEV. This in turn shows what verb patterns are used by the writers in that corpus in addition to how they use them.

Above and beyond the outcomes that were discussed in the preceding sub-sections, such as how these verb patterns are used by these learners, the PDEV entry of ALLOW allows me to see to what extent the Swedish learners make full use of the range of the patterns of this verb. Unlike the LOCNESS writers, the Swedish learners use quite a narrow range of patterns. This supports the findings of Altenberg and Granger (2001), Pu 2000 (cited in Guo 2006) and Ringbom (1998), where learners use a narrower range of patterns of multi-meaning verbs than native speakers. It should be noted that this study shows that such a phenomenon is not restricted to multi-meaning verbs since ALLOW is not one.

The method of comparison implemented in this study, using the PDEV as a reference, will be significant for teachers who wish to bridge the gap that exists between learners and the native-speaker students in terms of the range of verb patterns they are familiar and conversant with. In addition, if the learners show similarity to native-speaker students in terms of the use of the range of verb patterns, this would be a good indication of mastery by the learners. On the other hand, if the learners appear to use a smaller range of verb patterns than the native-speaker students do, it would indicate that they have not mastered the full range of the verb's patterns yet.

6.4 How accurate is the PDEV entry for ALLOW?

Three issues have been discovered with the PDEV entry for ALLOW. First, as shown in Table 6.2, the PDEV entry constantly treats the semantic types Human, Institution and Eventuality as alternatives, whereas learners do not, as discussed in section 6.2.4. A possible solution to resolve this, if the PDEV team wish to produce a version of PDEV that is suitable for learners, or if the teachers want to use the PDEV concept to identify patterns, is to treat each of these semantic types separately. For instance, in pattern 1 of ALLOW the PDEV entry can be altered to include what I call 'sub-patterns', in that each of the three semantic types that can populate the subject slot of ALLOW is typed in a new line, noting that they all have the same implicature. Doing so would help the learners realize that each of these semantic types can be used as a subject of ALLOW pattern 1. This kind of adjustment can be similarly applied to the rest of the patterns that show a similar issue.

Second, there are cases where the entry for ALLOW shows two different patterns but where they appear to have the same meaning. For instance, patterns 1 and 3 are treated as two distinctive patterns whose meanings are different; however, if we look at the descriptions of patterns 1 and 3 (Table 6.2) we will notice that they show the same meaning with different structures; pattern 1 is active while pattern 3 is passive. It is difficult to assert a reason why we should distinguish between these patterns. It would be related to whether the PDEV team themselves think that these two patterns should be included in the entry as two distinctive patterns. It would be better, though, for the team to indicate that pattern 3 is the passive version of pattern 1.

6.5   Conclusion

In this chapter I have examined the similarities and differences in uses of ALLOW produced by the Swedish learners and the native-speaker students. This has been done by implementing the PDEV where the full range of prototypical patterns of ALLOW, as observed in the BNC, is recorded. This method has helped me investigate the learners' use of the verb patterns of ALLOW and compare it with that of the native-speaker students. Each pattern has been discussed in detail. The key findings are that nominalization is less used by the Swedish learners in comparison with the native-speaker writers; some patterns are under-represented by the Swedish learners relative to the native writers; and the Swedish learners are likely to prefer use of Human subjects to Institution subjects, and in addition appear to prefer the use of Human to Eventuality subjects. This chapter concluded by answering the research questions and discussed the advantages of applying this method of comparison (i.e. the use of PDEV). Finally, several problems related to the PDEV entry for ALLOW were discovered and solutions were proposed.

Chapter 7   CPA analysis of the verb SUGGEST

7.1   Introduction

The overall aim of this chapter is to discover how feasible it is for an untrained individual to compile an entry for the PDEV project. This chapter demonstrates how corpus pattern analysis (CPA) works and how a PDEV entry can be established using it. This is done by analysing the usages of a verb. Since one of this chapter's aims is also to investigate the Chinese learners' usage of verbs, the verb to be investigated was selected according to two criteria: it had to be a verb that had not been analysed yet for the PDEV project, and it had to occur relatively frequently in the ICLE-Chinese. The verb SUGGEST fulfilled these criteria, since it occurs relatively frequently in ICLE-Chinese, with 253 instances, and it has not yet been analysed in the PDEV project.

This chapter has four aims: to discover how feasible it is for an individual not trained within the PDEV project to compile a PDEV entry; to demonstrate the processes involved; to compare the usage of SUGGEST in the two target corpora—the ICLE-Chinese and the BNC—and, to conclude this discussion some recommendations for researchers an teachers.

The structure of this chapter is as follows. It starts by analysing a sample of concordance lines of SUGGEST from the BNC. The outcomes of this analysis are then presented in a PDEV entry. This serves as a diagnostic tool for the identification of the usages of SUGGEST in ICLE-Chinese. After that, the chapter discusses how *suggest* can be identified and interpreted in the BNC following the CPA method. Next, the final version of the PDEV entry for SUGGEST will be shown, and a comparison between the target corpora will be conducted according to the outcomes of the PDEV entry for SUGGEST. After that, recommendations for researchers and teachers will be discussed. Finally, a conclusion is provided.

7.2   Identifying patterns of SUGGEST

The CPA method, developed by Hanks (2000), was employed to identify the patterns of SUGGEST (see Chapter 2 for details about the CPA). Following Hanks' method, the analysis of SUGGEST was based on a randomly selected sample of 300 concordance lines from the BNC. The number of concordance lines was set at 300 because all typical patterns of SUGGEST were identified within this number and no new patterns were observed beyond this number of lines. The concordance lines were then sorted into groups that had similar syntactic forms and the same meaning. Next, the implicature of each group was proposed. Finally, the relative frequency of each group in the 300 lines was calculated and presented as a percentage.

Processing the concordance lines shows two main phenomena. First, subjects of SUGGEST are found to be of two kinds: Human and non-Human. The following are some examples of the two subject types, with the subjects underlined:

1) Human

1	choose a light formulation. We	suggest	Vichy Restructure Eye Contour Gel (£14)
2	south in the winter, he does not	suggest	poverty, but dazzles us with
4	interpreter is. Herbert (1978)	suggested	that conference interpreting began only during
5	s about my work. My supervisor	suggests	I should ask Hermann Bondi to
6	The Regional Board representatives	suggested	that the council should move

2) Non-Human

1	. The existence of silica	<u>suggests</u>	that ammonia was not an important gas
2	. However, the <u>evidence</u> so far	<u>suggests</u>	that bats are using the technique
3	l categories of the words. <u>Syntax</u>	<u>suggests</u>	that an adjective should follow the word very
4	public transport. The <u>study</u> also	<u>suggests</u>	that there may be even more
5	The commission	<u>suggests</u>	that the employees should comply with the rules

Second, SUGGEST is found to be followed by three kinds of elements: noun phrases, *that-*clauses, and quotes. The following are some examples for each of the elements:

1) noun phrase

1	it would have been difficult to	<u>suggest</u>	any change in the administration
2	rest days. Both effects	<u>suggest</u>	a compensatory response to
3	education within schools and	<u>suggests</u>	a greatly increased emphasis
4	The difficulty in getting information	<u>suggests</u>	some state secret.
5	The Italian team went on to	<u>suggest</u>	a solution they had prepared

2) *that*-clause

1	mood, it is a small step to	<u>suggest</u>	**that** they should aim for constructive change
2	so on. Anecdotal evidence	<u>suggests</u>	**that** CATS have not sought out local-government
3	A fall in the price of meat	<u>suggests</u>	**that** there is no grain to feed livestock
4	Costings at November 1991 prices	<u>suggest</u>	**that** the scheme costs of £1.3 million.
5	Some commentators even	<u>suggest</u>	**that** the code should be put within a legal framework.

3) quote

1	one of the four girls in the crowd was heard to	**suggest** 'Let's cheer for the poms, you guys'!
2	The customer might **suggest**:	'I get the wrong amount of stationery '
3	The discovery may also be important medically',	Dr Zimmern **suggested**.
4	'Maybe we will find a willing girl at the wedding',	Allan **suggested**.

To this point, four things should be noted. First, in set 3 above, the quote can come either before (lines 3 and 4) or after (lines 1 and 2) SUGGEST. Second, in set 2, where *suggest* is followed by a *that*-clause, two meanings were identified within the set; one is with a modal verb such as *should* (lines 1 and 5), in which case the meaning is *Human puts forward a plan or idea for someone to think about*; the other is without a modal (lines 2 and 4), in which case the meaning is *Human says something which he or she believes is the case*. Third, when the subject of SUGGEST is either Human or Institution, the meaning is fairly close. For example, examples *The commission suggests that the employees should comply with the rules* and *The manager suggests that the employees should comply with the rules* both mean the same. The subject *commission* alternates with *manager* with no effect on meaning. Fourth, when the subject of SUGGEST is non-Human, that is, an Eventuality, the meaning is different. For example, an instance such as *The existence of silica suggests that ammonia was not an important gas* does not mean that *something puts forward a plan or idea for someone to think about*; it actually means that *something implies another thing (*that*-clause)* or *something indicates that it (*that*-clause) might be the case*.

The outcomes of the analysis of *suggest* are presented in Table 7.1 below.

Table 7.1 PDEV entry for *suggest*

Pattern number	Pattern and implicature	% frequency in BNC sample
1	**Pattern:** Human or Institution suggests Action or Plan or Proposition or *that*-clause including modal *should* or Quote **Implicature:** Human or Institution puts forward Action or Plan or Proposition or [*that*-clause] or [Quote] as a plan or proposition for someone to think about **Example:** The Antipodean panel suggests a plan for toxic waste	23
2	**Pattern**: Human suggests *that*-clause **Implicature**: Human says something which he or she believes is the case **Example**: Moscovici (1983) has suggested that the basic issue for social representation research is to understand 'the thinking society'	25
3	**Pattern**: Eventuality 1 suggests Eventuality 2 or *that*-clause **Implicature**: Eventuality 1 implies Eventuality 2 or *that*-clause, or indicates that Eventuality 2 might be the case **Example**: The extreme difficulty in getting information suggests some state secret	52

In the table, column 1 shows the pattern number; the first line of column 2 shows the patterns, the second line of column 2 presents an implicature of the pattern, and the third line gives an example of the pattern in use; and the third column shows the frequency rate of the pattern in the sample of data analysed, shown as a percentage.

As can be seen, based on the sample of 300 concordance lines, Table 7.1 shows three patterns of *suggest*. In spite of the fact that it is not the most frequent such pattern, pattern 1 has been set first since it represents the canonical meaning of the verb SUGGEST (i.e. *Human puts forward something as a plan or proposition for someone to think about*); patterns 2 and 3

follow. Pattern 3 has the highest proportion (52%) in the data analysed; pattern 2 comes second highest proportion (25%).

It is surprising that the canonical pattern (pattern 1) of SUGGEST has the lowest proportion of the three forms. However, a further investigation of the sample analysed revealed that most of the randomly selected instances came from scientific journals and academic books, in which results, evidence, studies, analysis, figures, etc. are reported to suggest some conclusion (e.g. "Evidence suggests that there is a sizeable amount of foreign currency held in Poland"); or in which authors suggest something which they believe is the case (e.g. "Moscovici (1983) has suggested that the basic issue for social representation research is to understand 'the thinking society'").

Having discussed the processing of the concordance lines and the established PDEV entry for SUGGEST, I will now move on to discuss every pattern presented in the table. As I do so, I also mention the challenges that I faced, and summarise them at the end of this chapter.

7.2.1   Pattern 1

As a reminder, and as Table 7.1 shows, SUGGEST in pattern 1 is followed by several elements: noun phrase, *that*-clause or quote (which may precede 'suggest'). The following three examples should make this clearer.

    7.1 *Dr Reid added that it would have been difficult to suggest any **change** in the administration of antibiotics. (BNC)*

    7.2 *Pusey suggested **that we should meet at his flat at six**. (BNC)*

    7.3 *'**How about bringing it to my table?**' he suggested at last. (BNC).*

Sorting these instances into one meaning group was challenging. This is because in each case SUGGEST is followed by more than one element and each element is different from each other. In addition, following the PDEV rules, only one example is allowed to be recorded for each pattern in the PDEV entry. Since SUGGEST in pattern 1 is followed by more than one element, this made it more difficult to determine which one should be chosen as the example. Nevertheless, this is ultimately one pattern, with the same implicature. Similarly, as one of the distinctive features of the PDEV entry is that an implicature is anchored to the pattern (i.e. the elements that accompany SUGGEST should be reflected in the implicature), it was also challenging as to which element should go in the implicature of pattern 1.

It was, therefore, not a straightforward decision whether to gather all the sub-patterns under one implicature and example.

Another challenging task was that of determining the frequency of pattern 1. In the PDEV entry for SUGGEST (Table 7.1), the proportion of each pattern in the sample of data is shown as a percentage. This is another distinctive feature of the PDEV. However, as mentioned earlier, since pattern 1 is followed by more than one element, it was not an easy decision as to whether the frequency rate of each sub-pattern[24] should be recorded in the PDEV.

For consistency, it was decided to follow the PDEV rules and keep Table 7.1 as it is.

---

[24] I have come up with this term *sub-pattern* to refer to elements following a pattern. For example, SUGGEST in pattern 1 is followed by three elements, and since all the elements that come before and after SUGGEST have the same implicature, they are regarded as sub-patterns. This decision will be discussed further in a later section in this chapter.

7.2.2 Pattern 3

As shown in Table 7.1, SUGGEST in pattern 3 is followed by one or both of two elements: Eventuality and *that*-clause. These two different elements nevertheless instantiate the same meaning because the subject of SUGGEST in pattern 3 is one kind, that is, Eventuality, and the implicature of this pattern is that *Eventuality 1 implies Eventuality 2 or* that-*clause; or Eventuality 1 indicates that Eventuality 2 or* that-*clause might be the case*. The same issue of determining what should go into the PDEV entry for pattern 3 was also challenging. In addition, finding the best level of semantic types that can group all lexical items that populate the subject and object slots of *suggest*. As shown in Figure 7.1, below, several lexical items fill the two slots of *suggest*.

Figure 7.1 Concordance lines for pattern 3

2	Zelinsky's description...at the present century,	suggests	a place with a clearly identifiable culture
5	Therefore, anergy of a given V T-cell subset strongly	suggests	previous activation by superantigen.
6	more or less smooth. The drawing shown here	suggests	sharp, step-little changes. These are
7	the decision … the Scots when they broke camp,	suggests	at the very least a lack of confidence
8	Studies of the in vitro … assays, however, have	suggested	hepatitis B virus capsid protein, rather
9	years in view of familial history. Barium enema	suggested	polyps, but
10	too.' 'Much?' Curtis's hands contrived to	suggest	a modest sum. 'Say forty pounds.
11	However, the pattern of merger/approximation that this	suggests	is complicated by the existence of the
12	but nothing on the map nor indeed on the ground	suggests	this. The dangers of 'dating' roads
13	and increased sleep time on rest days. Both effects	suggest	a compensatory response to shorter (3)
14	This view is rooted …and education within schools and	suggests	a greatly increased emphasis upon
15	Another popular novel called the Gospels into question by	suggesting	the existence of a new corpus of first-
16	an impression of good work. These indicators, however,	suggest	little of an officer's abilities in the field,
17	than that expected in women aged 45–60. This	suggests	considerable and largely unrecognised
18	The results...but would also provide fuel for debate,… and	suggest	areas for possible research

19	pottery and a few shards of imported wares also	<u>suggest</u>	some fifth-to sixth-century survival,
20	future. This is the Ricardo 'equivalence theorem', which	<u>suggests</u>	the equivalence of taxes and loans as
21	. The extreme difficulty in getting information	<u>suggests</u>	some state secret. In America, with
22	. Both phrases have spoken to me. The still small voice	<u>suggests</u>	a quiet listening, a quietening of the
23	with lower oesophageal sphincter deficiency. Our results	<u>suggest</u>	a link between smoking and alcohol,
24	and the views of the British government which has	<u>suggested</u>	a similar withdrawal from UNESCO,
25	. A glance at some of their fellow-passengers	<u>suggested</u>	to his mind several reasons why she a
26	of seeing is modified here. Formulated in words	<u>suggested</u>	by this quotation, the question I want to
27	in (69) below, where the phrase the notion that	<u>suggests</u>	surprise at someone having done such
28	is seen thus far south in the winter, he does not	<u>suggest</u>	poverty, but dazzles us with his beauty.
29	. Our results showed that this interpretation,	<u>suggested</u>	by a content-based representation,

Following the ontology developed by Hanks (2000), I linked every lexical items that

populates a subject or an object slot of SUGGEST in the figure to the suitable semantic types

given in the ontology. This practice revealed that different lexical items were falling under

different semantic types in the ontology. Table 7.2 below shows the lexical items that belong

to each semantic type:

Table 7.2 Semantic types and lexical items

Semantic Type	Subject slot	Object slot
Action	effect; interpretation	withdrawal; emphasis; notion; interpretation
Event	difficulty	
Activity		Response; use
State of		Areas; pattern; poverty; results

Affairs		
Investigation	study	
Speech Act	word	
Decision		decision
Property		probability
Process		response

This table shows that there are ten semantic types that represent the range of lexical items that are typically found in the subject and/or object slots of SUGGEST in pattern 3. If all ten of these semantic types were to be recorded in the PDEV entry for *suggest* pattern 3 accordingly, the user would be confused by such dense information. To overcome this issue, since all of the semantic types in Table 7.2 were found to be nested under one semantic type *Eventuality* in the ontology, I chose to record only one semantic type to represent all of the lexical items, found in the pattern, that is *Eventuality*. So, the final presentation of pattern 3 will be as follows:

Pattern 3: Eventuality 1 suggests Eventuality 2 or *that*-clause

Implicature: Eventuality 1 implies Eventuality 2 or *that*-clause, or makes you think that it might be the case

Better presentation of pattern 3, as shown above, would make it easier for the user to read through. In addition, grouping several semantic types under one semantic type saves space, and at the same time this allows user the opportunity to go back and search for more lexical items nested under Eventuality in the ontology.

7.2.3   Issues involved in interpreting concordance lines

The interpretation of the concordance lines using the CPA is not always straightforward. It is generally easy to identify Human or Institution by looking for human names or institution names; however, when it comes to identifying inanimate items, there is a need for deep interpretation. For example, in Figure 7.2, inanimate items are found in the subject and object slots as follows:

Figure 7.2 Inanimate items populating the subject and object slots

1 years in view of familial history. Barium enema   <u>suggested</u>   polyps

This example was taken from the concordance lines of pattern 3 (i.e. *Eventuality 1 suggests Eventuality 2*). Here, a *barium enema* is a diagnostic tool. A first glance at this example might not be enough to decide on the semantic type for a *barium enema*, and so its difficulty to identify the implicature of this pattern. It would be more surprising if *barium enema* became regarded as an Eventuality in this example. However, what is implied by *Barium enema* in this example is not just a name of a diagnostic tool but the investigation process in which a patient is given a barium enema to investigate their bowel problems. The outcome of that investigation is that polyps are present in the bowel. The following is an expanded version of the same example:

> 7.4 *submitted to screening aged 34 years in view of familial history. Barium enema suggested the presence of polyps, but sigmoidoscopy to 22 cm was normal. An incidental asymptomatic breast adenocarcinoma was discovered and she had a right simple mastectomy*. (BNC)

Another issue with interpretation of concordance lines is *ellipsis*. Hanks (2013:173) states that "[t]he regular patterns of usage in a language alternate with one another in three ways: lexical alternations … semantic alternations … and syntactic alternations'. The regularly observable alternation in the analysed sample of SUGGEST was a *syntactic alternation*. This type of alternation is divided into sub-types, one of which is ellipsis (i.e. omission of words), which "allows speakers to imply an argument without explicitly realizing it" (Hanks 2013:210). The commonly observable ellipsis in the analysed sample of SUGGEST is *clausal ellipsis*. In Pattern 1, for instance, *suggest* is followed by a *that*-clause (e.g. *He suggests that I ring her early morning*). There were some instances in the sample analysed where the word *that* was omitted, as shown below:

> 7.5 *Nevertheless, the findings suggest smoking may seriously reduce the chances of pregnancy.* (BNC)

In this example, *that* is omitted after *suggest*. The meaning, of course, is that the findings suggest **that** smoking may seriously reduce the chances of pregnancy. This kind of ellipsis is regarded as meaningful alternation, for the meaning in this example can still be understood.

Another type of syntactic alternation is between *active* and *passive*; this type of alternation was the most frequent type in the sample analysed. Hanks (2013:188) states that "[i]f the subject of a passive verb is the same as the direct object of the same word as an active verb, it is generally not necessary to propose separate patterns for the active and passive uses of that verb" … as they "represent normal uses with differences of focus" (ibid.:174). The following examples should illustrate this phenomenon:

7.6 *A possible way round this problem has been suggested by Sen and others*. (BNC)

(passive)

7.7 *Sen and others have suggested a possible way round this problem.* (Active)

Here, as noted, the former example represents the passive voice and second one the active voice. Both of these instances are regarded as examples of one pattern, that is, pattern 1, since the meaning is not affected by the alternation.

7.3   Missing patterns in the sample analysed

Since one of the PDEV project's aims is to record all prototypical patterns of English verbs, a reliable reference should be consulted; as the sample analysed here was only 300 lines, it was important to refer to a dictionary, built from a corpus, as a cross-check. Thus, the *Collins Cobuild English Dictionary* was consulted to look for any missing pattern of SUGGEST that did not find their way into the PDEV entry for SUGGEST. The consultation of the dictionary revealed that, in addition to the patterns identified in the PDEV (see Table 7.1 above), there is one more pattern to be added to the PDEV. The following is an example of the missing pattern in the PDEV, taken from the online Collins English Dictionary:

7.8 If you **suggest** the name of a person or place, you recommend them to someone.

(Cobuild)

Example*: Could you suggest someone to advise me how to do this?* [VERB noun]

The full BNC was searched for this pattern; some examples extracted from the BNC data are presented below:

7.9 *I suggest Stella Maris, Star of the Sea, since you've decided to make your dwelling place upon the face of the waters.* (BNC)

7.10 *I'd be happy to suggest such a place for Flora.* (BNC)

It was, therefore, determined to be best to add this pattern to the PDEV entry for SUGGEST, as in Table 7.1 above. A revised table presenting the PDEV entry for SUGGEST is shown below:

Table 7.3 A table revised after adding a missing pattern (pattern 4)

Pattern number	Pattern and implicature	% frequency in BNC sample
1	**Pattern:** Human suggests Action or Plan or Proposition or *that*-clause or Quote **Implicature:** Human puts forward Action or Plan or Proposition or *that*-clause or Quote as a plan or proposition for someone to think about **Example:** Antipodean panel suggests plan for toxic waste	23
2	**Pattern**: Human suggests *that*-clause **Implicature**: Human says something which he or she believes is the case **Example**: Moscovici (1983) has suggested that the basic issue for social representation research is to understand 'the thinking society'	25
3	**Pattern**: Eventuality 1 suggests Eventuality 2 or *that*-clause **Implicature**: Eventuality 1 implies Eventuality 2 or *that*-clause, or makes you think that it might be the case **Example**: The extreme difficulty in getting information suggests some state secret	52
4	**Pattern**: Human suggests Name **Implicature**: Human recommends Name of a person or place to Human 2. **Example**: Someone came from County Hall to ask for a list of folk for decoration. I suggested Bridget and Kitty	0

This table includes the added pattern (pattern 4). Certain descriptions were followed when adding this pattern. For instance, the semantic types that populate the subject and object slots of SUGGEST were added to pattern 4. The pattern is presented as follows:

Pattern 4: *Human suggests Name (to Human 2)*

Here, two semantic types were fitted to SUGGEST: Human and Name (i.e. Human populates the subject slot and Name fills the object slot of SUGGEST). After assigning the semantic types, the next step was to anchor the implicature to the pattern, as shown below:

Implicature: *Human recommends a Name of a person or place to Human 2.*

Finally, an example for the pattern was selected from the BNC to show the usage of the pattern, as presented below:

7.11 *Someone came from County Hall to ask for a list of folk for decoration.* ***I suggested Bridget and Kitty.*** (BNC)

7.4 SUGGEST in ICLE-Chinese

Making reference to the PDEV entry for SUGGEST, (see Table 7.3), it was possible to identify usages of *suggest* in ICLE-Chinese by matching the identified patterns in it to those in PDEV. The correct patterns of *suggest* in the ICLE-Chinese (i.e., those matched to the PDEV entry in Table 7.3) were sorted into groups accordingly. The divergent patterns were grouped according to their forms and meanings (see Chapter 5 for more detail on identifying patterns).

The outcomes of that analysis are presented and discussed in the following section.

7.4.1   Overall figures

Table 7.4 shows the frequencies of SUGGEST as identified in the ICLE-Chinese corpus.

Table 7.4 The frequencies of SUGGEST as identified in the ICLE-Chinese corpus

No.	Verb	Raw frequency	Correct usages	Divergent (incorrect) usages
1	SUGGEST	260	211 (81.15%)	49 (18.85%)

Here, the second column contains the target verb under study; the third column indicates the raw frequency of the verb; the fourth column shows the raw frequency and the rate of correct patterns of the verb (i.e. of patterns that correctly matched those in the PDEV); the fifth column includes the raw frequency and the percentage of verb usages that were divergent from those in the PDEV.

The divergent usages of *suggest* identified in the ICLE-Chinese are discussed in the following section.

7.4.2   Divergent usages of SUGGEST

Of 260 instances, 49 (18.85%) instances were regarded as divergent usages of *suggest* in the ICLE-Chinese (i.e. in these instances the usages of *suggest* did not match the patterns in the PDEV, shown in Table 7.3). Figure 7.3 shows 8 instances where SUGGEST is followed by a *to*-infinitive.

Figure 7.3 SUGGEST followed by to-infinitive in ICLE-Chinese

1	crossing, the SAR **Government**	suggests	to lengthen the operation
2	time. The **government**, therefore,	suggested	to construct another rail
3	in Hong Kong, some **developers**	suggest	to develop the country park
4	The Hong Kong **government** has	suggested	to import professionals d
5	Recently, the **government** has	suggested	to establish a law of
6	to solve this problem, **government**	suggested	to legalize soccer bettin
7	. In my opinion, **I** also	suggest	to use recycling to handle
8	few years, therefore some **people**	suggest	to construct a second railw
9	. In my opinion, **I** would	suggest	to develop the countryside.

This figure shows that the semantic types that populate the subject slot of *suggest* are *Human* or *Institution*. The structure of this divergent usage, following the CPA, would be as follows:

Human or Institution suggests to-infinitive

To make sure that such a pattern was divergent, an investigation of the full BNC was conducted; the result of the investigation indeed showed that this pattern is not found in the full BNC. The next step was to interpret the meaning of these divergent usages. This was achieved by consulting the PDEV entry for SUGGEST in Table 7.3. The examination of all instances in Figure 7.3 suggests that the approximate but close meaning of these instances would be *Human or Institution puts forward Action as a plan or proposition for someone to think about.* The reason the instances in Figure7.3 were regarded as divergent is that because *suggest* in all instances together is followed by a *to*-infinitive (e.g. **the SAR government suggests* to lengthen *the operation*). If all the divergent usages together in Figure 7.3 were to

be reproduced correctly, the *to*-infinitive would be omitted and *-ing* would be added to the verbs that follow *suggest*, as shown below:

Figure 7.4 Divergent instances before change

1	crossing, the SAR **Government**	suggests	to lengthen the operation
2	time. The **government**, therefore,	suggested	to construct another rail
3	in Hong Kong, some **developers**	suggest	to develop the country park
4	The Hong Kong **government** has	suggested	to import professionals d
5	Recently, the **government** has	suggested	to establish a law of
6	to solve this problem, **government**	suggested	to legalize soccer bettin
7	. In my opinion, **I** also	suggest	to use recycling to handle
8	few years, therefore some **people**	suggest	to construct a second railw
9	. In my opinion, **I** would	suggest	to develop the countryside.

Figure 7.5 Corrected version of Figure 7.4

1	crossing, the SAR **Government**	suggests	lengthening the operation
2	time. The **government**, therefore,	suggested	constructing another rail
3	in Hong Kong, some **developers**	suggest	developing the country park
4	The Hong Kong **government** has	suggested	importing professionals d
5	Recently, the **government** has	suggested	establishing a law of
6	to solve this problem, **government**	suggested	legalizing soccer bettin
7	. In my opinion, **I** also	suggest	using recycling to handle
8	few years, therefore some **people**	suggest	constructing a second railw
9	. In my opinion, **I** would	suggest	developing the countryside.

By taking these actions, I was able to provide an implicature for the reproduced instances in Figure 7.5: *Human or Institution puts forward Action for someone to think about*[25].

Another divergent usage of SUGGEST identified in ICLE-Chinese is shown in Figure 7.6 below:

Figure 7.6 SUGGEST followed by noun group and to-infinitive in the ICLE-Chinese.

1	students, they are always not	suggested	to bring much money to
2	. But recently, the government is	suggested	to legalize soccer bettin
3	various points of views, I	suggest	government to provide
4	nothing. In this case, I	suggest	myself to engage in some
5	debt. So, I do not	suggest	students to use credit card
6	a more efficient result. I	suggest	the government to do more
7	my opinion I do not	suggest	the student to use it.
8	the economic downturn. He also	suggests	the government to leave the

As can be seen from Figure 7.6, *suggest* in all instances together has the pattern *Human 1 suggests Human 2* to-*infinitive*; this pattern was not found in the full BNC. I would argue that the proper verb to replace *suggest* in these instances would be *advise* (e.g. *I do not advise students to use credit card*). To back up this argument, the original PDEV entry for ADVISE was thus investigated. The PDEV entry for ADVISE (pattern 1) is shown below:

---

[25] This is an amended version of implicature of pattern 1 in the PDEV. It has to be adjusted to anchor it to the reproduced instances in Figure 7.4.

Table 7.5 PDEV entry for ADVISE (pattern 1)

Pattern number	Pattern and implicature
1	**Pattern:** Human 1 or Institution 1 advises Human 2 or Institution 2 *to*-infinitive
	Implicature: Human 1 or Institution 1 **suggests** that it would be a good idea for Human 2 or Institution 2 to-infinitive
	Example: Asimov advised aspiring writers to take their models from history and the classics

Interestingly, as shown in Table 7.5, the implicature of ADVISE includes SUGGEST. In addition, the examination of all instances together in Figure 7.6 showed that they share the same semantic types with ADVISE: Human 1 or Institution 1 advises Human 2 or Institution 2 *to*-infinitive. This suggests that the implicature of all instances together in Figure 7.5 is the same as the implicature of ADVISE pattern 1. The following examples should make this clearer:

7.12   *I do not *suggest* student to use credit cards (incorrect use) (ICLE-Chinese)

7.13   I do not *advise* student to use credit cards (correct use)

Above, example 7.12, taken from Figure 7.6, represents a divergent pattern produced by the Chinese learners. Example 7.13 shows the same example with SUGGEST replaced with ADVISE.

7.4.3   Non-divergent usage of SUGGEST in ICLE-Chinese and BNC

Non-divergent usages of SUGGEST (i.e. the patterns that matched those in the PDEV entry) produced by the Chinese learners were examined and compared with those in the PDEV entry for SUGGEST, in Table 7.3. The results of this comparison are presented in Table 7.6 below

Table 7.6 PDEV entry for SUGGEST

Pattern number	Pattern and implicature	% frequency in ICLE-Chinese	% frequency in LOCNESS
1	**Pattern:** Human suggests Action or Plan or Proposition or *that*-clause or Quote **Implicature:** Human puts forward Action or Plan or Proposition or *that*-clause or Quote as a plan or proposition for someone to think about **Example:** Antipodean panel suggests plan for toxic waste	25.60	23.46
2	**Pattern**: **HUMAN suggests that-clause** **Implicature**: Human says something which he or she believes is the case **Example**: Moscovici (1983) has suggested that the basic issue for social representation research is to understand 'the thinking society'	26.10	25.7
3	**Pattern**: Eventuality 1 suggests Eventuality 2 or *that*-clause **Implicature**: Eventuality 1 implies Eventuality 2 or *that*-clause, or makes you think that it might be the case **Example**: The extreme difficulty in getting information suggests some state secret	17.37	53.8
4	**Pattern**: Human suggests Name **Implicature**: Human recommends Name of a person or place to Human 2 **Example**: Someone came from County Hall to ask for a list of folk for decoration. I suggested Bridget and Kitty	**0**	**0**

As a reminder, column 1 shows the pattern number; the first line of column 2 shows the pattern, the second line of column 2 presents an implicature of the pattern, and the third line gives an example of the pattern in use; column 3 shows the frequency rate of the pattern in the ICLE-Chinese; and column 4 shows the frequency rate of the pattern in the LOCNESS. The most striking observation regarding the percentages in this table is related to pattern 3. Pattern 3 comprises a high proportion of the BNC sample (53.8%) while this pattern accounts for the lowest proportion of the ICLE-Chinese (17.37%). This, I would argue, suggests this pattern would seem to unfold in a restricted register that some of the Chinese learners might not know. In addition, since this pattern (pattern 3) is different from the other patterns (e.g. patterns 1 and 2 in Table 7.6), in that the semantic type that populates the subject and object slot of pattern 3 is Eventuality, while the semantic type Human populates the slots in patterns 1 and 2, the Chinese learners may be less aware of this phenomenon; in other words, the Chinese learners might be more familiar with the fact that Human is the common semantic type used with SUGGEST. For the purpose of improvement, teachers may engage in activities to raise the learners' awareness of the existence of pattern 3.

Another observable phenomenon in Table 7.6 is that of pattern 4. As mentioned earlier in this chapter, pattern 4 is infrequent in the full BNC. If a pattern is infrequently used by the native speakers of English, it is assumed that there will not be enough opportunities for learners to write it. For the part of the Chinese learners, this pattern would again seem to have a restricted genre that some of the Chinese learners might not know.

7.4.4 Pattern found in ICLE-Chinese but not found in the BNC sample

The task of matching the patterns identified in the ICLE-Chinese to those in the PDEV entry for SUGGEST revealed a further interesting phenomenon: of 211 non-divergent patterns of

SUGGEST in ICLE-Chinese, 11 instances were found in the learners' sample but not in the

BNC sample. These 11 instances are shown below:

Figure 7.7 Instances of *suggest* found in ICLE-Chinese but not in the BNC

1	a large number of opponents	suggest	abolishing this cruel and b
2	banning smoking in restaurants is	suggested	. Although this suggestion
3	smoke. Some people even	suggested	banning smoking in all re
4	may be true because government	suggest	banning smoking in restaura
5	ion in the restaurants, Government	suggests	banning smoking in restaur
6	ultation document is raised, which	suggests	banning smoking in all ind
7	, nowadays, more and more people	suggests	banning smoking in restaur
8	Mainland China. Some people	suggest	developing Hong Kong's coun
9	reason why many union leaders	suggested	importing professionals f
10	Chief Executive Tung Chee hua	suggested	importing professionals i
11	betting. Some of the citizens	suggested	legalizing soccer betting and some

In this figure, all instances taken together resemble the pattern *Human or Institution suggest -ing-clause*; that is, the semantic type that fills the subject slot of *suggest* is Human or Institution, and SUGGEST is followed by an *-ing*-clause. To make sure that this pattern was non-divergent, the full BNC was manually searched for instances of this pattern. The results of this search showed that there were 287 instances of this pattern in the full BNC. A sample of concordance lines of this pattern taken from the full BNC is shown below:

Figure 7.8 Instances of SUGGEST found in the full BNC but not in ICLE-Chinese

1	may result.' Instead, the **company**	suggests	installing a £30,000 spray system
2	cannot prosper. The **Commission**	suggests	setting up red squirrel protection areas
3	The Charity Organization **Society**	suggested	segregating the degenerate residuum in
4	follow what the call was about. **She**	suggested	meeting Wainfleet at lunchtime to hear
5	what they should do. **Charlotte**	suggested	contacting the police, but Ursula would
6	ten layers to a press. Many people	suggest	using corrugated card for pressing and,

It can be argued that an *-ing*-clause that follows SUGGEST, as shown in Figures 7.7 and 7.8 above, can be labelled under the semantic type Action, which fills the object slot of SUGGEST pattern 1 (see Table 7.6 for pattern 1). However, although this view is reasonable and seems to hold true, this usage may not be clear to the learners. Thus, it was necessary to add the *-ing*-clause to the subject slot of SUGGEST pattern 1 in the PDEV entry. The final version of the PDEV entry for SUGGEST, after the addition of the *-ing*-clause, is shown in Table 7.7 below:

Table 7.7 Final version of the PDEV entry for SUGGEST

Pattern number	Pattern and implicature	% frequency in BNC sample
1	**Pattern:** Human suggests Action or Plan or Proposition or *that*-clause including model *should* or *-ing*-clause or Quote **Implicature:** Human puts forward Action or Plan or Proposition or [that-clause] or [Quote] as a plan or proposition for someone to think about **Example:** Antipodean panel suggests plan for toxic waste	23
2	**Pattern**: Human suggests that-clause **Implicature**: Human says something which he or she believes is the case **Example**: Moscovici (1983) has suggested that the basic issue for social representation research is to understand 'the thinking society'	25

	Pattern: **Eventuality 1 suggests Eventuality 2 or *that*-clause**	52
3	**Implicature**: Eventuality 1 implies Eventuality 2 or *that*-clause, or makes you think that it might be the case	
	Example: The extreme difficulty in getting information suggests some state secret	
4	**Pattern**: Human suggests Name	0
	Implicature: Human recommends Name of a person or place to Human 2	
	Example: Someone came from County Hall to ask for a list of folk for decoration. I suggested Bridget and Kitty	

7.5 SUGGEST in LOCNESS

The usage of SUGGEST was investigated in the LOCNESS for two reasons. First, since the PDEV entry for SUGGEST has been established based on a sample analysed from the BNC, and the genres in the BNC are different from those in LOCNESS (i.e. the BNC includes, for example, extracts from regional and national newspapers, specialist periodicals and journals, academic books and popular fiction, letters and memoranda, among many other kinds of text, whereas the LOCNESS is mainly argumentative essays), I speculated that there may be divergent patterns of SUGGEST produced by the native-speaker students. Second, as the ICLE corpus and the normative corpus, that is LOCNESS, are reported in the literature to be comparable, I speculated that the comparison of the usage of SUGGEST in the two corpora would yield some more interesting results.

To investigate the former hypothesis, the concordance lines of SUGGEST were analysed by matching the identified patterns in the LOCNESS to those in the PDEV entry for SUGGEST (see table 7.8 for the PDEV entry). A total of 51 instances of SUGGEST were retrieved from the LOCNESS using AntConc. All of the instances were matched to the PDEV entry for

SUGGEST. No atypical patterns were seen in the LOCNESS, nor new prototypical patterns found. This result refutes my first assumption mentioned above.

The second assumption was tested by employing a frequency comparison between the non-divergent patterns of SUGGEST in both corpora: the ICLE-Chinese and LOCNESS. The proportions of each pattern of SUGGEST in both corpora are shown in the PDEV entry below:

Table 7.8 PDEV entry for SUGGEST with frequencies of patterns in the two corpora

Pattern number	Pattern and implicature	% frequency in ICLE-Chinese	% frequency in LOCNESS
1	**Pattern:** Human suggests Action or Plan or Proposition or *that*-clause or -*ing*-clause Quote **Implicature:** Human puts forward Action or Plan or Proposition or [*that*-clause] or [Quote] as a plan or proposition for someone to think about **Example:**. Antipodean panel suggests plan for toxic waste	25.60	19.60
2	**Pattern**: HUMAN suggests *that*-clause **Implicature**: Human says something which he or she believes is the case. **Example**: Moscovici (1983) has suggested that the basic issue for social representation research is to understand 'the thinking society'	26.10	33.33
3	**Pattern**: Eventuality 1 suggests Eventuality 2 or *that*-clause **Implicature**: Eventuality 1 implies Eventuality 2 or *that*-clause, or makes you think that it might be the case **Example**: The extreme difficulty in getting information suggests some state secret	17.37	39.21
4	**Pattern**: Human suggests Name **Implicature**: Human recommends Name of a person or place to Human 2 **Example**: Someone came from County Hall to ask for a list of folk for decoration. I suggested Bridget and Kitty	**0**	**0**

This table provides at least two important insights. First, as can be seen in Table 7.8, the proportions of the patterns are not the same in each corpus. Pattern 1 (25.60%) and pattern 2 (26.10%) are more dominant patterns of SUGGEST in comparison with pattern 3 (17.37%) in ICLE-Chinese. In contrast, pattern 3 (39.21%) is the dominant pattern relative to pattern 1 (19.60%) and pattern 2 (33.33%) in LOCNESS. The reason for this difference, I would argue, is the semantic type that populates the subject slot in all three patterns. For instance, according to the PDEV entry for SUGGEST, the semantic type that populates the subject slots of patterns 1 and 2, as shown in Table 7.8, is Human (e.g. <u>Michael</u> *suggested that we visit London*; <u>I</u> *am not suggesting that is what is happening*.). In contrast, the semantic type that populates the subject slot of pattern 3 is Eventuality (e.g. <u>This chain of events</u> *suggests that Britain got into medium- and high-level waste technology early*). This difference between these three patterns suggests that the Chinese learners may lack the knowledge that the semantic type Eventuality can also be used as a subject of SUGGEST (i.e. *Eventuality 1 suggests Eventuality 2 or* that-*clause*). It would also suggest that they are using the word for different functions: the Chinese writers are reporting what someone said/wrote, whereas the LOCNESS writers are constructing an argument.

Second, however, at the same time, this table shows similarity between the ICLE-Chinese and LOCNESS, as far as the use of SUGGEST is concerned. On the one hand, the learners and the native-speaker writers use the same range of patterns. Of the 4 patterns recorded in the PDEV, 3 patterns were identified in the two corpora, that is, patterns 1, 2 and 3. On the other hand, pattern 4 was not found in either corpus.

7.6   Sub-patterns of SUGGEST in ICLE-Chinese and LOCNESS

According to the PDEV entry for SUGGEST, it is followed by more than one element in at least two patterns, namely, patterns 1 and 3 (see Table 7.8 for the final version of this entry in the PDEV). For instance, pattern 1 in the PDEV shows that SUGGEST is followed by a semantic type (e.g. Action, Plan or Proposition), *that*-clause, *-ing*-clause or quote, and by a semantic type (i.e. Eventuality) or *that*-clause in pattern 3. That observation led me to ask this question: to what extent is each of these elements used in ICLE-Chinese and in LOCNESS, as far as SUGGEST is concerned? To answer this question, each of these elements was re-sorted and counted. On this point, it is important to note that each of these elements was regarded as part of the patterns of SUGGEST. In addition, the label 'sub-pattern' was given to each element. For example, pattern 1 in the PDEV was divided into 4 sub-patterns: a) Human suggests Action *or* Plan *or* Proposition; b) Human suggests that-clause; c) Human suggests -*ing*-clause; d) Human suggests Quote; similarly, pattern 3 was divided into 2 sub-patterns: a) Eventuality 1 suggests Eventuality 2; b) Eventuality suggests that-clause.

The results are shown in Table 7.9 below:

Table 7.9 Relative frequency of sub-patterns of *suggest* pattern 1 in ICLE-Chinese

No.	PDEV	%Raw frequency in ICLE-Chinese	%Raw frequency in LOCNESS
Pattern 1	A1. Human suggests **Action or Plan or Proposition**	27.53%	37.5%%
	B1. Human suggests *that*-**clause**	50.72%	50%
	C1. Human suggests -*ing*-**clause**	21.73%	12.5%
	D1. Human suggests **Quote**	0%	0%

As can be seen in table 7.9, the proportions across the sub-patterns of SUGGEST pattern 1 are unequal in both the ICLE-Chinese and the LOCNESS corpora. Sub-pattern B is the dominant sub-pattern in both corpora, with the highest proportion (50%) of uses. Sub-pattern A is the next most frequent, in both corpora, and sub-pattern C the next. However, the difference in frequency between A and C is larger in the LOCNESS corpus than in the ICLE corpus. Neither corpus contains any instances of sub-pattern D.

The implicatures of all the sub-patterns are the same (i.e. Human puts forward Action or Plan or Proposition or [*that*-clause] or [Quote] as a plan or proposition for someone to think about). The dominance of sub-pattern B is evidence that this is the first pattern the learners learn

Similarly, the instances of pattern 3 identified in ICLE-Chinese and LOCNESS were re-sorted and counted. The result is provided in Table 7.10 below:

Table 7.10 Relative frequency of sub-patterns of *suggest* pattern 3 in ICLE-Chinese

No.	PDEV	%Raw frequency in ICLE-Chinese	%Raw frequency in LOCNESS
Pattern 3	A3. Eventuality 1 suggests **Eventuality 2**	100%	85%
	B3. Eventuality suggests *that*-clause	0%	15%

Table 7.10 shows another striking phenomenon. No occurrence of sub-pattern B3 was found in the ICLE-Chinese sample; the only sub-pattern found in ICLE-Chinese is A3 (i.e. *Eventuality 1 suggests Eventuality 2*). This indicates, I would argue, that the Chinese learners

might not be aware of the existence of sub-pattern B3, that is, of the fact that SUGGEST, in pattern 3, can be followed by a *that*-clause.

Table 7.10 also shows that both sub-patterns A3 and B3 were found to be produced by the native writers. However, the proportion of occurrences of sub-pattern B3 (15%) is low compared to that of sub-pattern A3 (85%). This suggests that the learners learn the most frequent pattern first.

7.7   Discussion

7.7.1   Establishing the PDEV entry for SUGGEST

In the first half of this chapter, the usage of *suggest* was analysed following the CPA method. Since *suggest* has not been analysed yet in the PDEV project, it gave me the opportunity to show how this can be done, and how it can be translated into a PDEV entry. I would like to conclude this process with some recommendations for an untrained individual who wishes to compile a PDEV entry. The recommendations can be summarised as follows:

A. The verb to be analysed should occur relatively frequently in both the learners' corpus and the native speakers' corpus, to yield solid results. However, it is also important to note that high raw frequency of a verb does not necessarily mean that there are a high number of patterns associated with it: for example, AFFECT has a frequency of 8643 in the BNC50, but the PDEV entry for AFFECT records only 2 patterns for this verb, whereas 25 patterns are recorded for BANG, which occurs only 205 times in the BNC50.

B. A selected sample of a verb from a normative corpus (e.g. the BNC) may not include all prototypical 'patterns' of a verb. Hence, it is advisable for analysts (e.g. teachers,

researchers, or material writers) to consult a reliable dictionary for cross-check. The online *Collins Cobuild English Dictionary* is among the highly recommended dictionaries that can be consulted[26].

C. Sometimes, the 'meaning' of some patterns of a verb may be difficult to identify; a consultation of reliable dictionaries is needed. When it comes to word definition, consulting more than one dictionary is highly recommended to achieve a sound judgment on the implicature of the pattern.

D. When divergent patterns of a verb have been identified in the learner corpus (by matching the patterns in the learners corpora to those in the PDEV), it is important to search for them in a normative corpus such as the British National Corpus (BNC) or the Bank of English (BoE) corpus. This would make the analyst confident enough to regard the patterns as divergent and thus as recordable in the PDEV.

E. However, it should be pointed out not all the verb usages in the BNC are always correct; there are some instances where usages of verbs are incorrect. For example, the following instance is found in the BNC: '*I would* suggest *you to ask that question in fact you'll ask Gordon afterwards*'. Here, *suggest* is followed by a noun group (pronoun) and *to*-infinitive. This pattern would sound peculiar to a native speaker of English, as it is not typical that *suggest* is followed by a noun group and *to*-infinitive. If an expert analyst is in doubt about a particular usage of a verb, he or she may consult a native-English-speaking human rater about it. In other words, he or she should not be totally dependent on the corpus data.

F. As shown in section 7.2.2, several lexical items are identified in the object slot of SUGGEST pattern 3 (see Table 7.2, section 7.2.2 for details). Each of these lexical items is found in nine different semantic types (see table 7.2 for these semantic types)

---

[26] https://www.collinsdictionary.com.

in the ontology of the original PDEV. It would be impractical to record all nine

semantic types in the subject slot of SUGGEST pattern 3 in the PDEV entry for

SUGGEST pattern 3. Hence, an alternative way of presenting all 9 semantic types in

the PDEV is to group them into one semantic type, that is, Eventuality (e.g.

Eventuality 1 suggests Eventuality 2), as the semantic type Eventuality represents a

plausible, and indeed the best, level of generalisation for the range of all lexical items

observed in the object slot of SUGGEST pattern 3. In other words, having one

semantic type populating the subject slot of SUGGEST, as long as all the lexical items

belong to that semantic type, is more practical. This is the fruitfulness of the PDEV

ontology's hierarchal nature, as "a hierarchical ontology enables relevant

generalizations to be made and implicatures stated" (Hanks and Ježek 2008:393).

7.8   Conclusion

In the first half of this chapter, I have demonstrated how the CPA method works by analysing

the verb SUGGEST, which had not been completed in the original PDEV. Some challenges

appeared in relation to the analysis of the sample from the BNC. Interpreting the concordance

lines was not straightforward, and some lexicographical skills were needed to make sound

judgments on certain interpretations. The outcomes of this analysis have been presented in the

PDEV entry for SUGGEST in accordance with the principles of the original PDEV. Existing

dictionaries were frequently consulted throughout the process of establishing the new PDEV

entry for SUGGEST. This helped me decide on the best level of implicature to adopt for each

pattern of *suggest* presented in the PDEV. In addition, consulting dictionaries has enabled me

to add a pattern that did not initially find its way into the PDEV entry for SUGGEST.

In the second half of the chapter, the established PDEV entry for SUGGEST was used to identify the patterns of SUGGEST in ICLE-Chinese. This showed that there exist divergent usages of SUGGEST in the ICLE-Chinese corpus. Furthermore, a prototypical usage of SUGGEST that did not find its way into the PDEV has been found in the ICLE-Chinese sample (namely, *Human or Institution suggests* -ing-*clause*), and has been added to pattern 1 in the PDEV entry for SUGGEST. This should make it clearer for the learners that pattern 1 of SUGGEST can also be followed by an -*ing*-clause. Furthermore, a frequency comparison of the non-divergent patterns of SUGGEST between ICLE-Chinese and LOCNESS was also conducted. This revealed some similarities and disparities between the two corpora. Furthermore, another frequency comparison was done, of sub-patterns in ICLE-Chinese, and it too revealed some interesting findings.

Finally, this chapter discussed some issues related to the establishment of a PDEV entry for SUGGEST. It then concluded with some recommendations for an untrained individual who wishes to compile a PDEV entry.

Chapter 8   Discussion of Themes Emerging from the Results

8.1   Introduction

This chapter discusses themes that have emerged from the three findings chapters, just presented. The three main research questions this thesis intends to answer will be dealt with in Chapters 9 and 10.

These themes are also expressed as questions, as follows:

1. How does PDEV compare to other similar models for describing language?

2. To what extent do learners use PDEV patterns correctly?

3. What can learners' preferences in pattern use tell us about the competence of those learners and about English as a global language?

It should be noted that question 1 here does not emerge from the findings, as it discusses the theoretical standing of PDEV in relation to other similar theories. The aim of including this question in this chapter is to determine how good CPA/PDEV is for learners in comparison with other approaches of describing language. In other words, do these approaches offer new information that can be included in the PDEV for the benefit of learners?

8.2   How does PDEV compare to other similar models for describing language?

8.2.1   Lexis and syntax

Recent views of language stress the inseparability of lexis and syntax, such as works by Sinclair (1991, 2004), Hunston and Francis (2000), and Goldberg (1995, 2006). Sinclair, for instance, proposes that units of meaning, composed of collocation, colligation, semantic preference, or prosody, are a central model in the field of corpus linguistics studies (see

Chapter 2), and the UoM model sees lexis (semantics) and syntax (grammar) as two inseparable components of forming meaning. Hunston and Francis's view of the inseparability of lexis and syntax is based on the investigation of the grammatical behaviour of particular words as observed in the texts (see Francis et al. 1996, 1998 for examples).

The inseparability of lexis and syntax has become a central point not only in corpus linguistics but also in cognitive linguistics (for example, Goldberg 1995, 2006). The CPA/PDEV method chimes with this inseparability, particularly Sinclair's take on it, in several aspects. First, patterns of verb that are recorded in PDEV are identified in their context; in other words, different usages of verb are realized by the words surrounding them. For example, the outcome of CPA investigation of the BNC shows that AGREE is found to be followed by grammatical components such as *that*-clauses, *with*, and *to*-infinitives, with semantic preferences (types) such as Human, Eventuality, Proposition, Activity, etc. (see Chapter 2 for details). The way the CPA/PDEV is set to look at verb usages is in this way, to some extent, a manifestation of the 'lexico-grammatical' phenomenon.

Many of the observations found in PDEV can be described in terms of two models: units of meaning (Sinclair 2004) and pattern grammar (Hunston and Francis 2000). However, there are some places where the models diverge. These are:

1. Introductory *it* as Subject (*it be* V-*ed that*-clause; *it be* V-*ed to*-inf)
2. Word-form and lemma

Each of these aspects will be discussed below, respectively.

8.2.1.1 Introductory *it* as subject (*it be* V-*ed that*-clause; *it be* V-*ed to*-inf)

Francis et al. (1996) is a good treatment of the lexico-grammatical phenomenon (that lexis and syntax are not separable). However, some patterns identified in this pattern grammar book are not described as patterns in PDEV. I will take some examples from this book as counter-examples indicating what CPA/PDEV might not consider as a pattern in relation to the lexico-grammatical phenomenon.

It is reported in Francis et al. (1996) that there are some verbs observed to frequently occur with the pattern *it be* V-*ed* clause in the Bank of English corpus. Among these verbs are *accept*, *agree*, *admit*, *argue*, *suggest*, *claim*, *point out*. Some examples taken from Francis et al. (1996) are provided below:

8.1 *Eventually it <u>was agreed</u> that the present laws would continue to apply in the same areas for two years*. (Francis et al. 1996)

8.2 *It <u>is suggested</u> that teachers should design activities which will keep the class actively and constructively engaged*. (Francis et al. 1996)

8.3 *When it <u>was pointed out</u> that she would need considerable journalistic experience she agreed she didn't have it*. (Francis et al. 1996)

8.4 *It <u>can be argued</u> that human health would not suffer if we were to stop most animal research*. (Francis et al. 1996)

8.5 *It <u>is claimed</u> that running helps to unleash hidden energies, both psychic and physical*. (Francis et al. 1996)

As reported in Francis et al. (1996:527), "[t]his pattern is used to indicate that something is said, thought, or discovered, without indicating who said, thought, or discovered it." An

investigation of the description of patterns in the PDEV entry for the above-mentioned verbs (e.g. *accept*, *agree*, *admit*, *argue*, *suggest*, *claim*, *point out*) shows that no pattern of this kind (*it be* V-*ed that*-clause) has been recorded in the PDEV entries for these verbs. However, a further search of the annotated concordance lines in the PDEV entry for each of these verbs reveals that there are instances where each of these verbs is in fact found to occur in the above-mentioned pattern, that is, *it be* V-*ed that*-clause. The following are examples of some of these verbs:

8.6 AGREE

Pattern 3: Human or Institution agrees *that*-clause

Implicature: Human or Institution indicates that he, she, or it accepts *that*-clause

Example: *It was agreed, however, that the OAU Secretariat should present to the 1993 summit the results of a feasibility study*. (PDEV)

8.7 SUGGEST

Pattern 1: Human suggests that-clause.

Implicature: Human puts forward [that-clause] as a plan or proposition for someone to think about.

Example: *Third, it has been suggested that government provision may encourage an unhelpful separation between payment and cost in the minds of consumers*. (PDEV)

8.8 ADMIT

Pattern 1: Human 1 or Institution 1 admits quote or *that*-clause (to Human 2 or to Institution 2)

Implicature: Human 1 or Institution 1 reluctantly says (to Human 2 or Institution 2) that Eventuality = Bad is a true fact

Example: *It must be admitted that Château Meloch was a great deal better some years than others*. (PDEV)

One possible explanation why the pattern *it be* V-*ed that*-clause is not recorded in the PDEV, as far as the above-mentioned verbs are concerned, is that the PDEV treats these instances as syntactic alternations, where such alternations have no effect on the meaning (see Chapter 2 for details on alternations). If we take AGREE as an example, the example above where *agree* occurs in the pattern *it be* V-*ed that*-clause can be alternated into the active voice with no change in meaning, as follows:

8.9 *It was agreed, however, that the OAU Secretariat should present to the 1993 summit the results of a feasibility study*. (BNC)

8.10 *The committee agreed, however, that the OAU Secretariat should present to the 1993 summit the results of a feasibility study*. (Amended)

The meaning in both examples is the same, even though their patterns are different.

To put it simply, I would argue that the reason these few instances (with introductory *it*) are found in the PDEV annotated lines is that these instances are treated as instances of the passive voice in contexts where the distinction between active and passive voice has no effect on meaning. On the other hand, another possible explanation why the pattern *it be* V-*ed that*-clause has not been recorded in a PDEV entry might be that the number of instances of the pattern is too few, in the sample analysed by the PDEV team, to be recorded.

From my point of view, it would have been better if the PDEV were to have considered *it be* V-*ed that*-clause as a pattern for the given verbs. This is because my further inspection of the full BNC revealed that there are quite a number of instances where the pattern occurs with the given verbs in the BNC, which means that this pattern is 'typical' of the given verbs.

8.2.2 The word-form and lemma

Sinclair (1991:41) observed that the word-forms that constitute a lemma do not occur with equal frequency across various patterns and meanings. For example, *yielded* prominently occurs with the meaning 'led to'; *yields* mainly occurs with the meaning 'produces'; *yield* mostly occurs with the meaning 'give way'. He further concludes that 'there is a strong tendency for sense and syntax to be associated' (Sinclair 1991:65).

The CPA/PDEV does not take differences in word-forms into account, but instead treats all word-forms that constitute a lemma as the same. For example, the concordance lines for ALLOW pattern 1, as an example, include four different word-forms of ALLOW (*allow*, *allows*, *allowed*, and *allowing*), but most instances of pattern 1 are passives with *allowed*. However, what is typically recorded in the PDEV entry is a pattern in one word-form (e.g. Human *allows* Human *to*-infinitive). One possible explanation is that since PDEV is a dictionary it may follow the conventions of a dictionary in that it treats every part of the lemma as the same, an approach I would agree with—among other reasons because if PDEV were to do what Sinclair did with *yield*, the size of PDEV would be huge, as every verb entry would consist of multiple separate word-forms whose meanings would be the same.

For the benefit of learners, I would argue that such distinction should be made clear in that there should be another pattern where *allowed* constitutes the following pattern, following the PDEV convention:

> **Pattern**: Human or Institution or Eventuality be allowed *to*-infinitive
>
> **Implicature**: Human or Institution or Eventuality is given opportunity, time, or permission *to*-infinitive.
>
> **Example**: *After all, I am not <u>allowed to drive</u> a motorcar without having proved my ability to a 'competent authority' — the driving test examiner.*

8.2.3 Constructions and CPA/PDEV: The form–meaning pairing

Goldberg (2003:219) defines constructions as "stored pairings of form and function, including morphemes, words, idioms, partially lexically filled and fully general linguistic patterns." These morphemes (e.g. the comparative *-er*), words (e.g. *orange*), idioms (e.g. *kick the bucket*), partially lexically filled linguistic patterns (e.g. verb–argument constructions) and fully general linguistics patterns (e.g. negative *not*) are all treated as constructions that are stored in the mind and each have their own function and meaning. The major area where the PDEV approach and construction grammar, particularly under this definition, have similarities and differences is 'partially lexically filled linguistic patterns'. An example of this meeting point is verb–argument constructions, as noted above. Goldberg (2006) argues that verb–argument constructions have meanings in themselves regardless of the words that occur in these constructions. For example, although the verb *sneeze* is typically intransitive, it works as transitive in the caused-motion construction '*Sam sneezed the napkin off the table*' (Ellis et al. 2016:32). This view is not accounted for in the CPA/PDEV approach, for several reasons. First, unlike the verb–argument approach, CPA/PDEV investigates 'patterns' of English verbs individually (i.e. verb by verb). Second, the PDEV records only 'typical' patterns of each verb

in an individual inventory, as observed in a corpus (i.e. the PDEV does not look for 'possible'

usages of a verb).

I would argue that the concept of CPA/PDEV is more adequate to the linguistic facts than that

of construction grammar for a number of reasons. First, CPA uses a corpus-based approach,

in that it studies what is observed in the corpus collectively; it takes frequency of occurrence

into account. Second, CPA/PDEV investigates every verb succinctly, in that all prototypical

patterns of each individual verb are identified and each pattern is given an implicature.

Another key difference between construction grammar (verb–argument constructions) and the

CPA/PDEV is that the CPA/PDEV makes use of an ontology of semantic types (see Chapter

2 for the ontology). To illustrate the difference, consider the follow table:

Table 8.1 PDEV vs. verb–argument constructions: Subject and object slots

PDEV	Verb–argument constructions
Human 1 or Institution 1 or Eventuality gives Human 2 or Institution 2 the opportunity, time, or permission *to*-inf [verb]	Subject–Verb–Object construction X allows Y to do something Or V n to-infinitive (in Francis et al.'s (1996) convention)
Human or Institution or Eventuality causes or provides the opportunity for Physical Object to be V-*ed*	Subject–Verb–Object construction X causes Y to be Z V n *be* v-*ed*

Although the two approaches overlap in the structure of verb (e.g. subject–verb–object; transitive, etc.), the CPA/PDEV went a step further by describing the implicature of such a structure. For pattern 1, the subject slot is filled with one of the semantic types Human, Institution or Eventuality; the object is replaced with Human or Institution followed by a *to-*infinitive. Similarly, the subject of ALLOW in pattern 3 is replaced by any of the same labels, and the object is replaced by the semantic type Physical Object followed by *to be* V-*ed*. This means that the subject and object slots for each of the two patterns are 'limited' to those semantic types, and the meaning of each of these patterns is different, as shown in the table above.

To conclude, although construction grammar and CPA/PDEV overlap in that they both treat lexis and syntax as one unit of meaning CPA/PDEV is differentiated from construction grammar in that it uses the ontology.

8.3   Do learners use PDEV patterns correctly?

One of the main questions asked by this study is the extent to which learners use PDEV patterns correctly. In order to answer this question, the usages of the 16 target verbs were identified in both learner sub-corpora: ICLE-Chinese and ICLE-Swedish. The results are presented in Table 8.2 below.

Table 8.2 Frequencies of divergent and non-divergent patterns of target verbs in learners and native-speaker writers

no.	verb	Number of patterns recorded in the PDEV entry	Chinese		Swedish	
			Non-Divergent	Divergent	Non-Divergent	Divergent
1	ADMIT	13	40	0	20	0
2	ADVISE	9	11	2	1	0
3	AFFORD	5	189	3	35	0
4	AGREE	10	157	51	42	0
5	ALLOW	8	164	35	73	1
6	APPEAR	12	52	3	28	0
7	ASK	12	66	2	52	0
8	AVOID	3	71	33	34	1
9	CLAIM	6	173	0	52	0
10	DENY	9	35	0	21	0
11	ENCOURAGE	3	158	16	43	0
12	MAINTAIN	5	61	0	18	0
13	PLAN	4	50	2	15	0
14	PROPOSE	8	63	3	3	0
15	TELL	21	55	4	74	1
16	URGE	6	30	3	3	0
Total			1304	231	474	3

The most significant information this table reveals is that the proportions of non-divergent patterns of target verbs in both corpora outnumber those of divergent patterns, which in turn means that the learners most of the time produce non-divergent (correct) patterns.

In addition, however, this table reveals three meaningful phenomena. First, the proportion of non-divergent patterns is not the same for each verb and/or for each group of students. For instance, for the verb AGREE, only 75% of the total instances in the Chinese corpus are

correct, whereas for CLAIM all (100%) instances are correct; for the verb ALLOW the Chinese students are incorrect in 17%, whereas Swedish students are correct in nearly all instances. This in principle raises two questions: a) why the proportion of non-divergent patterns is not the same for each verb; b) why the proportion of non-divergent patterns is not the same for each group of students (Chinese and Swedish). A possible explanation for the second question is that the Swedish language may share a considerably greater number of commonalities with the English language than the Chinese language does, and that the Swedish students may thus do better than the Chinese students. In addition, English is widely used in Sweden, and Swedish speakers regularly interact with foreigners, often using English as a lingua franca.

The question of the difference in proportions among the different verbs is rather more interesting. There are at least two possible explanations for this phenomenon. One of the possibilities is that the number of divergent patterns is a measurement of the difficulty of the verb concerned. The other is that the proportion of non-divergent patterns of the 16 verb lemmas in each of the two learners groups outnumbers that of the divergent ones, suggesting that the learners produce correct patterns most of the time.

More interestingly, however, the qualitative analysis of these divergent instances shows three phenomena: 1) the learners use recognisable patterns; 2) several learners use the same divergent pattern; 3) the divergent pattern is often one that is used with another verb with a similar meaning (see Tables 5.1, 6.1, 7.1 for details). The following is an example of the third-mentioned phenomenon taken from ICLE-Chinese, with a corrected version of the example:

8.11 *It may **lead** the situation in Hong Kong **worse**. (ICLE-Chinese)

8.12 It *may* **make** *the situation in Hong Kong* **worse** (corrected).

In this example, a number of learners were found to use the verb *lead* in the pattern that belongs to the verb *make*. This is interesting in two respects. First, it is not difficult to describe such divergent patterns using the PDEV description. Second, this shows that the learners knew the meaning of the pattern and verb even if they have produced divergent instances.

8.4   What can learners' preferences in pattern use tell us?

It is not only divergent uses that are of interest when we look at the learners. In terms of the use of a range of non-divergent patterns of verbs, the investigation shows that the learners do not attempt to use the full range of patterns of verbs concerned. For example, of 8 patterns of ALLOW recorded in the PDEV entry, only 3 patterns are used by the Swedish learners: patterns 1, 4, and 7; and the proportions of these patterns are not the same as in the PDEV entry. Two general possibilities can be sketched as to why the Swedish learners do not use the full range of patterns of ALLOW. First, this may be because the learners do not learn them, or because they do not have opportunities to use them. For instance, the essay topics posed to them may not lead to full use of the available range of patterns. The second group of possible speculations might be related to the pattern itself. For example, the description of pattern 8 of *allow*, according to the PDEV entry, is Human *allows that*-clause, whose implicature is *Human admits the truth of that*-clause. This pattern seems old-fashioned and rare, as it only occurs 3 times in the BNC sample of ALLOW. The following is an example of pattern 8 taken from the BNC:

8.13 *In fairness to Rollin he does **allow** that such rights may be breached, but only for*

life-or-death reasons. (BNC)

It is possible that learners would substitute another verb such as *admit*, which occurs in the

implicature statement of *allow*. If so, the learners would produce the same example but with

admit, as follows:

8.14 *In fairness to Rollin he does **admit** that such rights may be breached, but only for*

life-or-death reasons

The alternative to this pattern is pattern 1 of ADMIT, in which *Human admits* that-*clause*.

The proportion of ADMIT pattern 1 in PDEV is 44.4%, according to the PDEV entry,

whereas the proportion of ALLOW pattern 8 is only 1.6%; in other words, ADMIT pattern 1

is a very common use of this verb whereas ALLOW pattern 8 is a very uncommon use of the

verb. Similarly, pattern 1 of ADMIT has the highest proportion rate in the two learner groups

(ICLE-Swedish and ICLE-Chinese) as well as the native-speaker writers (LOCNESS). It is

therefore not surprising that pattern 8 of ALLOW does not occur in any of the target corpora.

Another possible explanation may lie with patterns of a given verb in relation to each other.

That is, the existence of some patterns might affect the use of others. If we take AGREE as an

example, several instances are found in the ICLE-Chinese where AGREE is followed by a *to-*

infinitive, while no occurrence is found with AGREE followed by the preposition *to* followed

by a noun group. Both of these patterns, according to the PDEV entry for *agree*, carry the

same meaning, that a Human or Institution indicates that he or she or it is willing to undertake

something. In the following two examples; the former is taken from the ICLE-Chinese and the latter is an amended version:

8.15 *The government agreed to construct a new railway*. (Pattern 4; ICLE-Chinese)

8.16 *The government agreed to the construction of a new railway*. (pattern 5, amended version)

It is interesting that the PDEV entry for AGREE shows an alternative to pattern 4. Learners who can achieve a particular meaning with one pattern will be unlikely to expend the effort to learn a second pattern with a similar implicature. It would therefore be good if teachers would take action to raise their students' awareness of the existence of pattern 5.

Another interpretation of the above-mentioned phenomenon (the *to* + noun pattern) is offered. Some patterns might require a more metaphoric approach to grammar—using Halliday's (1993) terms, 'nominalization' or 'non-congruent'. For example, according to its PDEV entry, the semantic type that fills the object slot of pattern 4 of ALLOW is Eventuality as in *the teacher allows <u>the use of mobile phones</u> in the classroom*, where *the use of mobile phones* is a non-congruent or nominalization pattern. The congruent version of the same example is *The teacher allows pupils to use <u>the mobile phones</u> in the classroom*. Changes that take place in the two examples are verb change and noun change: *allows the use of mobile phones* (V n) has been changed to *allows pupils to use the mobile phones* (V n *to*-infinitive). Another example is that of ENCOURAGE. The object slot of pattern 3 of ENCOURAGE is filled with one semantic type, Eventuality, as in *The government <u>encourages the use of public transport</u>*. The non-congruent version of this example is *The government <u>encourages people to use public transport</u>*.

196

Similarly, the phenomenon of nominalization or non-congruent usage can be found in the subject slots. For example, the semantic type that fills the subject slot of pattern 1 of ALLOW is Human/Institution or Eventuality. An example of this pattern is as follows:

8.17 *The change in the law allowed people to smoke in restaurants.* (non-congruent)

This example represents the non-congruent version. The congruent version of the same example is as follows:

8.18 *When the law was changed, people were allowed to smoke in restaurants.* (congruent)

The changes between these two examples take place in the subject slots: the noun *change* in example 8.17 becomes the verb *change* in example 8.18.

Even though these two examples are both correct and carry the same meaning, the learners preferred to produce instances similar to example 8.17; the subject slot of the former is filled with Human/Institution, while that of the latter is filled with Eventuality or non-Human, as underlined.

8.5   Conclusion

This chapter has addressed some themes that have emerged from the results chapters. With regard to what we have learned about learners from the results, the main findings are as follows: a) they mostly produce correct (non-divergent) patterns, and when they produce incorrect (divergent) patterns they mostly produce them in sensible ways, that is, they still

produce patterns that are recognisable; b) they often use analogy with another verb; c) they tend to use a restricted range of patterns that involve avoidance of nominalization.

Concerning the comparison between CPA/PDEV and other related concepts, it can be concluded that although there are some slight differences between CPA/PDEV and the other concepts, discussed in this chapter, they all correlate with the concepts of CPA/PDEV in that they all look at lexis and syntax as inseparable parts.

The next two chapters will answer the main research questions of this thesis: What are the pedagogical implications of applying PDEV to L1-Chinese and L1 Swedish learners? (Chapter 9); Can the concept of CPA/PDEV be used to successfully describe learner corpora? (Chapter 10); and, How useful is PDEV to learners and teachers? (Chapter 10).

Chapter 9   **Pedagogical Implications**

9.1   1 Introduction

This chapter presents the pedagogical implications of the results of this thesis. Specifically, it answers the research question *What are the pedagogical implications of applying PDEV to L1-Chinese and L1 Swedish learners?* It is divided into two parts. The first part (section 9.2) focuses on the divergent uses of verbs produced by the Chinese learners; this section contains five sub-sections, each of which deals with one verb (9.2.1 AGREE; 9.2.2 AVOID; 9.2.3 LEAD; 9.2.4 SUGGEST; 9.2.5 ALLOW). The second part (section 9.3) is devoted to the non-divergent patterns of ALLOW identified in the Swedish sub-corpus.

When considering how PDEV might be made relevant to language learners, an obvious connection to draw is with data-driven learning (DDL). A popular, innovative method of exploring language in the classroom, DDL was proposed by Tim Johns (Johns 1988, 1991, 1994, 1996, and 2002). DDL, as defined by Johns and King (1991:iii, cited in Granger and Tribble 1998:200), is "the use in the classroom of computer-generated concordances to get students to explore regularities of patterning in the target language and the development of activities and exercises based on concordance output." DDL has become popular in ELT (cf. Johns 2002; O'Keeffe et al. 2007; O'Keeffe and McCarthy 2010; Charles 2011; Boulton 2012). The literature on DDL reports several ways of implementing DDL in the classroom: on-line corpora use, hands-off concordances (e.g. printed concordance lines), or even isolated corpus-driven examples (Frankenberg-Garcia 2014; Chujo et al. 2015).

This chapter, therefore, will be offering teachers ways of presenting the outcomes of investigations of the uses of target verbs identified in the Chinese and Swedish sub-corpora to learners. In general, these ways involve the following: the teachers show the learners the

PDEV entry and instances produced by learners and talk them through the differences between the two. In some cases, the teachers will rewrite what the learners have produced.

9.2 Some pedagogical implications for Chinese learners: Divergent uses of five verbs

9.2.1 AGREE

As seen in chapter 5, the investigation of the target verbs shows that there are a number of divergent usages that can be identified for most of the target verbs in ICLE-Chinese. Sets of instances of these divergent usages are produced by several learners. Furthermore, the learners appear to produce several types of divergent patterns for each of the target verbs. For example, 4 different sets of instances of AGREE are identified in ICLE-Chinese (see section 5.7.6). In one of these sets, 12 instances (Figure 9.1 below) instantiate the following divergent pattern:

Pattern: Human agrees *to*-infinitive

Implicature: Human indicates that he or she has a favourable attitude towards a suggested proposition that is proposed by another human.

Example: In conclusion, I agree to construct a second railway link.

The PDEV entry for 'Human agrees *to*-infinitive' is shown below in Table 9.1.

Table 9.1 PDEV entry for AGREE (pattern 4)

Pattern number	Pattern and implicature	% BNC sample
4	**Pattern**: Human or Institution agrees *to*-infinitive **Implicature**: Human or Institution *indicates that he, she or it* is willing to undertake [V]	27.8%

	Example: We agreed to meet at school the following week	

In comparison with pattern 4 (Table 9.1), what makes the learners' uses divergent is that there is a mismatch between the implicature of these instances and the implicature of the pattern in standard English as shown in the PDEV entry. In other words, the learners do not mean that they are willing to undertake an action (e.g. to construct a second railway); rather, they indicate that they have a favorable attitude towards an idea or proposition that is proposed by another human, as shown in the implicature above. Such a mismatch would not be captured without consultation of the PDEV entry for AGREE.

Figure 9.1 Divergent instances of agree produced by the Chinese learners

1	and cons in each aspect, **I** don't	agree	**to construct** a second railway in this
2	an economic benefits. In conclusion, **I**	agree	**to construct** a second railway link to
3	would be much more better. Anyway, **I**	agree	**to construct** a second railway link.
4	mentally feeling. In my opinion, **I** am not	agree	**to build** a second railway in Hong Kong
5	its advantages and disadvantage. For me, **I**	agree	**to ban** smoking in restaurants as I realize
6	Hong Kong. However, in my view, **I** totally	agree	**to use** the recycling as a method of
7	may increase. In my opinion, **I** do not	agree	**to ban** smoking in restaurants. It is
8	in restaurants. From my point of view, **I**	agree	**to** ban smoking. Smoking is really a bad
9	play and bet on them. So, **I**	agree	**to legalize** soccer betting in Hong Kong.
10	will complaint the government. **I** don't	agree	**to ban** smoking in restaurants and it is not
11	affect others' health. However, **I** do not	agree	**to ban** smoking in bars because smoking
12	smoke-free establishments. In my opinion, **I**	agree	**to do** a favour by banning smoking in

Interestingly, an alternative pattern to the divergent one, found in the PDEV entry for AGREE, can be used, as shown in Table 9.2 below:

Table 9.2 Alternative to the divergent pattern of AGREE produced by the learners

Pattern number	Pattern and implicature	% BNC sample
3	**Pattern**: Human or Institution agrees *that*-clause **Implicature**: Human or Institution *indicates that he, she or it accepts that*-clause	26.0%

As seen above, in this alternative AGREE is followed by *that*-clause, and the implicature fits the instances produced by the learners.

What the teachers need to do in order to correct learners' divergent instances by exploiting PDEV is to reproduce the learners' instances (such as in Figure 9.1) with the corrected version (in this case using pattern 3) and provide the students with copies of these corrected instances and the corresponding divergent instances. To illustrate this, one example is given below (Figure 9.2).

Figure 9.2 Corrected version of learners' instances using pattern 3 in the PDEV entry

1	I don't	agree	**that a second railway should be constructed.**
2	In conclusion, I	agree	**that a second railway should be constructed.**
3	Anyway, I	agree	**that a second railway link should be constructed** .
4	In my opinion, I do not	agree	**that a second railway should be built in Hong Kong.**
5	For me, I	agree	**that smoking should be banned in restaurants.**
6	However, in my view, I totally	agree	**that the recycling should be used as a method of waste management**.

7	In my opinion, I do not	agree	**that smoking should be banned in restaurants**.
8	From my point of view, I	agree	**that smoking should be banned**.
9	So, I	agree	**that betting should be legalized in Hong Kong**.
10	I don't	agree	**that smoking should be banned in restaurants**.
11	However, I do not	agree	**that smoking should be banned in bars**.

The corrected parts of the items in Figure 9.2 are shown in bold. In these instances, what comes after AGREE has been rephrased so that it conforms to pattern 3 and its implicature. The teachers can then talk their students through the two versions, showing them the differences in the usage and meaning of AGREE in the two versions. Then, the teachers can talk learners through the corrected version and show them how items in bold can be rephrased using passive voice (e.g. *I agree to construct a railway* can be paraphrased to *I agree that a second railway should be constructed*).

Another way of using the PDEV to correct the learners' usages of AGREE is to provide them with a printed screenshot of the PDEV entry showing patterns 3 and 4 of *agree* and encourage them to notice the difference in the description of the two patterns. A printed screenshot would be useful for the learners for several reasons. First, the PDEV entry has been recently updated in terms of the way the description of patterns are presented. Two examples of the PDEV entry before and after, are shown below (Table 9.3) (Figure 9.3).

Table 9.3 Adapted old version of the PDEV entry for AGREE (patterns 3 and 4)

Pattern number	Pattern and implicature
3	**Pattern**: [[Human]] or [[Institution]] agrees *that*-clause **Implicature**: Human or Institution indicates that he, she or it accepts *that*-clause

4	**Pattern**: [[Human]] or [[Institution]] agrees to-inf
	Implicature: [[Human]] or [[Institution]] indicates that he, she or it is willing to undertake [V]

Figure 9.3 Screenshot from the new PDEV entry for AGREE (patterns 3 and 4)

As can be seen, the PDEV entry has been further articulated and made more effective. In Table 9.3, patterns are presented all in black and semantic types are put between double brackets. In Figure 9.3, however, it can be seen that the new version is more user-friendly, especially for learners; several colours have been used and small caps are used as well. Each colour represents a specific function. For example, semantic types are in yellow, the node word (verb) in green, pronouns in black, and functional grammar items in purple and typed in small caps. This helps learners to notice the differences between the two patterns (provided they are not colour-blind). Another advantage of providing the learners with the screenshotted version is that it includes the implicature. For example, in pattern 3, the notion of 'someone accepts an idea' is made clear in the implicature, while the notion of 'someone is willing to do something' is expressed in the implicature of pattern 4. More interestingly, the items found in each of the two pattern descriptions are reflected in their implicatures, such as semantic types that fill the subject slots (e.g. Human or Institution) and the functional grammatical items that follow the node (e.g. *that*-clause or *to*-infinitive). Finally, each pattern entry includes an example taken from the BNC sample. This is useful to help the learners observe the patterns in authentic language.

204

Another set of divergent usages identified in ICLE-Chinese for AGREE is provided below:

Figure 9.4 Divergent instances of *agree* produced by Chinese learners

1	On one hand, I	agree	on this attitude.
2	That is true, I	agree	on that absolutely.
3	But in my opinion, I	agree	on the implementation of MPF in Hong Kong

In these instances, the subject slot of AGREE is filled with a singular Human *I*, and *agree* is followed by preposition *on* followed by a noun group. One pattern can be used to draw students' attention to why these instances are divergent. This pattern is recorded in the PDEV entry for AGREE pattern 7, shown as Figure 9.5 below.

Figure 9.5 PDEV entry for AGREE (pattern 7)

> 7   Pattern: Human_Group agrees on or upon Action or on or upon Proposition
> Implicature: members of Human_Group reach a consensus regarding a Proposition for some future Action
> Example: The European Council agreed on the creation of a European police office (Europol)

As seen in Figure 9.5, the semantic type that fills the subject slot of AGREE here (pattern 7) is Human_Group, and the node word (*agree*) is followed by prepositions *on* or *upon* followed by a noun group. This means that in instances where AGREE is followed by *on/upon* and a noun group, the subject slot should be filled with a noun phrase indicating more than one human being. The implicature of pattern 7 makes this phenomenon clear, in that only 'members of a human group' (but not one human) can reach a consensus regarding a proposition for some future action. The teachers would make this phenomenon clear to the students by providing them with a copy of this entry in Figure 9.5 and letting them observe

the pattern and its implicature. After that, they[27] would ask students to go back to the

divergent instances in Figure 9.4 and explain to them that in this pattern the subject slot

cannot be filled with a singular subject (the use of 'I' in this case) because one human cannot

reach a consensus; rather, only a group of people can agree on something. The teacher then

provides them with alternatives to these divergent instances. These alternatives can be found

in the PDEV entry for *agree*, as shown below.

Figure 9.6 PDEV pattern for AGREE (pattern 2)

2   Pattern: Human or Institution agrees (with Proposition or with Activity)
      Implicature: Human or Institution indicates that he, she, or it has a favourable attitude towards Proposition or Activity
      Example: Councillors agreed with the recommendation by Darlington Borough Council not to have special exemptions for disabled drivers.

As seen in Figure 9.6, the semantic type that fills the subject slot for AGREE pattern 2 is not

restricted to a noun phrase indicating more than one human (contrary to pattern 7), and

AGREE is followed by *with* and a noun group (e.g. Proposition or Activity). In this way, the

implicature of pattern 2 is different from that of pattern 7. Pattern 7 is argued to be the one

that the students were aiming for in the instances above (see Figure 9.4). Thus, the teacher

would provide the students with a copy of pattern 2 of AGREE (Figure 9.6) and ask them to

consider the pattern description and read the implicature to understand more deeply what this

pattern means. After that, the teacher would ask the students to amend instances like those in

Figure 9.4 by replacing the preposition *on* with *with*. The teacher can also take some instances

that instantiate pattern 2 (e.g. someone agrees with something) from the BNC, and let the

students observe the pattern in these instances while he explains to them the meaning of these

instances. Another set of divergent instances produced by the Chinese learners is shown in

Figure 9.7.

---

[27] Throughout this chapter, I am using the gender-neutral *they* to include teachers of all genders.

Figure 9.7 Divergent instances of AGREE produced by the learners

1	most of Hong Kong **women** do not	agree	**abortion**, however, because of
2	**woman** in Hong Kong almost does not	agree	**abortion**. But in fact, abortion is
3	of the middle class **women** in Cosmo	agree	**abortion**. They believe that they
4	the pros and cons of abortion. **People**	agree	**abortion** because they think it is
5	On the other hands, **others** do not	agree	**abortion** since it may harm the
6	do not agree at that. Those **people** who	agree	**abortion** are called pre-choice, it
7	to , most middle class **women** in Cosmo	agree	**abortion** as they feel abortion is
8	and right of living. In my opinion, **I** do not	agree	**abortion** in most of the cases. I think
9	of abortion is rise. In my opinion, **I** do not	agree	**abortion** because I feel that pregnancies

Here *agree* is followed by a specific noun, that is, *abortion*. It is grammatically correct that *agree* is followed by noun groups, as shown in the PDEV entry for *agree* (see Figure 9.8 below).

Figure 9.8 PDEV entry for AGREE (pattern 6)

6 Pattern: Human_Group agrees Activity or Proposition
Implicature: members of Human_Group reach a consensus about Activity = Future or Proposition
Example: *The council have agreed a 10 % salary supplement for staff with language skills'.*

If we look at the implicature of pattern 6, we can observe that *agree* is followed by a noun (e.g. an Activity or Proposition). This is similar to the instances produced by the students, as shown in Figure 9.7 above. However, the meaning in students' instances is different from that of pattern 6. The meaning of pattern 6 is that *members of a Human group reach a consensus about a future activity or proposition* (e.g. *The council have agreed a 10% salary supplement for staff with language skills*), whereas the intended meaning of all instances produced by the learners together is that a Human agrees with or has a favorable attitude towards proposition, that is, *abortion*. In other words, the Humans that fill the subject slots of instances in Figure

9.7 are in a position of not 'reaching a consensus about the future of abortion' (e.g. *In my opinion, I don't agree abortion in most of the cases*). In addition, regardless of whether the subject is a group of people or singular, the use of this particular word, *abortion*, in this pattern is not semantically correct (i.e. the meaning is not correct in that *abortion* is not a future activity nor a proposition that someone can reach a consensus about).

In such a case, the teacher would explain the difference to the students and provide them with an alternative pattern such as pattern 2 of AGREE, as shown in Figure 9.6 above, that is, *Human agrees with proposition*. The teacher can amend his/her students' instances in Figure 9.7 to use *agree* followed by *with*, as shown in Figure 9.9 below. The teacher can then encourage the students to notice the pattern in these instances while the teacher explains the meaning of these instances.

Another way of explaining this pattern is that the teachers select some of the annotated concordance lines of pattern 2 from the PDEV entry and provide the students with copies of them, and encourages the students to observe the pattern in these lines. Teachers should be warned that not all annotated concordance lines are easy to understand, especially for beginner learners. Thus, the teacher should be selective and try to avoid instances they see as difficult for the learners to understand.

Figure 9.9 Corrected instances of AGREE produced by the learners

1	most of Hong Kong **women** do not	agree	with **abortion**, however, because of
2	**woman** in Hong Kong almost does not	agree	with **abortion**. But in fact, abortion is
3	of the middle class **women** in Cosmo	agree	with **abortion**. They believe that they
4	the pros and cons of abortion. **People**	agree	with **abortion** because they think it is
5	On the other hands, **others** do not	agree	with **abortion** since it may harm the

6	do not agree at that. Those **people** who	agree	<u>with</u> **abortion** are called pre-choice, it
7	to , most middle class **women** in Cosmo	agree	<u>with</u> **abortion** as they feel abortion is
8	and right of living. In my opinion, **I** do not	agree	<u>with</u> **abortion** in most of the cases. I think
9	of abortion is rise. In my opinion, **I** do not	agree	<u>with</u> **abortion** because I feel that pregnancies

9.2.2 AVOID

In chapter 4, we have seen several sets of divergent instances of AVOID produced by the Chinese learners. One of these sets is provided below as Figure 9.10:

Figure 9.10 Divergent instances of AVOID produced by the learners

1	debt can evenly distributed **which** can	avoid	great debt burden. By contrast
2	the benefits of economy, **recycling** can	avoid	much pollution. Also, I would
3	industrial use. Then the urban **area** can	avoid	over-crowding. Hong Kong
4	**choosing** the sex of the baby can	avoid	some hereditary diseases.
5	of to burn it or landfill it, **this** can	avoid	the harmful by-products.
6	space. **Recycling** of Plastic products	avoid	the above problems so that it
7	globally and in some cases, **it**	avoids	the troubles of carrying too
8	to saving money, **recycling** also	avoid	the production of pollutants.
9	**This** can also ensure your safety and	avoid	the high unemployment rate
10	**Recycling** also provides an advantage to	avoid	the problems of toxic by-products
11	of saving our lives from hell, **it** can	avoid	any inherited diseases such as

What makes these instances divergent from the norm is the semantic types that fill the subject slot of *avoid*. All lexical items that fill the subject slot in of these instances are related to Eventuality (and as such are non-Human), such as *recycling*, *credit cards*, *choosing the sex of the baby*, etc. No pattern in the PDEV entry of AVOID conforms to this divergent pattern,

and it was not difficult to confirm that these instances are divergent by consulting the PDEV

entry for AVOID. This consultation revealed that the PDEV entry for AVOID has a pattern

that stands as an alternative to the divergent one, as shown in Figure 9.10 below:

Figure 9.11 Screenshot of PDEV entry for AVOID (pattern 1)

1 Pattern: Human or Institution avoids Eventuality
   Implicature: Human or Institution takes action to prevent Eventuality from occurring
   Example: If the applicant has survived this far and **avoided** a pre-hearing assessment, the case will be listed for hearing.

This figure shows that the semantic types that fill the subject slot of pattern 1 of AVOID can

be either Human or Institution. In this case, the teacher would provide their students with the

instances they produced in Figure 9.9 along with a copy of concordance lines based on the

entry. They would then draw students' attention to their mistakes in these instances, and talk

them through the entry in Figure 9.10. The teacher could ask the students to read the pattern

description first, and then ask them to notice what precedes *avoid* in the entry, that is, Human

or Institution. At the same time, they could ask the learners to find or underline the lexical

items that fill the subject slot in the instances in Figure 9.10. After that, the teacher could ask

a question like 'Are these lexical items human?' If some students respond that they are not

sure, the teacher could then ask them to read the implicature. The implicature would dispel

any doubt, as it includes the phrase 'takes action to prevent', as in 'Human or Institution takes

action to prevent Eventuality from occurring'. The teacher then would stress that lexical items

such as *recycling* or *credit cards* does not take action; human is the one who can act to take

action. In other words, an Eventuality cannot 'typically' avoid another Eventuality.

9.2.3   LEAD

Six different sets of divergent instances of LEAD are found in ICLE-Chinese. One of the sets

is shown in Figure 9.12 below:

Figure 9.12 Divergent instances of LEAD produced by the learners

1	need to spend. As a result, it	leads	**to rise up** the tax payment of
2	heard objection to importing professionals	leads	**to increase** the unemployment rate. According
3	the Hong Kong's local workers. It	leads	**to increase** the unemployment rate and
4	the job. Although using credit card may	lead	**to affect** one's study, indicates that
5	Smoke- free establishments would then	lead	**to increase** the economy in Boston. In
6	.K. If smoking is prohibited, this may	lead	**to reduce** the business of catering industry.
7	. Importing professionals, however, may	lead	**to worsen** the economic situation in Hong
8	Professional Scheme does not necessary	lead	**to increase** the unemployment rate in Hong
9	job. On the other hand, It can	lead	**to hurt** the local experts' prospects and
10	problem of the decline of population that	lead	**to weaken** strength, reduce the economy

In these 10 instances, LEAD is followed by a *to*-infinitive, and the subject slots are filled with lexical items that belong to the Eventuality category (e.g. *smoking*, *using credit cards*). This set is labeled divergent because LEAD is followed by a *to*-infinitive, which is not possible according to the PDEV entry for LEAD. Rather, LEAD can be followed by the preposition *to* (e.g. *This leads to death*), noun group (e.g. *In 1986 the governor led a successful campaign by the state's voters*), or noun group followed by *to*-infinitive (e.g. *This leads her to mention that her mother's name is Wanda*).

An alternative to this divergent pattern (LEAD followed by to-infinitive) is pattern 1, as shown in Figure 9.13 below.

Figure 9.13 PDEV entry for LEAD (pattern 1)

1 Pattern: Eventuality 1 leads to Eventuality 2
Implicature: Eventuality 1 is the cause of Eventuality 2
Example: This may lead to problems with regard to business efficiency.

This figure shows that LEAD can be followed by preposition *to* followed by Eventuality. Thus, the teacher would want to rephrase the learners' relevant instances (see Figure 9.12) by using pattern 1 (Figure 9.13). Such rephrased instances are shown in Figure 9.14 below.

Figure 9.14 Rephrased instances of *lead*

1	need to spend. As a result, it	leads	to a rise in the tax payment of
2	heard objection to importing professionals	leads	to an increase in the unemployment rate. According
3	the Hong Kong's local workers. It	leads	to an increase in the unemployment rate and
5	Smoke- free establishments would then	lead	to an increase in the economy in Boston. In
6	.K. If smoking is prohibited, this may	lead	to a reduction in the business of catering industry.
7	. Importing professionals, however, may	lead	to a situation in which the economic in Hong Kong gets worse
8	Professional Scheme does not necessary	lead	to an increase in the unemployment rate in Hong

As seen, the paraphrased items are underlined; these underlined phrased have been rephrased from verb with *to*-infinitive to the form *to*-preposition + Eventuality (e.g. **this may lead to reduce the business of catering industry* is rephrased to *this may lead to a reduction in the business of catering industry*), following pattern 1 of LEAD.

There are at least two possible ways of encouraging learners to observe this correct pattern. First, the teacher would want to provide the students with the two versions: version 1, covering their divergent instances, as in Figure 9.12, and version 2, which includes the paraphrased ones, as in Figure 9.14. The teacher would then talk the students through the two versions, showing them the differences in patterns and implicatures between the two versions. Then the teacher would explain to the students what has been changed from the divergent to the standard version. The second way of presenting this difference would be to select some concordance lines representing pattern 1 from the PDEV entry where *lead* is followed by *to* and a noun group (e.g. *It also led to a confusion between those buildings with a social or civic*

importance, and those without). The teacher can create an exercise using these examples by removing the *to*-prepositional phrase and replacing it with a gap indicator and the relevant verb, as shown here: *It also led ……. (*confuse*) between those buildings with a social or civic importance, and those without*. The teacher would then ask the students to complete this sentence with a correct pattern of LEAD. This would be done by changing the verb between the two brackets, *confuse*, into a noun, (*a*) *confusion*, preceded by the preposition *to*, as shown below:

9.1 *It also led <u>to a confusion</u> between those buildings with a social or civic importance, and those without.*

By doing so, the students would get the chance to practice paraphrasing and, at the same time, observe the difference between patterns.

Another phenomenon identified in the data is that the Chinese learners tend to use a recognisable pattern with the wrong verb, as shown in Figure 9.15.

Figure 9.15 Divergent instances of *lead* produced by the learners

1	information in cyber cafes. says It	leads	a teenager's life more **colourful**
2	cafes can really improve our life and	lead	it more **colourful** and **meaningful** way.
3	problem of lacking skilful workers and	lead	the company more **competitive**. It also
4	during the economic downturn, it may	lead	the situation in Hong Kong **worse** and
5	Import of specialists from the mainland may	lead	the jobless rate **higher**. The next

As can be seen from Figure 9.15, LEAD, in all these instances, is followed by a noun group followed by adjective (e.g. *it leads a teenagers' life more colourful*), which is not possible

and which makes these instances divergent. This particular pattern belongs to another verb, that is, *make*. It would have been useful for present purposes if the PDEV entry for *make* had been complete, to see what alternatives it might offer; however, it had not been completed at the time of writing this thesis.

Nevertheless, the teacher could use the same instances produced by the learners in a task, as follows. The teacher would want to reproduce all the instances in Figure 9.15 but replace the node word LEAD with MAKE, as follows:

Figure 9.16 Instances of *make*

1	information in cyber cafes. says It	makes	a teenager's life more **colourful**
2	cafes can really improve our life and	make	it more **colourful** and **meaningful** way.
3	problem of lacking skilful workers and	make	the company more **competitive**. It also
4	during the economic downturn, it may	make	the situation in Hong Kong **worse** and
5	Import of specialists from the mainland may	make	the jobless rate **higher**. The next

Here, the node word MAKE in all instances is followed by a noun group followed by an adjective. The teacher would provide the students with two versions: version 1, including the students' instances (Figure 9.15), and version 2, including the same instances with *make* as the node word (Figure 9.16). The teacher would then encourage the students to observe the patterns in the two versions and ask them to focus on what comes after the node words in the two versions. The teacher would then tell them what had been changed in version 2 (namely, taking out the node word in version 1, LEAD, and replacing it with the correct word that belongs with the pattern, MAKE).

9.2.4 SUGGEST

A similar phenomenon is found with SUGGEST. The following are some divergent instances produced by the Chinese learners:

Figure 9.17 Divergent instances of *suggest* produced by the learners

1	students, they are always not	suggested	to bring much money to
2	. But recently, the government is	suggested	to legalize soccer bettin
3	various points of views, I	suggest	mass media and government to provide
4	debt. So, I do not	suggest	students to use credit card
5	a more efficient result. I	suggest	the government to do more
6	my opinion I do not	suggest	the student to use it.
7	the economic downturn. He also	suggests	the government to leave the

Figure 9.17 shows that SUGGEST is followed by lexical items that belong to the Human or Institution types (e.g. *students*, *government*) followed by a *to*-infinitive (e.g. *to use*, *to leave*, etc.).In addition, the lexical items that fill the subject slot of SUGGEST belong to Human or Institution categories (e.g. I, He, government). Such a pattern, though inappropriate to SUGGEST, is appropriate to another verb, that is, ADVISE. The PDEV entry for ADVISE confirms this argument; the following is the entry (pattern 1):

Figure 9.18 PDEV entry for ADVISE (pattern 1)

1 *Pattern:* Human 1 *or* Institution 1 advises Human 2 *or* Institution 2 TO-INFINITIVE
Implicature: Human 1 *or* Institution 1 suggests that it would be a good idea for Human 2 *or* Institution 2 TO/INF [VERB]
Example: Asimov advised aspiring writers to take their models from history and the classics.

Figure 9.18 shows that the semantic types that populate subject and object slots of ADVISE are Human and Institution, and that the verb is followed by a noun group followed by a *to*-infinitive (e.g. *The teacher advised the students not to smoke inside the building*).

To correct students' instances with SUGGEST given above, the teacher can reproduce them in Figure 9.17 with ADVISE in the centre, as shown below:

Figure 9.19 Corrected instances with ADVISE replacing SUGGEST

1	the government is	advised	*to legalize* soccer betting
2	various points of views, I	advise	mass media and government to provide
3	In this case, I	advise	myself *to engage* in some
4	So, I do not	advise	students *to use* credit cards
5	a more efficient result. I	advise	the government *to do* more
6	my opinion I do not	advise	the student *to use* it.
7	the economic downturn. He also	advised	the government *to leave* the

The teacher can use these figures in the classroom in several ways. First, the teacher can provide the students with these three figures (Figures 9.17, 9.18, and 9.19). Then they can ask the students to look at the two versions of instances (that is, version 1, in Figure 9.17, and version 2, in Figure 9.19). Next, they can explain that version 1 includes instances with the wrong verb, SUGGEST, in the centre, as produced by students, and version 2 includes corresponding instances with the right verb, ADVISE, in the centre. Then they can explain to the students why these changes to arrive at version 2 are needed. They can point out that SUGGEST does not fit with the instances in Figure 9.19 because SUGGEST cannot be followed by a noun group followed by a *to*-infinitive (e.g. *I do not* suggest students to use

credit cards); rather, this pattern, where a node word is followed by a noun followed by a *to*-infinitive, occurs with ADVISE (*I do not* advise students to use *credit cards*). Then, the teacher can introduce the PDEV entry for ADVISE (pattern 1) to the students, as shown in Figure 9.18. The teacher can talk them through the description of ADVISE pattern 1 and encourage them to notice the elements that come before and after the verb in that pattern, *Human 1 or Institution 1 advises Human 2 or Institution 2 to-infinitive*. The teacher should explain what these semantic types and the *to*-infinitive mean, referring to instances in Figure 9.19 as examples of such conventions.

In addition, if the teacher would like to show that the divergent instances produced by the students in Figure 9.17 can be amended and correctly used with the verb SUGGEST, they can could use the PDEV entry for SUGGEST, as in Chapter 7. An amended entry is shown in Table 9.4 below:

Table 9.4 An amended entry for SUGGEST (pattern 1)

Pattern number	Pattern and implicature
1	**Pattern:** Human or institution suggests *that*-clause
	Implicature: Human puts forward *that*-clause as a plan or proposition for someone to think about
	Example: He suggested that we should get together and do some writing

This table shows that SUGGEST (pattern 1) is followed by a *that*-clause, and the semantic types that populate the subject slot are Human or Institution.

The teacher would then revise students' instances in Figure 9.17, where SUGGEST is the node word, following pattern 1 of SUGGEST as presented in Table 9.4. The following are instances after correction:

Figure 9.20 Corrected instances of SUGGEST

1	people	suggested	**that the government should legalize soccer betting.**
2	I	suggest	**that mass media and government should provide more guidelines**
3	I do not	suggest	**that students should use credit card.**
4	I	suggest	**that the government should do more promotions on recycling.**
5	I do not	suggest	**that the student should use it.**
6	He also	suggests	**that the government should leave the decision of banning to the industry.**

In this figure, the elements that come after SUGGEST have been paraphrased so that the pattern matches the description of SUGGEST pattern 1 in Table 9.4. For example, instance 4 has been rephrased from *So, I do not suggest* <u>*the students to use credit cards*</u> to *So, I do not suggest* <u>*that the students should use credit cards*</u>. The paraphrasing involves the elements that come after *suggest*, as underlined. Before paraphrasing, SUGGEST was followed by a noun group followed by a *to*-infinitive; after paraphrasing, suggest is followed by a *that*-clause.

The teacher would present these changes to the students by providing the students with the PDEV entry for SUGGEST (pattern 1), as shown in Table 9.4, along with the two sets of instances in Figures 9.17 (as version 1) and 9.18 (as version 2) (i.e. sets before and after paraphrasing). The teacher would explain to the students that version 1 includes instances produced by the students while version 2 is the corrected version. The teacher would then stress that SUGGEST cannot be followed by a noun followed by *to*-infinitive, but that one correct usage of SUGGEST is followed by a *that*-clause. Next, the teacher would ask the students to look at the PDEV entry for SUGGEST, as shown in Table 9.4, and encourage

them to observe the pattern (e.g. Human or institution suggests *that*-clause) and match it with the paraphrased instances in Figure 9.20. In addition, the teacher would encourage the students to read the implicature of pattern 1 of SUGGEST and look for it in the paraphrased instances. This is important because the main elements that are presented in the pattern (e.g. semantic types such as Human or Institution and grammatical-functional elements such as *that*-clause) are reflected in the implicature. By doing this, the students will thus link the pattern with the implicature. This in turn would reinforce the form–meaning (or pattern– implicature) pairing of pattern 1 of SUGGEST. In other words, the students will not only observe the pattern (syntax), but also know the meaning of the given pattern (semantics).

In addition, the comparison between the instances of use of the verb, SUGGEST, in the BNC and ICLE-Chinese shows another divergent usage of SUGGEST in the ICLE-Chinese: **Human suggests* to-*infinitive* and **Human suggests Human* to-*infinitive*. If the teacher wishes to address this phenomenon in the classroom, they can provide the learners with this list of divergent usages, as in Figure 9.22, along with a corrected version of the same list, as shown in Figure 9.21. The teacher can then review or discuss the lists and the PDEV with the learners.

Figure 9.21 Correct uses of the divergent instances of SUGGEST

1	crossing, the SAR **Government**	Suggests	lengthen**ing** the operation
2	time. The **government**, therefore,	suggested	construct**ing** another rail
3	in Hong Kong, some **developers**	Suggest	develop**ing** the country national park
4	The Hong Kong **government** has	suggested	import**ing** professionals
5	Recently, the **government** has	suggested	establish**ing** a law of
6	› solve this problem, the **government**	suggested	legaliz**ing** soccer betting

7	In my opinion, **I** also	suggest	us**ing** recycling to handle
8	few years; therefore, some **people**	suggest	construct**ing** a second railway
9	In my opinion, **I** would	suggest	develop**ing** the countryside

Figure 9.22 **Divergent instances of SUGGEST in the ICLE-Chinese**

1	crossing, the SAR **Government**	suggests	to lengthen the operation
2	time. The **government**, therefore,	suggested	to construct another rail
3	in Hong Kong, some **developers**	suggest	to develop the country park
4	The Hong Kong **government** has	suggested	to import professionals
5	Recently, the **government** has	suggested	to establish a law of
6	› solve this problem, the **government**	suggested	to legalize soccer betting
7	In my opinion, **I** also	suggest	to use recycling to handle
8	few years; therefore, some **people**	suggest	to construct a second railway
9	In my opinion, **I** would	suggest	to develop the countryside

By doing this, the teacher can raise the learners' awareness of the phenomenon in question. Since this way of raising awareness is mainly focused on the pattern of SUGGEST (i.e. followed by the *-ing*-clause), the learners will still need to know the meaning of SUGGEST in this particular pattern. Hence, showing the PDEV entry for SUGGEST pattern 1 to the learners enables them to understand the interaction between form and meaning, since the implicature is anchored to the pattern, as shown in the PDEV entry for SUGGEST pattern 1 below.

Table 9.5 The PDEV entry for SUGGEST (pattern 1)

Pattern number	Pattern and implicature
1	**Pattern:** Human suggests -ing-clause
	Implicature: Human puts forward -ing-clause as a plan or proposition for someone to think about
	Example: Some people suggest constructing a second railway

As can be seen, the PDEV entry presented in Table 9.5 displays the interaction between pattern and implicature (i.e. form and meaning), whereby the Human and *-ing*-clause appearing in the implicature are reflected in the pattern, as highlighted. This phenomenon is also illustrated by providing an example from the same entry.

For example, Figure 9.22 suggests that the presented instances have no objects. Providing the learners with the implicature of pattern 1 as presented in the PDEV will enable them to learn that what comes after SUGGEST is regarded as an object. To make this clearer, I provide an amended PDEV entry for SUGGEST pattern 1 as follows:

Table 9.6 An amended PDEV entry for SUGGEST pattern 1 as

Pattern number	Pattern and implicature
1	**Pattern:** *Human* suggests *-ing*-clause
	Implicature: *Human* puts forward *-ing*-clause as a plan or proposition for someone to think about
	Example: Some people suggest constructing a second

railway

This table shows that the implicature is anchored to the pattern by the *-ing*-clause, which is presented in both the pattern and the implicature. Most importantly, the implicature shows that there is an implicit agent in the element that comes after SUGGEST, that is, Human. For instance, in the example, *Some people suggest constructing a second railway*, the meaning is that, following the implicature of pattern 1 in the table, someone puts forward the idea of constructing a second railway for someone (e.g. the government) to think about. In this way, the learners will realize that there is a Human suggesting to another Human an idea for consideration.

Regarding the non-divergent usages of SUGGEST identified in the ICLE-Chinese corpus, the frequencies of the non-divergent usages of SUGGEST in the ICLE-Chinese are compared to those in the PDEV entry for SUGGEST. The results of the comparison are shown below.

Table 9.7 Percentage frequencies of the non-divergent patterns of SUGGEST in the ICLE-Chinese, BNC and LOCNESS

Pattern	ICLE-Chinese	BNC (sample)	LOCNESS
Pattern 1	25%	23%	19%
Pattern 2	26%	25%	33%
Pattern 3	**17%**	**52%**	**39%**
Pattern 4	0%	0%	0%

This table shows that pattern 3 is the dominant pattern in each of the two normative corpora, that is, the BNC and LOCNESS, whereas patterns 1 and 2 are the dominant patterns in the

learners' corpus. As discussed earlier, a possible explanation of this difference is the different semantic types that populate the subject slot of each pattern, as shown in Table 9.8 below. This table shows that pattern 1 and 2 share the same semantic type, Human, but pattern 3 does not.

Table 9.8 The non-divergent patterns of SUGGEST in the BNC sample

Pattern	Semantic type populating the subject slot
Pattern 1	Human
Pattern 2	Human
Pattern 3	Eventuality

If the teacher wishes to address this phenomenon, they might show the PDEV entry for SUGGEST pattern 3 to the learners along with simple examples of pattern 3 from the BNC. The teacher may also ask them to note what comes before and after *suggest* and then ask them to observe the implicature of pattern 3 that is recorded in the PDEV entry.

In addition, Table 9.8 above can be shown to the writers of the curriculum materials. Based on the statistics provided in this table, they may design tasks that focus more on the use of pattern 3 and incorporate them in the in-house materials. Careful attention to clarity and comprehensibility when selecting examples from the BNC is highly recommended.

Another phenomenon found relates to how patterns 1 and 3 both have sub-patterns, as shown respectively in Tables 9.9 and 9.10 below. In each pattern, SUGGEST may be followed by more than one different element. Each of these sub-patterns is identified in the ICLE-Chinese

and their respective proportion rates are counted. The results are presented below for easy

reading and understanding.

Table 9.9 Proportions of the sub-patterns of SUGGEST in the ICLE-Chinese (pattern 1)

No.	PDEV	%Raw frequency in the ICLE-Chinese
Pattern 1	A1. Human suggests **Action, or Plan or Proposition**	27%
	B1. Human suggests *that*-clause	**50%**
	C1. Human suggests -*ing*-clause	21%
	D1. Human suggests Quote	**0%**

One striking phenomenon is observed in Table 9.9: variation exists in the frequencies within

the pattern (i.e. the frequencies across the sub-patterns are not the same). Sub-pattern B1 is

the dominant one (50%), whereas sub-type D1 has no occurrence (0%). This information is

useful for the teacher to identify which sub-pattern is used more frequently than the others

(e.g. here, sub-pattern B1) and which, if any, are not used at all (e.g. sub-pattern D1). The

teacher may wish to provide the learners with this table and draw their attention to these two

phenomena. Since the meaning of all these sub-patterns is still the same, according to the

PDEV entry (i.e. *Human puts forward Action, or Plan or Proposition, or* that-*clause, or -*ing-

clause or Quote as a plan or proposition for someone to think about), the teacher may

emphasise this point and simultaneously provide the learners with a list of concordance lines

that represent the use of each of these sub-patterns.

Similarly, the percentage frequencies of the sub-patterns of pattern 3 are not the same in the ICLE-Chinese, as shown in Table 9.10 below.

Table 9.10 Proportions of the sub-patterns of *suggest* in the ICLE-Chinese (pattern 3)

No.	PDEV	%Raw frequency in the ICLE-Chinese
Pattern 3	A3. Eventuality 1 suggests **Eventuality 2**	100%
	B3. Eventuality suggests *that*-clause	**0%**

Table 9.10 shows that sub-pattern B3 scores 0% in the ICLE-Chinese. A possible explanation may be that the learners think that a *that*-clause is restricted to the pattern where the semantic type, *Human*, fills the subject slot of SUGGEST (e.g. ***He suggested that we visit London***); thus, they solely rely on the use of sub-pattern A3. It can therefore be useful if the teacher follows the same steps mentioned above to raise the learners' awareness of the existence of sub-pattern B3.

9.2.5 ALLOW

A similar phenomenon to that observed above, of using the wrong verb with a pattern that belongs to another verb, is found with ALLOW, as follows:

Figure 9.23 Divergent instances of ALLOW produced by the learners

1	gliffs can be 'painted' will	allow	**developers** <u>create</u> things like two-dimensional
2	Access Manager 200, which	allows	**users** <u>secure</u> single point of login
3	changing those structures by	allowing	**them** <u>make</u> such changes directly through
4	Our data do not	allow	**us** <u>determine</u> whether HBIG is beneficial

5	, he determined never to	allow	**anyone** <u>develop</u> power independent of the
6	I can't see him	allowing	**me** <u>take</u> the children so far away from

In this figure, the students use the wrong verb, ALLOW, with a pattern that belongs to another verb, that is, LET. Unlike ALLOW, LET can be followed by a noun group followed by a bare infinitive (i.e. without *to*) (e.g. *Our data do not <u>allow</u> us determine whether HBIG is beneficial* vs. *Our data do not <u>let</u> us determine whether HBIG is beneficial*).

Two alternatives can be used by the teacher to correct this divergent usage of *allow*. First, the teacher can reproduce all instances in Figure 7.21 with the node word replaced with LET, as shown below:

Figure 9.24 Divergent instances of ALLOW repaired with LET

1	gliffs can be 'painted' will	let	**developers** <u>create</u> things like two-dimensional
2	Access Manager 200, which	lets	**users** <u>secure</u> single point of login
3	changing those structures by	letting	**them** <u>make</u> such changes directly through
4	Our data do not	let	**us** <u>determine</u> whether HBIG is beneficial
5	, he determined never to	let	**anyone** <u>develop</u> power independent of the
6	I can't see him	letting	**me** <u>take</u> the children so far away from

Then the teacher can provide the students with the two versions and explain to them that LET is the right verb in these instances (as it can be followed by noun group followed by a bare infinitive.

Second, another alternative usage to the divergent instances in Figure 7.21 is found in the PDEV entry for ALLOW, as shown below:

Figure 9.25 PDEV entry for ALLOW (pattern 1)

1 Pattern: Human 1 *or* Institution 1 *or* Eventuality allows Human 2 *or* Institution 2 *or* Self TO-INFINITIVE
Implicature: Human 1 *or* Institution 1 *or* Eventuality gives Human 2 *or* Institution 2 *or* Self the opportunity, time, or permission TO/INF [VERB]
Example: the government allows certain individuals affected by the wrong to the public interest to bring action to redress the wrong.

This figure shows pattern 1 of ALLOW as recorded in the PDEV entry. Pattern 1, as seen in Figure 9.23, is followed by a noun group followed by a *to*-infinitive, and the semantic type that populates the subject slot can be Human, Institution, or Eventuality (e.g. *A 'browsing' system allows the user to explore the whole direction of facilities*). This pattern can be used to reproduce the divergent instances in Figure 9.21, as shown below:

Figure 9.26 Corrected version of learners' productions for ALLOW pattern 1

1	gliffs can be 'painted' will	allow	developers **to** create things like two-dimensional
2	Access Manager 200, which	allows	users **to** secure single point of login
3	changing those structures by	allowing	them **to** make such changes directly through
4	Our data do not	allow	us **to** determine whether HBIG is beneficial
5	, he determined never to	allow	anyone **to** develop power independent of the
6	I can't see him	allowing	me **to** take the children so far away from

In this figure, the *to*-infinitive has been added in the concordance lines, as shown in bold. This figure can be presented to the students along with the original, so that they notice the changes. In addition, the teacher can draw the students' attention to the fact that if they want to say that someone gives someone else the opportunity, time, or permission to do something, they should use pattern 1 of ALLOW, as in Figure 9.24 above.

9.3 Non-divergent uses of five ALLOW patterns produced by Swedish learners: Some pedagogical implications

Since most of the discussion in the previous subsection has focused on divergent (including non-divergent usages of SUGGEST) instances produced by Chinese, this section will focus on non-divergent patterns identified in the Swedish sub-corpus. The focus on non-divergent as opposed to divergent patterns is because Chapter 6 only focuses on the over-/under-representation of non-divergent ALLOW in the ICLE-Swedish as divergent instances were too few to investigate. This section starts with patterns that are found in both corpora, ICLE-Swedish and LOCNESS: patterns 4 and 1.Then it moves on to patterns that are found in ICLE-Swedish but not in LOCNESS: patterns 2, 3 and 7, respectively.

As was discussed in Chapter 6, the results show that some patterns of ALLOW are under-represented in ICLE-Swedish relative to the native-speaker writers, as shown in Table 9.5 below:

Table 9.11 Log-likelihood values of verb patterns of ALLOW

	BNC	ICLE-Swedish		LOCNESS			
		RF.	NL.	RF.	NL.	Over-/under-representation	Log-likelihood
pattern 1	54.8%	52	25.99	149	46.20	-	**13.85**
pattern 2	4.4%	0	0	14	4.34	-	**13.51**
pattern 3	7.6%	0	0	7	2.17	-	**6.76**
pattern 4	17.2%	3	1.49	66	20.46	-	**44.79**
pattern 5	2.0%	0	0	0	0		
pattern 6	2.8%	0	0	2	0.62	-	1.93
pattern 7	8.0%	0	0.99	15	4.65	-	**6**

pattern 8	1.6%	0	0	0	0		

This table can be used by the teacher in several ways. First, the teacher can find patterns that

are underrepresented in the ICLE-Swedish sub-corpus and see what alternatives are offered in

the PDEV entry. For example, pattern 4 is found to be under-represented in ICLE-Swedish in

relation to LOCNESS. The description of pattern 4 as per PDEV is shown in Figure 9.25

below:

Figure 9.27 PDEV entry for ALLOW (pattern 4)

4 *Pattern:* Human or Eventuality 1 or Institution **allows** Eventuality 2
Implicature: Human or Institution or Eventuality 1 gives the opportunity, time, or permission for Eventuality 2 to occur
Example: there are a number of people who, though fully Christian, have never allowed justification by faith to be an experiential reality in their life.

Only 3 instances of pattern 4 are found in ICLE-Swedish. Meanwhile, however, pattern 1 is

found to be frequently used by the Swedish learners (52 times out of 57). Pattern 1 is shown

in Figure 9.26 below:

Figure 9.28 PDEV entry for ALLOW (pattern 1)

1 *Pattern:* Human 1 or Institution 1 or Eventuality **allows** Human 2 or Institution 2 or Self TO-INFINITIVE
Implicature: Human 1 or Institution 1 or Eventuality gives Human 2 or Institution 2 or Self the opportunity, time, or permission TO/INF [VERB]
Example: the government allows certain individuals affected by the wrong to the public interest to bring action to redress the wrong.

The teacher would encourage the students to paraphrase some of their instances by changing

them to pattern 4. The following are some Swedish students' instances of pattern 1:

9.2 *I advocate totally free speech because I think it is a human right to be allowed <u>to</u>*

<u>express your opinions</u>. (ICLE-Swedish)

9.3 *They should certainly be allowed <u>to practice their religion</u>.* (ICLE-Swedish)

In these two examples, ALLOW is in passive voice and followed by a *to*-infinitive. The teacher can provide the students with rephrased versions of these two examples using pattern 4, as shown below:

9.4 *The government should allow <u>the freedom of speech</u>.*

9.5 *The Swedish government should allow <u>the practice of multi-religion</u>.*

In these two examples, ALLOW is followed by what Halliday (1993) refers to as nominalized phrases (e.g. *the freedom of speech*; *the practice of multi-religion*). The teacher can encourage the students to practice this operation of nominalizing a *to*-infinitive that comes after ALLOW. Encouraging students to use pattern 4 will make their language more sophisticated.

The students can be given a copy of the PDEV entry for *allow* (pattern 4) (see Figure 9.25) along with selected concordance lines that represent pattern 4, and asked to realize the pattern in the concordance lines provided. Some instances that represent pattern 4 of ALLOW are presented in Figure 9.27, taken from the PDEV entry for ALLOW (pattern 4):

Figure 9.29 Instances of pattern 4 of ALLOW taken from the PDEV entry

improvements in the rail network ,	allowing	**more <u>movement of goods</u> and passengers by**
transfer tax. In 1990 the Gift Aid Scheme	allows	**<u>tax relief</u> on single cash gifts to charities**
straps between alternate segments which	allowed	**<u>the achievement of still greater power</u>**
of the engine altered the budgeting and	allowed	**<u>the building of a new model</u>, the Senior**
democracies, within France, governments	allowed	**<u>the emergence of a private</u>, alongside the**
newly-cut coupe for about seven years to	allow	**<u>the establishment of new seedlings</u>. The**
fundamental causes of deforestation and to	allow	**<u>the inclusion of forest-dwellers</u>**

educationally most important. Obviously it   allows   **the production of a very large number of**

the demand for private farmers to be   allowed   **the use of firearms to combat rising crime**

In these instances, ALLOW is followed by lexical items that belong to the semantic type Eventuality, as underlined. The teacher can use these instances along with the PDEV entry for ALLOW pattern 4 (Figure 9.25) to raise students' awareness of the existence of this pattern (i.e. *Human or Eventuality or Institution allows Eventuality* 2).

The teacher would distribute copies of this set of instances (Figure 9.27) and ask the students to notice the underlined phrases. They would then explain to the students that these phrases represent the semantic type Eventuality 2 in the object slot of ALLOW pattern 4. The students then would be asked to look at the PDEV entry of ALLOW pattern 4 (seen in Figure 9.25) and read the description of the pattern. The teacher can pick up any of the instances in Figure 9.27 and ask the students to realize Eventuality 2 in the instances. For example, the teacher could select one of these instances, say, *He supported the demand for private farmers to be allowed the use of firearms to combat rising crime against them*, and rephrase it to pattern 1 of ALLOW, following the PDEV entry for ALLOW. The following are examples before and after paraphrasing:

> 9.6 *He supported the demand for private farmers to be allowed the use of firearms to combat rising crime against them*. (pattern 4) (PDEV)
>
> 9.7 *He supported the demand that allows private farmers to use firearms to combat rising crime against them*. (pattern 1)

Here, example 9.6 represents pattern 4 of ALLOW, where it is followed by Eventuality, that is, *the use of firearms*, whereas example 9.7 represents pattern 1, where ALLOW is followed by a noun followed by a *to*-infinitive.

The teacher would apply the same procedure to the rest of the instances. They could also use instances taken from other native-speaker corpora such as the Bank of English corpus. The degree of difficulty of examples would depend on what level of students they were working with. Instances suitable for advanced learners would not be suitable for beginners, for example.

Another phenomenon observed in ICLE-Swedish is that the Swedish learners tend to prefer the use of Human to non-Human or Eventuality in the subject slot of pattern 1. This is not necessarily an error, but a matter of semantic interpretation and emphasis. The following examples represent this phenomenon:

> 9.8 *The teacher does not allow people to use their mobile phones during exams.*
> 9.9 *The exam regulations do not allow students to use their mobile phones during exams.*

The lexical item that fills the subject slot in example 1 is Human (*The teacher*), while the lexical item that populates the subject slot of example 2 is an Eventuality (e.g. *The exam regulations*). Both examples are correct and meaningful; however, there are cases where example 1 would be a better choice than example 2, and vice versa, depending on the event and context.

One possible pedagogical application to encourage Swedish learners to use an Eventuality in the subject slot of pattern 1 might involve selecting some students' instances of pattern 1 and replacing the lexical items belonging to the Human category with those that belong to Eventuality. The following is one of the Swedish learners' examples instantiating pattern 1:

> 9.10 *Swedish customs are permitted to stop you if they suspect you, and they do not even have had to seen you earlier. <u>They are allowed to stop anyone.</u>* (ICLE-Swedish)

The underlined phrase is the target one, which needs to be amended so that it instantiates pattern 1, as follows: *Eventuality allows Human* to-*infinitive* (see Figure 9.26 for the PDEV description of pattern 1). The following is the underlined phrase after amendment:

> 9.11 *Swedish customs are permitted to stop you if they suspect you, and they do not even have had to seen you earlier. <u>The law allows them to stop anyone.</u>*

Here, the pronoun *they* (Human) that populates the subject slot of ALLOW has been replaced with *The law* (Eventuality), and the whole phrase has be changed to active voice, in that *allows* is followed by pronoun *them* followed by to-infinitive.

Another activity that Table 9.5 shows that the Swedish students could benefit from is looking at the patterns of ALLOW that are found in the LOCNESS but not found in the ICLE-Swedish sub-corpus. For example, 14 instances of pattern 2 of ALLOW are found in LOCNESS, while no occurrence is found in ICLE-Swedish. The description of pattern 2 of ALLOW is provided in Figure 9.28 below.

Figure 9.30 PDEV entry for ALLOW (pattern 2)

2 Pattern: Human 1 or Institution 1 or Eventuality allows Institution 2 or Self or Human 2 Privilege
   Implicature: Human 1 or Institution 1 or Eventuality gives Human 2 or Institution 2 or Self permission or the opportunity to have Privilege
   Example: Another great spiritual advantage was that the people's occupation allowed them the opportunity to read good books or to engage in godly conversation

Figure 9.28 shows the PDEV entry for ALLOW (pattern 2). In this pattern, ALLOW is followed by two semantic types: Human and Privilege. I would argue that the absence of this pattern in ICLE-Swedish might be due to students being unaware of such a pattern or to the difficulty of this pattern, in that in it ALLOW is followed by two noun groups. In other words, it could be due to the lack of a *to*-infinitive as in pattern 1 (e.g. *The teacher allows the students to have a break*). The PDEV entry for *allow* would be useful to help students learn more about and become more familiar with pattern 2, particularly by the use of the ontology, as shown in Figure 9.29 below.

Figure 9.31 A screenshot of the PDEV ontology for semantic type Privilege[28]

This figure shows the hierarchical ontology of the PDEV. On the far right of this figure, there is a box including several lexical items that belong the semantic type Privilege. The box lists items such as *privilege*, *rights*, *access*, *freedom*, and others. These words would be useful for

---

[28] Taken from http://pdev.org.uk/#onto.

the learner to practice the use of pattern 2 (see Figure 9.29). The following are examples of

pattern 2 taken from the BNC:

9.12 *It is very rare for a farmer to allow <u>us</u> **access** to his fields once he has seeded them.*
   (BNC)

9.13 *He allowed <u>her</u> more **freedom** of choice than many of his clients.* (BNC)

9.14 *The purpose of this is usually to allow <u>students</u> the **opportunity** to use language they*
   know in a less controlled situation. (BNC)

9.15 *In fact sections 3 and 4 allow <u>him</u> quite generous **rights** of resale.* (BNC)

It is important to mention that of these four examples, only one was taken from the annotated

concordance lines of the PDEV entry for ALLOW (pattern 2). The rest of the examples were

taken from the full BNC, for two reasons. First, there are only 11 instances of pattern 2 in

PDEV. Second, most of these 11 instances are too difficult to be handled by students.

Going back to these four included examples, the teacher would provide the students with a

copy of this set of examples along with the PDEV entry for ALLOW (pattern 2), as shown in

Figure 9.28. He would then explain that these four examples represent pattern 2 of *allow* as

shown in Figure 9.28, stressing that, in this pattern, ALLOW is ditransitive (*allows <u>him</u>*

<u>freedom of choice</u>). Then he would draw their attention to the lexical items that fill the direct

object slot of ALLOW, as highlighted in the examples, and explain to them that these words

belong to the semantic type Privilege.

Assuming that the students and the teacher had been trained to use the PDEV, it would then

be useful to get the students to use the online version of ontology and search for other lexical

items that belong to the semantic type Privilege, as presented in Figure 9.29, and to try to put

them into meaningful sentences using pattern 2 of ALLOW, employing the description of the

pattern (see Figure 9.28).

It should be noted here, however, that not all lexical items that are listed in the ontology under

Privilege can be fitted to pattern 2 of ALLOW, as some of these items may look vague. In

addition, not all lexical items listed under Privilege are found in the annotated concordance

lines in the PDEV entry for ALLOW pattern 2. Therefore, the whole process of the search

should be done under the teacher's supervision, since some students may encounter some

difficulty working it out. In addition, the teacher could produce materials that are based on

PDEV but not exactly the same as it.

Another pattern of *allow* that is found in LOCNESS and PDEV/BNC but not in the ICLE-

Swedish sub-corpus is pattern 3 (see Table 9.5). The description of pattern 3, according to the

PDEV entry, is shown below:

Figure 9.32 PDEV entry for ALLOW (pattern 3)

This figure shows that ALLOW can be followed by noun (physical object) followed by *to be*

V-*ed* (e.g. *The banking service allows funds to be deposited in numerous foreign currencies*).

The lack of instances of pattern 3 in the ICLE-Swedish sub-corpus is interesting because most

of the instances that instantiate passive voice identified in the sub-corpus belong to pattern 1

(e.g. *The Indians were not allowed to buy liquor for about one hundred years*). Moreover,

some of these instances, identified in the ICLE-Swedish sub-corpus, can be paraphrased to instantiate pattern 3. These instances are as follows:

Figure 9.33 Paraphraseable instances of ALLOW (pattern 1)

1	should we	allow	them to keep their culture
2	There are at least three endings which	allow	the reader to draw his or her own conclusion.
4	The situation in Sweden is that they are	allowed	to practice their own religion whatever it is.
7	neo-Nazism if they are	allowed	to express their views.
8	The Indians were not	allowed	to buy liquor for about one hundred years
9	Moreover, they are not	allowed	to see their relatives.

Figure 9.31 shows some instances of ALLOW (pattern 1) produced by the Swedish learners. Most of the instances are in passive voice (e.g. *They are not allowed to see their relatives*), while others are in active voice (e.g. *Others read and then allow the stars to govern their lives*).

The teacher can use these instances in several ways to raise the students' awareness of the existence of pattern 3, which is not found in ICLE-Swedish. These instances can be paraphrased so that they instantiate pattern 3; the same instances shown in Figure 9.31 are shown below, after paraphrasing:

Figure 9.34 Paraphrased instances of ALLOW (pattern 1→pattern 3)

1	Should we	allow	**immigrants' culture to be kept**?
2	There are at least three endings which	allow	**reader's conclusion to be drawn**.
3	The Swedish government	allows	**immigrants' religion to be practiced** whatever it
7	fighting neo-Nazism if the government	allows	**their views to be expressed**.

237

8	The law did not	allow	**liquor to be bought by Indians** for about one
9	Moreover, they do not	allow	**their relatives to be seen.**
10	Others read and then	allow	**their lives to be governed.**

As seen in Figure 9.32, all of these instances together conform to pattern 3 (*Human or Institution allows Physical Object to be* V-*ed*). Instances that were in passive voice have been changed into active voice by adding a subject at the beginning of the sentence, and then paraphrased so that they instantiate pattern 3. The dominant changes apply to the elements that come after ALLOW, as highlighted in Figure 9.32. For example, in instance 2, *There are at least three endings which allow <u>the reader to draw his or her own conclusion</u>*, the noun phrase *reader's conclusion* is followed by *to be* followed by *drawn*, as follows: *There are at least three endings which allow <u>the reader's conclusion to be drawn</u>*. It should be acknowledged that there are cases where pattern 1 would be a better choice than pattern 3. Nonetheless, it would be useful for the students to practice an alternative use, pattern 3, that would make their language more sophisticated.

This activity can be used as an exercise in the classroom in two ways. First, the teacher can provide the students with two versions of the instances (before and after paraphrasing); in this case, Figure 9.31 (before paraphrasing, pattern 1) would be version 1 and Figure 9.32 (after paraphrasing, pattern 3) would be version 2. Then, the students would be asked to consider the two versions and compare between them. The teacher would ask questions such as 'Can you tell the difference between the two versions?' or 'Can you find the differences particularly in those elements that follow *allow* in each of the two versions?' The teacher would next talk the students through the two versions and elaborate more on these differences in pattern and meaning.

Another way of getting the students to practice paraphrasing can be employed. The students

would be given version 1 (instances produced by the students) along with the PDEV entry for

ALLOW pattern 3, as shown in Figure 9.30. Then the teacher would pick up an instance from

version 1 and paraphrase it following the description of pattern 3. Paraphrasing should be

done step by step and the teacher should refer back to the description of pattern 3 while

paraphrasing. This is important since the students will be paraphrasing the rest of the

instances in version 1. To do so, they will be given the PDEV entry for ALLOW pattern 3, as

shown in Figure 9.30, along with version 1, including instances that instantiate pattern 1, as

shown in Figure 9.31. Next, the students will be asked to paraphrase these instances,

following the description of pattern 3; however, before that, as some conventions presented in

the description of pattern 3 might be new to the students (Figure 9.30) (e.g. Physical object, *to*

be V-*ed*, etc.), the teacher should explain to the students what these elements mean. Then the

students can work the instances out under the teacher's supervision.

Another pattern of ALLOW that is not found in ICLE-Swedish is pattern 7. The description

of pattern 7 is provided below:

Figure 9.35 PDEV entry for ALLOW (pattern 7)

As shown, ALLOW here is followed by preposition *for* followed by Eventuality (noun

group); and the semantic type that populates the subject slot is Human or Eventuality (e.g. *But*

once <u>we</u> <u>allow for</u> <u>the shuffling of genes</u>, *there is a whole new set of possibilities*). It is not

surprising that the Swedish learners use no instances of pattern 7. I would argue that the

reason behind this avoidance of pattern 7 is that, unlike pattern 1, pattern 7 lacks a *to-*

infinitive. Encouraging Swedish learners to use pattern 7 of ALLOW would make their language more sophisticated. The teacher could select a set of instances that instantiate pattern 7 from a native-speaker corpus and present it to the students along with the PDEV description of ALLOW pattern 7, as shown in Figure 9.33. The following is a selected set of instances of ALLOW pattern 7 taken from the full BNC:

Figure 9.36 Instances of ALLOW pattern 7 (BNC)[29]

This is too restricted to	allow	**for** much ***expansion***
This will	allow	**for** the ***co-ordination*** of school development
professional readership would	allow	**for** informed ***debate*** of issues amongst teachers
which	allows	**for** the ***creation*** of new terms,
faire approach (which	allows	**for** any ***combinations*** of dishes, even the
in particular to	allow	**for** the ***creation*** of an AACR2-conformant

All these instances together are followed by *for* followed by Eventuality (noun group). This set of instances can be distributed to the students along with the PDEV entry for ALLOW pattern 7 (see Figure 9.33). The students then can be asked to observe the pattern in all of these instances, particularly *for* and the nouns that follow it (underlined in Figure 9.34). Meanwhile, the teacher can explain to the students that *allow* can be followed by *for* followed by Eventuality, as presented in Figure 9.33. The teacher can then draw students' attention to the PDEV description in Figure 9.33 and let them look at these grammatical and semantic categories in the description. The teacher should explain to the students that the doubled-underlined words in Figure 9.34 refer to Eventualities. He then asks them to read the

---

[29] The doubled-underlined words are meant to represent Eventuality, following PDEV concept.

implicature of pattern 7. He would elaborate more on this implicature, since the phrase *make provision for*, as mentioned in the implicature, would be difficult for the students to understand. He could pick up any of the instances in Figure 9.34, for example, and say to them, explaining the meaning of pattern 7, that 'if you *allow for* certain problems or expenses, you include some extra time or money in your planning so that you can deal with them if they occur'[30]. He could then talk the students through these instances and show them how they are reflected in the PDEV entry for pattern 7 of ALLOW.

The teacher can also use examples to encourage learners to change pattern 7 to pattern 1. The following is an example of pattern 7:

> 9.16 *Human rights organizations opposed the emergency powers, which **allowed for** indefinite detention without trial* (Pattern 7). (BNC)

This example can be changed to pattern 1 as follows:

> 9.17 *Human rights organizations opposed the emergency powers, which **allowed the government to detain prisoners** without trial* (Pattern 1).

In this example, ALLOW is followed by a noun, *government*, followed by a *to*-infinitive with the verb, *detain.* Again, there are cases where the use of one pattern is better than the use of another. If we want to place the stress on who takes the action of detention, for instance, then pattern 1 (i.e. example 9.17) is better than pattern 7 (i.e. example 9.16). If, on the other hand, we want to obscure who took the action of detention and who were detained, then pattern 7 would be a better choice.

---

[30] This definition was obtained from www.collinsdictionary.com.

This practice of transforming one pattern to another would encourage learners to not only understand how to transform pattern 1 to pattern 7, but also see why the use of one pattern is better than the use of another.

It would therefore be beneficial for learners to practice transforming one pattern into another. In some cases, this would allow them to write with more sophistication; in other cases, this practice could demonstrate that one pattern is better in one context than another, for instance, if the transformation does not work.

9.4   Conclusion

This chapter has discussed the pedagogical implications of this thesis. The first part of this chapter focused on Chinese learners; the second part focused on Swedish learners. Various ways of presenting the outcomes as errors and clarify how to get to correct alternatives have been offered. However, in general these ways have in common the following approach: the teachers show the learners the PDEV entry and the BNC and learner corpora instances and talk them through the differences between the two. In some cases, the teachers will rewrite what the learners have produced. In addition, the teachers will ask the learners to conduct repairs.

Chapter 10 **Conclusion**

10.1 Introduction

This final chapter will first answer research questions 1 (section 10.2) and 2 (section 10.3),
respectively. It will then review the thesis and discuss the main conclusions and contributions
of this work. Next, it will go on to assess the limitations of this thesis, and some directions for
future research will be discussed. This chapter will close with some concluding remarks.

10.2 Can CPA/PDEV be used to describe learner corpora successfully?

This study has demonstrated that CPA/PDEV can be a successful method for describing
learner corpora. Sorting annotated concordance lines of each target verb in the target corpora
into 'semantically motivated syntagmatic patterns' (Hanks 2004:88), I was able to identify
several phenomena concerning uses of verb by the learners. All of these phenomena have
been discussed as research findings in Chapter 8; they are summarised here, as follows:

1. *Phenomenon 1:* the use of the wrong verb in a recognisable pattern (see
   Chapters 5 and 7);

2. *Phenomenon 2:* mismatch between form and meaning in some verb patterns
   (see Chapters 5 and 7);

3. *Phenomenon 3:* the use of the wrong semantic type to fill the subject slot of
   some verb patterns (see Chapter 5);

4. *Phenomenon 4:* lack of nominalization in some verb patterns (see Chapter 6);

5. *Phenomenon 5:* the preference for congruent words (Human) to non-congruent
   ones (non-Human) in some verb patterns relative to the normative language
   (see Chapter 6);

6. *Phenomenon 6:* the preference for particular patterns of a given verb over others within a corpus (see Chapters 6 and 7);

7. *Phenomenon 7:* the preference by learners for particular verb patterns over others relative to the normative language (see Chapters 6 and 7); and

8. *Phenomenon 8:* the use of a narrower range of verb patterns by learners relative to the normative language (Chapters 6 and 7).

The conclusions arising from phenomena 1–3 are reliable, as the analysis does not involve frequency comparison; they are the outcomes of qualitative analysis. On the other hand, phenomena 4–8 are the outcomes of frequency comparison. The use of PDEV as a tool for evaluating comparative frequency between learners and native speakers may not give wholly reliable information, because the target corpora used in this study are not perfectly comparable, as discussed in Chapter 4. The ICLE data are from a particular genre, that is, argumentative texts, while the data used to build the PDEV come from a general corpus, the BNC, that includes various genres. Thus, the influence of genre difference on the research findings of comparative analysis may affect the conclusions on comparison of verb patterns between the learners and the native speakers. The absence of a particular verb pattern in the learner writing, for instance, may be due to the task the learners were asked to complete (e.g. writing an argumentative essay) rather than a lack of knowledge on their part. Thus, the conclusions from the comparative frequency analysis, using PDEV as a measure of difference between learner corpora and native-speaker corpora, should be treated with caution.

In terms of the learners' preference for particular patterns of a given verb, this phenomenon was based on the concept of under-/over-representation, using PDEV and/or LOCNESS as a reference for comparison between learners and native speakers. Such a method is effective in

showing which patterns are more/less frequently produced by the learners relative to the native speakers. However, since the frequency of each verb in the PDEV entry is based on a limited sample (ranging between 250 and 500 concordance lines per verb), the data may not properly reflect the real frequency of use of verb patterns by native speakers. In addition, the PDEV is based on an old corpus, that is, the BNC.

The PDEV can successfully reveal learners' preferences for particular patterns of a given verb within a given corpus. For instance, we have seen, in Chapter 6, that the Swedish learners preferred the use of pattern 1 of ALLOW to pattern 3, producing instances such as *The government does not allow people to use weapons*. In this instance, *allow* is followed by a noun followed by a to-infinitive. This instance has an alternative in the PDEV entry, where ALLOW is followed by noun Eventuality, which would produce *The government does not allow the use of weapons*. Both instances have the same meaning, but with different structures: the former occurs with a *to*-infinitive (*to use weapons*), while the latter occurs with a nominalization (*the use of weapons*).

It is difficult to know why the learners prefer one pattern to the other, and the avoidance of one pattern does not necessarily imply a restriction in competence. There are instances where pattern 1 (e.g. *The government does not allow people to use weapons*) would be better than pattern 3 (e.g. *The government does not allow the use of weapons*), as well as vice versa. Regardless of that, although the Swedish learners produced perfectly accurate sentences, the fact that they did not use the alternative pattern suggests that they have a limited range of patterns at their disposal.

In addition, the CPA/PDEV successfully identifies the dominant patterns of a given verb within a corpus. For example, pattern 1 of SUGGEST was extensively produced by the Chinese learners relative to the other patterns of SUGGEST. In some cases, however, the use of the PDEV entry for comparison may hide some patterns identified in the learner corpora. In Chapter 7, for example, it was noted that pattern 1 of SUGGEST contains sub-patterns (i.e. different patterns having the same meaning) such as *Human suggests* that-*clause*, *Human suggests Eventuality*, or *Human suggests* -ing-*clause*. The analysis revealed that the learners extensively produced *Human suggests* that-*clause*, while *Human suggests Eventuality* was not produced at all, and a large number of instances of *Human suggests* -ing-*clause* were identified. In this case, the PDEV would have been more useful to learners if it divided patterns that had different forms into sub-patterns, showing the frequency of each one. The same suggestion is offered to researchers who use the concept of CPA/PDEV to describe learner corpora.

10.3 How useful is PDEV to learners?

This research question will be answered in relation to four themes, as follows:

1. Semantic types;

2. PDEV annotated concordance lines;

3. The description of patterns in the PDEV; and

4. The frequency of each pattern of a given verb.

Each of the above themes will be dealt with individually.

10.3.1 Semantic types

One of the key elements that distinguish PDEV from other resources is the concept of semantic types (see Chapter 2 for details). A corpus-based dictionary that appears to be

similar to the PDEV in that it includes patterns of a given verb in its entries is the on-line *Collins Cobuild English Dictionary*. However, it is a general dictionary, in that it defines all part of speech (e.g. noun, adjective, verb, adverb, etc.), whereas PDEV is restricted to one parts of speech, that is, verbs. Both on-line and hard-copy version of the *Collins Cobuild English Dictionary* define verbs using the *if someone or something* convention. The following are two examples of a definition of one verb pattern taken from the *Collins Cobuild English Dictionary* and its counterpart in the PDEV:

10.1 If <u>you</u> are *allowed* <u>something</u>, you are given permission to have it or are given it.
    (on-line *Collins Cobuild English Dictionary*)

10.2 Human 1 *or* Institution 1 *or* Eventuality *gives* <u>Institution 2</u> *or* <u>Self</u> *or* <u>Human</u> 2 <u>*permission or the opportunity to have* Privilege.</u> (PDEV)

Example 10.1 introduces a definition of a pattern of ALLOW taken from the on-line *Collins Cobuild English Dictionary*, while example 10.2 introduces the corresponding definition of ALLOW taken from PDEV entry for ALLOW (pattern 2). The underlined words in example 10.1 (*you* and *something*) are reflected in example 10.2 (*Institution 2* or *Self* or *Humans* and *Privilege*). In example 10.1, the underlined words are general or unspecific; in contrast, the corresponding ones in example 10.2 are specific in that *you*, in example 10.1, could mean three things: Institution, Self, or Human, and in that *something* is restricted to Privilege, as presented in the PDEV entry for ALLOW (pattern 2).

In this case, I would argue, the implicature of ALLOW in example 10.2 (the PDEV) is more beneficial to the learners than that in example 10.1 (the *Collins Cobuild English Dictionary*), for several reasons. First, the semantic types in example 2 open several choices for the learners, in that they become aware that the direct object slot of *allow* (pattern 2) can be filled

with lexical items that belong to three specific types of elements: Institution, Self, or Human. Similarly, the semantic type that fills the indirect object of *allow*, as in example 10.2, is narrowed down to one specific type, Privilege. More interestingly, each of these semantic types includes a pool of lexical items. This would be useful for learners, who could search the ontology of each of these semantic types and see what types of lexical items each semantic type can offer. In terms of the semantic type Privilege, for example, the learners can go online and search the entry for Privilege to see what other lexical items it includes. The learners can then choose some words from the entry for *Privilege* and practice using them with pattern 2 of ALLOW (see Chapter 6 for examples).

The learners can benefit from semantic types in other ways as well, as provided for instance in the following PDEV description of CLAIM:

> 10.3 Pattern: Human *claims* **Entity**
>
> Implicature: Human *asserts confidently that* Human *has received or is entitled to receive* **Entity = Benefit** *as a benefit*
>
> Example: *The wife also claimed <u>damages</u> for mental stress, vexation and strain caused by the defendants' negligence.* (BNC)

The highlighted words in the description above represent the semantic type that populates the object slot of *claim* (e.g. Entity = Benefit). The learners can benefit from the entry for semantic type Entity in that they can go online and search for lexical items that belong to Entity, assuming the learners have been trained to use the PDEV. Some of these lexical items are as follows: *benefit(s)*, *allowance*, *compensation*, *reward*, *right(s)*, etc. If the learners are not trained to use the online PDEV, the teacher can select some lexical items from the entry

for Privilege and present them to the learners. Then the teacher can explain to the learners that these words belong to the Entity semantic type, which populates the object slot of CLAIM (pattern 3), as shown in the description of pattern 3 above.

Although the use of semantic types in the PDEV is argued to be beneficial to the learners, it has some limitations. First, the language of semantic types might be unfamiliar to the learners. For example, some patterns of given verbs (e.g. *agree*, *lead*, *encourage*) have semantic types such as Proposition, Concept, or Eventuality as their object (e.g. *Human agrees with Preposition*; *Eventuality 1 leads to Eventuality 2*; *Human accepts Concept*). The learners may start to question what these words mean. In addition, some lexical items that belong to these semantic types are not straightforward; for example, the entry for Eventuality includes many words, including words such as *use*, *lack*, that cannot be used with verbs as they are but rather need to be used with the head noun of a noun group, such as *the use of a* <u>noun group</u> or *the lack of a* <u>noun group</u>, where the noun slot can be filled with different words (e.g. *the use of mobile phones*, *the use of the emergency door*, *the lack of resources*, *the lack of evidence*) to form meaningful sentences when used with verbs, as follows:

> 10.4 *<u>The lack of evidence</u> led to difficulty in identifying the suspect*. (BNC)
> 10.5 *The teacher did not allow <u>the use of mobile phones</u> inside the classroom.* (ICLE-
> Swedish)

In example 10.4 above, *lack* used in the noun group *the lack of evidence* qualifies the Eventuality, just as *use*, in example 10.5, does when used with the noun group *the use of mobile phones*. Without doing so, the meaning, and of course the pattern, of the two examples would not be complete.

All semantic types in the PDEV include only single words. This may confuse the learners, who may assume that these single words are used with verbs as presented in the ontology (i.e. as single words with no addition to them). Two possible solutions can be suggested. First, learners may refer to the sets of annotated concordance lines recorded in the PDEV entry, to which each verb pattern is linked. These sets include a number of instances, taken from the BNC, which represent a given pattern (see Chapter 2 for examples). The good thing about these instances is that the single words that belong to semantic types are no longer single words, but are used with the head noun of a noun group. Therefore, the learners will see these words in their correct and meaningful usage. The following is an example taken from the annotated concordance lines for the PDEV entry of LEAD (pattern 1):

Figure 10.1 Concordance lines of pattern 1 of LEAD (PDEV entry)

of UK exports and imports are likely to	**lead**	1	to the development of a UK balance of payments
the adaptive expectations hypothesis that	**led**	1	to the development of the rational expectations
through them and NRA officials hope it will	**lead**	1	to the development of new saltmarsh to
contact with the Minoan civilization which	**led**	1	to the development of the Phoenician civilization
their attitude to their role and this has	**led**	1	to the development by political scientists
James) Alfred Ewing [q.v.] and Thomas Gray,	**led**	1	to the development of the first true seismograph
symptoms of the disorder. This may well	**lead**	1	to the development of more effective drugs
morphological and structural evidence has	**led**	1	to the development of numerous models of

This set of instances represents pattern 1 of LEAD: Event*uality1 leads to Eventuality 2*. Looking at this figure shows the object slot of *lead* is populated with the head noun of a noun group (e.g. *the development of a UK balance of payments*). This would be beneficial for learners who wish to check how lexical items that belong to semantic types are used in a meaningful way. In addition, if these lines are too short to determine the meaning of a sentence, the learners will simply click on any of these incomplete citations and the full citation will pop up at the bottom of the screen, as shown below:
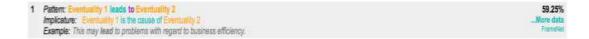
Figure 10.2 Concordance lines of pattern 1 of LEAD with full citation pop-up (PDEV entry)

10.3.2 The annotated concordance lines of PDEV

As mentioned in the previous section, the PDEV entry for each verb pattern is linked to a set of annotated concordance lines. This feature is useful for learners in several ways. First, all annotated concordance lines are corpus-based. This means that examples are not made up; rather, they are authentic examples taken from a native-speaker corpus, the British National Corpus. Second, while the learners are checking the description of a pattern of a given verb (e.g. *Eventuality 1 leads to Eventuality 2*), they do not need to take the trouble of looking for other resources to find examples. Rather, they can simply click at the far right of the pattern description, labelled *More data*, as shown in Figure 10.3 below:

Figure 10.3 PDEV entry for LEAD (pattern 1): *More data*

This will take the learner to a new window that includes a number of annotated concordance lines linked to the given pattern (see Figures 10.1 and 10.2 for examples).

Attention should be drawn here to several points. First, all annotated concordance lines in the PDEV were based on a randomly selected sample from the BNC. This selection was made electronically, in that the researcher gave orders to the concordance software to select 250 concordance lines, which the software then did. The outcomes of this automatic selection have two major limitations. First, some of the selected instances might be too difficult to be understood by the learners. This is because the BNC contains specialized texts in complex fields such as natural science, finance, and others, and the selected sample may include some such citations. For example, the following citation is taken from the PDEV entry for ENCOURAGE (pattern 3):

10.6 *Ironically, the Government* **encouraged** *the SERPS exodus because the projected costs of the scheme were too high*

Some words mentioned in the above example, such as *SERPS* and *exodus*, are likely to be too specialized for learners. *Exodus* is a relatively infrequent word, so learners are unlikely to have met it before. *SERPS* is an old term relating to pensions that very few people would understand nowadays. This issue is in this sense related to the increasingly elderly status of the BNC.

Issues like these could prevent learners from understanding the meaning of the example, which would result in failure to acquire the pattern. To overcome this problem, I would suggest that the BNC be replaced with another corpus which is current and contains non-

252

specialized texts. Another solution is that the teachers can use another corpus and select examples that are more relevant to their students and thus easier for them to understand. Another limitation of the annotated concordance lines recorded in the PDEV is that the citations are in different forms, such as passive voice or gerund. Forms like these may obscure the actual lexical items that fill subject slot of a given verb. To illustrate this phenomenon, an example is provided below:

> 10.7 *But in practice these rules were not properly enforced, **allowing** banks to 'pretend' that loans to state-owned enterprises were not loans to a 'country'.*
> (BNC)

This example represents pattern 1 of ALLOW, following the PDEV entry. The description of pattern 1 is: *Human 1 or Institution 1 or Eventuality allows Human 2 or Institution 2 or Self to-infinitive* (e.g. *The system allowed her to download several files*). In example 1, I argue, the subject of ALLOW is not made explicit, and so the learner may get confused about which of these three semantic types fills the subject slot of ALLOW (i.e. Human or Institution or Eventuality). In addition, some learners may be familiar with the usual usage of ALLOW, in which its subject is usually Human, thus avoiding the use of the other two types (e.g. Institution and Eventuality).

We have seen this phenomenon in Chapter 6, where most lexical items that fill instances of pattern 1 of ALLOW belong to Human (e.g. *The teacher allowed the students to use their mobile phones inside the classroom*); and very few belong to Eventuality (e.g. *The low taxes in Switzerland allowed the woman to stay at home*).

A similar phenomenon was observed with some usages of AVOID in the ICLE-Chinese sub-corpus. A number of divergent instances appeared (see Chapter 4) where lexical items that fill the subject slot of AVOID pattern 1 belong to Eventuality (e.g. *_Recycling can avoid much pollution_). This usage is incorrect because, according to the PDEV entry for AVOID pattern 1, it is not typical of pattern 1 of AVOID to have Eventuality as its subject. Rather, only two semantic types can fill the subject slot of AVOID, that is, Human or Institution.

Another observation is that the set of annotated concordance lines for each pattern entry is shuffled randomly. It would have been more useful for the learners if instances were sorted according to active and passive voice. This would make it easier for the learners if they wished to observe the pattern in a particular voice.

Therefore, if teachers want to use the annotated concordance lines from the PDEV entry as examples for their learners, they should take the above-mentioned limitations into consideration and be selective when using examples for their learners. The following points are suggestions for teachers using the annotated instances in PDEV:

1. A fairly large number of annotated instances in the PDEV entry are too difficult for the learners. These instances either include terms that are too difficult or sentences that are too long. Teachers can overcome this by being selective about the instances they adopt.

2. If teachers want to present the complete set of annotated instances of a pattern of a target verb, they should re-sort the instances into active and passive voice. This can be done, for example, by copying the target set to an Excel sheet and sorting them using Excel program functionality.

3. Teachers should avoid selecting instances where the subject of the node word is elliptic; examples are better if they explicitly include all main elements of a pattern.

4. Selecting examples that are in active voice is highly recommended when introducing the pattern for the first time. The teacher then can branch out to use other examples employing different structures.

5. There are some cases where the PDEV entry of a pattern of a verb has few instances and, in addition, most of these instances are too difficult for learners. In this case, the teachers are advised to use the full BNC corpus to look for more instances that are easier for the learners to understand. Alternatively, the teachers can use another corpus which is current and contains non-difficult language.

10.3.3 The description entry for patterns

The PDEV entries for all completed verbs are consistent. The entry of every pattern includes 4 elements: pattern, implicature example, and proportion for each pattern.

Figure 10.4, below, is an example of such a PDEV entry.

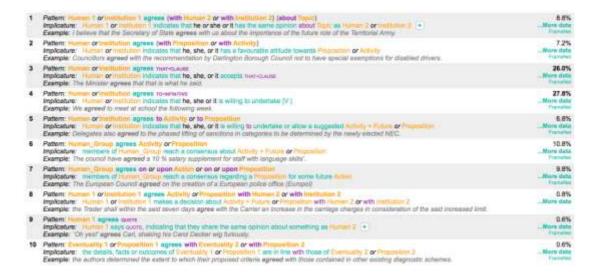**Figure 10.4 PDEV entry for AGREE (patterns 1 to 10)**

Figure 10.4 is a screenshot of the PDEV entry for AGREE. This entry includes 10 prototypical patterns of AGREE. The description of each pattern includes four elements: pattern, implicature, example and proportion of the pattern. This combination of elements is argued to be useful to learners for several reasons, which will be explained along with the discussion of these four elements below.

10.3.3.1 Pattern

The pattern of verbs in the PDEV entry generally consists of several elements, such as the node word (the verb itself), or collocational information, which includes collocation with classes of nouns, colligational information, semantic information, and grammatical information. This information can be seen throughout the ten pattern descriptions for AGREE in Figure 10.4, for example. However, there is no consistency within the pattern. For instance, some of the labels are semantic (e.g. 'Eventuality') and some are structural (e.g. 'to-infinitive'); in other words, there is a mixture of types of information within the pattern, including semantic and grammatical information. Teaching patterns of verbs that include all

256

pertinent types of information (e.g. collocation, colligation, semantic types) is more useful for learners than teaching words in isolation.

In addition, information about types of patterns is added at the beginning of each pattern, whether the pattern is an idiom, phrasal verb, slang, or what, as illustrated in Figure 10.5. Such extra information would be useful for the teachers as well as the learners; for instance, if the teacher wishes to teach their students some phrasal verbs or idioms of the target verb, such extra information makes it easy for the teachers to choose material from the relevant entry. Similarly, some learners may be less knowledgeable about which types of phrases are phrasal verbs, idioms, slang, etc. If, for example, they are looking at the PDEV entry for TELL, they will be aware that patterns 19, 20, and 21 are special types of patterns that they can use if they wish to sound more native-like. This is because the PDEV explicitly presents this information.

Figure 10.5 PDEV entry for TELL (patterns 19, 20, and 21)

Another aspect of PDEV entries which is useful relates to format, specifically to their use of colours. Figure 10.4 shows the different colours used for different functions. Semantic types are presented in yellow (e.g. Human, Institution, Activity, Proposition); grammatical words are in purple (e.g. *with, about*, *that*-clause, *to*-infinitive); and the node word (the verb) is in green (*agree*).

This way of using colours is useful to help the learners distinguish between patterns. For example, in Figure 10.4, *with* occurs in several patterns of AGREE: patterns 1, 2, 8, and 10. In some of these patterns (patterns 1, 2, and 10), *with* is followed by semantic types, whereas in pattern 8 it is preceded by semantic types. This kind of coding helps the learners spot the differences between patterns more quickly. Similarly, two semantic types, Activity and Proposition, appear in two different patterns: patterns 5 and 6 (see Figure 10.4). In pattern 5, these two semantic types are preceded by preposition *to* (e.g. *He agreed to requests*), while in pattern 6 the same semantic types occur without preposition *to* (e.g. *They agreed terms*).

10.3.3.2 Implicature

Another valuable feature of the PDEV entries is the implicatures, which function as definitions of a pattern. To illustrate this function, the example of pattern 5 of AGREE is shown in Figure 10.6 below:

Figure 10.6 PDEV entry for AGREE (pattern 5)

5 Pattern: Human or Institution agrees to Activity or to Proposition
   Implicature: Human or Institution indicates that he, she, or it is willing to undertake or allow a suggested Activity = Future or Proposition
   Example: Delegates also agreed to the phased lifting of sanctions in categories to be determined by the newly elected NEC.

Figure 10.6 shows a screenshot of the PDEV entry for pattern 5 of AGREE, in which AGREE is followed by preposition *to* followed by Activity or Proposition and the semantic type that populates the subject slot is Human or Institution (e.g. *He agreed to terms*). The implicature of pattern 5 is: *Human or Institution indicates that he, she, or it is willing to undertake or allow a suggested Activity = Future or Proposition*. This means that if someone or some institution agrees to something, s(he) or it will be willing to do it. This implicature is applied to any instance that instantiates pattern 5.

258

Two key features of the PDEV implicature hold. First, key elements presented in the

description of a pattern are reflected in the implicature. To illustrate this, the following show a

pattern and its implicature:

Figure 10.7 PDEV entry for AGREE (pattern 5)

A comparison between pattern 5 of AGREE shown above and its implicature shows that some

key elements that occur in the pattern, such as Human, Institution, Activity, Proposition and

to-preposition are found in the implicature in the same order (e.g. Human or Institution is

followed by AGREE followed by *to* followed by Activity or Proposition). Reflecting the

same order of pattern elements in the implicature is important for learners, as it allows the

learners to 'link' the usage of the pattern with its implicature (definition). This linkage

facilitates easier learning of form–meaning pairs.

The implicatures in the PDEV have advantages and disadvantages. The first advantage, as

mentioned earlier, is that the main information that occurs in the description of the pattern is

reflected in the implicature, which clarifies the pattern. Second, the implicature includes more

information that what the pattern itself shows; in other words, the implicature uses more

statements to define the meaning of the pattern: the node word is not just repeated in the

implicature, but rather is expanded as a statement. This helps learners understand the pattern

better. Third, the implicature provides information about what genres some patterns are likely

to be found in. For example, Figure 10.8 shows some patterns of *call*. Pattern 15, according to

PDEV, is most likely to be found in law discourse; pattern 22 is most likely to be found in

military discourse; and pattern 18 is more likely to be used in formal language. Such extra information is useful for learners.

In addition, PDEV implicatures of patterns help learners to get more information on the pattern. Almost no implicatures define semantic information. In Figure 10.8, for example, there are implicatures that further define semantic types, such as patterns 15 and 17. For instance, the implicature of pattern 15 further defines the semantic types Institution as Participant in Law Case, and Human 2 as Witness; similarly, the implicature of pattern 17 defines the semantic type Money further as Loan. Since semantic information is abstract, such elaboration on semantic types would be more useful for learners, as it would help learners understand these abstract words further. Sometimes semantic types are too abstract to define by implicature, such as Concept, Proposition, State of Affairs, etc. These terms are too difficult for learners to understand. It would be best if the implicature included some lexical items that belong to these terms so that the learners can comprehend, at least, what these semantic types mean.

Figure 10.8 PDEV entry for CALL

10.3.3.3 Examples

The inclusion of examples in the PDEV entry is important to the learners. As mentioned earlier in this chapter, some words that occur in the description of the pattern, particularly semantic types, may be too difficult for the learners to understand (e.g. Proposition, Entity, Concept). Even though the implicature of a pattern defines the meaning of the pattern, the usage of a given pattern may still be beyond learners' competence. Examples, in such a case, can play a most helpful role. For instance, the following is an amended PDEV entry for AGREE pattern 4, with the example deleted:

Pattern 4: Human or Institution agrees *to*-infinitive

Implicature: Human or Institution indicates that he, she, or it is willing to undertake V

In this amended entry, some coding conventions may be difficult for the learners, such as *to*-infinitive or V, highlighted above. To make these conventions clear to learners, an example for this pattern can be included. The amended entry is shown below with an example added:

Pattern 4: Human or Institution *agrees to*-infinitive

Implicature: Human or Institution indicates that he, she, or it is willing to undertake V

Example: She agrees to have a dinner.

This example serves as a concrete illustration of *agree* pattern 4. This example contains some abstract codes the PDEV used, including, but not limited to, *to*-infinitive and V. When the learners, for example, read the example they would realise that conventions such as *to*-infinitive and V mean *to* followed by verb (e.g. *to have*). They would then be able to observe this usage in a meaningful way.

However, throughout this study, it has been noticed that examples included in the PDEV

entry for patterns have some limitations that could impede learners' understanding.

10.3.3.4 The length of example

Some patterns' entries include examples that are too long. For instance, the following figure

presents a screenshot of a PDEV entry that includes an example that is too long:

Figure 10.9 PDEV entry for SEE (pattern 12)

This figure represents an entry of PDEV for SEE pattern 12. As seen, it includes a long

example:

> 10.8 *At Barclays de Zoete Wedd, whose report on the sector is due out this week,*
>
> *analyst Jane Anscombe points out that there are currently some worries on the*
>
> *advertising revenue front, and that some television people have **seen** a bit of*
>
> *pressure*. (BNC)

This example is intended to show the usage of pattern 12 of *see*: Human or Institution sees

Emotion or Attitude or Responsibility or Obligation. There are at least two possible

explanations why such a long example has been selected. One is that particular words only

occur in very long sentences; and, the team who have been establishing the PDEV are

following the *Cobuild* principle in that they include the whole sentence and keep the example in its original form.

In this particular example, it is actually pointless to give the whole example, and it would be better to cut it down. On the other hand, the part the pattern occurs in in the example, *some television people have **seen** a bit of pressure*, is quite difficult to understand in itself. This is because the meaning of the example depends on the aggregate whole of the example. For example, the sentence *there are currently some worries on the advertising revenue front* helps the reader to know that this sentence is talking about business or finance, because it is to do with the advertising, and thus to know that *television people* too is to do with the business of television. The context can also be manipulated (e.g. *There are currently some worries about income from advertising, and some television people have seen a bit of pressure*).

Another criticism relating to this particular example is related to culture. If Saudi learners, for instance, read this example, they may find it difficult to understand, because they may lack cultural knowledge of how television in some countries is financed.

There are at least two possible ways to overcome these problems. First, it would be better if this particular example were replaced with another example that was easy to understand for any learner. Second, if this particular example could not be replaced for any reason, it would be useful for the learners if the whole phrase of the pattern were underlined.

Inclusion of such a long example in the PDEV entry would be pointless for learners for the reasons mentioned above. Too long an example would divert learners from the target goal: observing patterns of a given verb. Looking at the example in Figure 10.9, it shows that the

target pattern comes at the end of the example (see example above) (e.g. *some television people have **seen** a bit of pressure*).

10.3.3.5 The structure of examples

Most pattern entries give examples in the active voice, but in some cases the examples are passive. This creates difficulty when the pattern is basically expressed in the active. Figure 10.10 illustrates this issue, as follows.

Figure 10.10 PDEV entry for ENCOURAGE (pattern 1)

This figure presents the PDEV entry for ENCOURAGE pattern 1. The description of pattern 1 is shown in active voice: *Human 1 or Institution 1 or Eventuality encourages Human 2 or Institution 2* to-*infinitive*. However, the example is shown in passive voice: *People are encouraged to question why things must be done*. Again, the lexical items that are supposed to fill the subject slot of ENCOURAGE are implicit. This results in the learner's question which of the semantic types that are shown in the description of pattern 1 of ENCOURAGE (e.g. Human 1, Institution 1, or Eventuality) is supposed to fill the subject slot in this example.

Another issue relating to examples is that some examples are given in an imperative mood. For example, in the following entry the pattern is presented in active while its example is expressed in the imperative structure:

Figure 10.11 PDEV entry for ALLOW (pattern 6)

> **6** *Pattern:* Human allows Resource (for Activity)
> *Implicature:* When planning, Human ensures that a sufficient amount of Resource is made available to complete Activity
> *Example: allow plenty of time in your total schedule for the physical typing and reproduction of your report.*

As seen, this figure includes a description of pattern 6 of ALLOW: the pattern is described in active voice—*Human <u>allows</u> Resource*. If we look at the example, however, we will notice that this structure is not actually reflected; rather, the example is shown in the imperative: *allow plenty of time in your total schedule*…. In this example, the lexical item that fills the subject slot of ALLOW is thus implicit. This would prevent the learner from understanding how the description of a given pattern is reflected in the example.

The reason the passive and imperative examples are argued to be difficult is that, unlike the pattern description itself, subjects are eliminated. To solve the above-mentioned problems of examples, in which a particular pattern is often passive, it would be better to give an active and passive example of each pattern so that each example can be matched to the pattern.

Another instance of a difficult example that may get learners confused is the order of lexical items that populate subject and object slots. For example, Figure 10.12 shows pattern 8 of ADOPT. The pattern is expressed as *Animate adopts Property*. The information that the pattern includes is shown in the typical order, that is, Subject Verb Object. On the other hand, the same information is expressed in different order in the example: *An alternative is a special posture that the prey adopts when approached by the predator, suddenly showing off its warning colours*. Both subject *posture* and object *prey* of the verb *adopt* come before the node word. This discrepancy may make learners confused about which word fills which role. In addition, the learners may not be familiar with the semantic types Animate and Property. In this case, it would be better to show which part of the example represents which part of the

pattern. A possible suggestion is to use different colours to map the main bits of the example onto those of the pattern. For example, both *Animate* and *prey* can be marked in one colour, while *Property* and *posture* can be put in another colour.

Figure 10.12 PDEV entry for ADOPT (pattern 8)

Another minor phenomenon observed in the PDEV is that some entries do not include examples. The following is one such case:

Figure 10.13 PDEV entry for ACCEPT (pattern 9)

Figure 10.13 shows the PDEV entry for pattern 9 of ACCEPT. As seen, no example is shown in the entry. The absence of an example will make it difficult for the learner to learn how a given pattern is used, especially when learners themselves are using the online PDEV. Thus, this entry would be impractical if the teacher wants to present it to the students as it is (without concocting an example to go with it).

10.3.3.6 One example for several semantic types of a given pattern

Some patterns in the PDEV entry occur with several semantic types, as shown in Figure 10.14 below:

Figure 10.14 PDEV entry for ENCOURAGE (pattern 1)

1 Pattern: Human 1 or Institution 1 or Eventuality encourages Human 2 or Institution 2 TO-INFINITIVE
Implicature: Human 1 or Institution 1 or Eventuality has the effect of causing Human 2 or Institution 2 to want to do something
Example: At school and university people are encouraged to question why things must be done, rather than accept orders passively.

Figure 10.14 shows the PDEV entry for ENCOURAGE pattern 1. Three semantic types can populate the subject slot of pattern 3 of ENCOURAGE: Human, Institution, or Eventuality 1; similarly, two semantic types can populate the object slot: Human 2 or Institution 2. Only one example is included in the entry (see Figure 10.14). Inclusion of one example, while the pattern occurs with several semantic types that fill the subject/object slots of a given verb would make learners ask questions such as 'Which of these three semantic types is reflected in this example? There are 3 semantic types populating the subject slot of ENCOURAGE but there is only one in the example; how about the others?' Such questions could be answered if the PDEV entry for every pattern were to include a number of example each representing the usage for each semantic type.

The following is a suggested version of the PDEV entry for ENCOURAGE pattern 3 (Figure 10.14):

Pattern: Human or Institution or Eventuality encourages Eventuality 2

Implicature: Human or Institution or Eventuality has the effect of causing Eventuality 2 to be more likely

Example: John encourages recycling

The government encourages the use of public transports

Attractive advertising encourages excessive consumption

The above amended entry includes 3 examples, each of which represents the usage of one of the semantic types that can populate the subject slot of ENCOURAGE with pattern 3, as underlined in the figure. By doing so, it makes it easier for learners to see how each of these semantic types is used.

10.3.3.7 Frequency of patterns in the PDEV entry

One key feature of the PDEV is the inclusion of frequency information. Each pattern recorded in the PDEV entry contains frequency information, shown in a percentage at the far right of the row for a pattern, as shown below:

Figure 10.15 PDEV entry for ENCOURAGE

As seen, this figure includes the frequency of each of the 3 patterns of ENCOURAGE as percentage rates: pattern 1 (46.4%), pattern 2 (5.6%), and pattern 3 (47.6%). Inclusion of frequency (percentage) for every pattern would not be useful for learners, however, as what the learners need to know is not the frequency of the pattern, but rather the circumstances in which a pattern is used, in order to produce it appropriately when (for example) writing an essay. For some researchers, however, knowing the frequency or proportion of patterns is very useful.

10.3.3.8 Range of 'typical' patterns of a given verb

One key features that the PDEV has is that it identifies only the 'typical' patterns of English verbs (Hanks 2000). This feature is useful for learners; if the learner wants to focus on how a given verb is 'typically' used by native speakers, the PDEV would be the best reference to consult.

The PDEV entry for each verb includes a range of patterns of verb that the learner may wish to learn. So if the learner' knowledge is limited to, say, one usage of ALLOW (e.g. *Human allows Human to-infinitive = The teacher allowed the learners to use their mobile phones in the classroom*), (s)he can check the PDEV entry for other alternative way of saying the same thing. For example, among the 8 patterns recorded in the PDEV entry for ALLOW, there is one pattern that can serve as an alternative to the usual usage: *The teacher allowed the learners to use their mobile phones in the classroom*. The alternative pattern that the PDEV entry for *allow* offers is as follows:

   10.9 Pattern 4: *Human allows Eventuality*

      Example: *The teacher allows the use of mobile phones in the classroom*. (BNC)

This alternative shows how ALLOW can be followed by Eventuality, such as *the use of mobile phones*. The learner will know that there may be an alternative to his/her conventional usage of ALLOW (e.g. *Human allows Human to-infinitive*), and may find that pattern 4 of ALLOW, *Human allows Eventuality*, serves as a good alternative. Thus, the learner would apply pattern 4 to their sentence, as follows:

     Student's usual usage:

10.10 *The teacher allowed the learners to use mobile phones in the classroom.* (ICLE-

Swedish)

Alternative usage (PDEV):

10.11 *The teacher allowed <u>the use of mobile phones</u> in the classroom.*

The use of pattern 4 as an alternative would make students' language more sophisticated.

The reader should be warned here that not just any pattern in the PDEV works as an alternative. In addition, if a given pattern, say, pattern 4 of ALLOW, works as an alternative to a given sentence (e.g. *The teacher allowed <u>the learners to use mobile phones</u> in the classroom*; *The teacher allowed <u>the use of mobile phones</u> in the classroom*), this does not mean that it can also work for any analogous sentences (e.g. *The company allowed us to travel* vs. **The company allowed the travel*).

10.4 Advantages of applying the concepts in PDEV for teachers

The method that I have piloted and demonstrated in this thesis could be used by appropriately trained teachers, who could follow these steps:

1. Load students' texts into concordance software to generate concordance lines of a target verb in KWIC (key word in context) form.

2. Print the outcomes of the target verb KWIC.

3. Get the PDEV entry of the target verb ready.

4. Start matching students' correct usages of the target verb with those of the PDEV entry by annotating the concordance lines with the numbers of their corresponding patterns in the PDEV (pattern 1, pattern 2, etc.).

5. Mark those usages that appear to be divergent and put them aside for further action.

6. Search the BNC for these divergent usages to make sure that they do not actually exist.

7. Continue to iteratively complete the previous steps until one is done annotating all the concordance lines.

8. Group lines of correct usages that have the same annotated numbers together (e.g. lines for pattern 1, lines for pattern 3)

9. Count those matched patterns identified in the learners' texts and record them in a table

10. Group divergent usages according to their types of divergence (e.g. Type 1 = *Human agrees Human to-infinitive*; Type 2 = *Human agrees Human infinitive*).

11. Count instances of the divergent usages identified in the learners' texts and record them.

Having carried out these steps, the teacher should be able to come up with a good learner's version of PDEV. This kind of version can be effectively used by the teacher to a) identify divergent usages of a target verb among learners; b) find alternatives to these divergent usages, if applicable; and c) know what rate of 'typical' usages of a target verb the learners have achieved so that they can expand the learners' knowledge concerning the target verb. This can be done by introducing those patterns of the target verb that are recorded in the PDEV entry but not found in learners' texts.

If the teacher wants to know how much typical usage of a given verb has become a part of learners' English and what is yet to be learned and used by the learners, (s)he can use the PDEV entry as a measure of the range of usages of a given verb identified in the learners' writing. An example of how the PDEV entry can be used as a measurement can be found in Chapters 6 and 9, where the Swedish learners show a smaller range of usages of ALLOW relative to the range of patterns recoded in the PDEV entry of the same verb.

In Chapters 6 and 9, the result shows that of 8 typical patterns of ALLOW recorded in the PDEV entry, only three patterns were identified in the ICLE-Swedish sub-corpus, meaning that only three, a relatively small proportion, had become a part of the learners' competence, and the other patterns are yet to be used by the learners. The teacher may wish to expand his/her students' knowledge concerning the range of uses of ALLOW, and could, for instance, introduce the patterns of ALLOW that were not found in the students' writing (5 patterns).

At this point, it should be noted that teachers who wish to create a learners' version of PDEV should be familiar with PDEV conventions such as semantic types, ontology, frequency, or implicature, among others.

10.5 Review of the thesis

This thesis intended to see how useful the Pattern Dictionary of English Verbs (PDEV) could be to learners; to explore what the use of corpus pattern analysis of verbs (CPA) can tell us about learners' uses of English verbs; to see how PDEV can be translated into pedagogical implications; and to explore how CPA/PDEV is different/similar to other related concepts.

Chapter 2 introduced the notion of PDEV/CPA, and discussed how it is related to other language phenomena such as collocation, colligation and semantic preference. It showed that CPA/PDEV overlaps with all of these language phenomena in that the CPA/PDEV identifies such phenomena. In other words, when CPA is applied to texts to identify patterns of verb it does not look at patterns as constituents of one unit (e.g. collocation, colligation, semantic preference); rather, it identifies the verb pattern as an integrated combination of these units.

Chapter 3 introduced two concepts relevant to this thesis: error analysis (EA) and learner corpora. It discussed weaknesses of EA and how a new era of describing learners' language (CL) has emerged. In addition, some work done in learner corpous research was considered.

The fourth chapter introduced the methods for corpus selection and data extraction. The size and structure of the corpora were discussed, and issues with the target corpora comparability described. The reasons for the selection of the target corpora and the target verbs were given. The extraction and interpretation of the target verbs was explicated step by step. Chapter 4 set out to apply the CPA/PDEV to two corpora from the ICLE data (ICLE-Chinese and ICLE-Swedish sub-corpora) and to LOCNESS (The Louvain Corpus of Native English Essays). It looked at sixteen verb lemmas in each target corpus; this was done following the CPA method described in Chapter 2. The concordance lines for every target verb in each target corpus were annotated and grouped accordingly. Perhaps the most important finding here is that the learners produced correct patterns most of the time. Of these 16 verbs, 5 verb lemmas were selected for further qualitative study. The results show that when the learners produce divergent patterns they still produce patterns that are recognisable and shared by several learners. The PDEV entry served as a good resource to identify divergent usages and find alternatives to these usages.

Since Chapter 5 mostly focused on identifying divergent usages found in ICLE-Chinese, the chapter that followed, Chapter 6, was devoted to studying the ICLE-Swedish sub-corpus in relation to the phenomenon of over-/under-representation. The reason for selecting the Swedish sub-corpus for this study is that it encourages the use of the concepts of over-and under- representation rather than the concept of divergence. That is, the quantitative analysis of the 16 verbs in Chapter 4 showed that for the most part errors in ICLE-Swedish are simple typographical errors, suggesting that a study of divergence would be unproductive. The aim of Chapter 6 was to complement the study of divergence in Chapter 5 with a study of non-divergent differences between ICLE-Swedish and LOCNESS using the PDEV as a benchmark for the range of patterns of target verb identified in the two corpora, that is, *allow*. The results show that the Swedish learners did not use the full range of patterns of ALLOW *recorded in the PDEV entry*. It was argued that this phenomenon had two main possible explanations: a) the Swedish learners did not learn them; or b) they did not have the opportunity to use them.

Another important finding was that the Swedish learners preferred some patterns of ALLOW to others. Patterns where *allow* is followed by a *to*-infinitive (e.g. *The teacher allows <u>students to use mobile phones</u> in the classroom*) was found to be preferred more than when ALLOW lacks a *to*-infinitive (e.g. *The teacher allows <u>the use of mobile phones</u> in the classroom*). It was argued that some of the Swedish learners may lack sufficient knowledge of nominalization as there were only a small number of cases where nominalization was used in the Swedish sample. Finally, the results show that most Swedish learners avoided the use of non-congruent uses of the verb. For instance, the Swedish learners preferred *The <u>director</u> of school allows the use of mobile phones in the classroom* to *The <u>rules and regulations</u> allow the use of mobile phones in the classroom*.

Unlike the two preceding chapters, Chapters 5 and 6, Chapter 7 studied a verb that was not completed in the PDEV, that is, SUGGEST. The aim of this chapter was to discover how feasible it is or is not for an individual not trained within the PDEV project to compile a PDEV entry, and to learn more about native-speaker writers' and learners' uses of *suggest*. In addition, this chapter was meant to offer some recommendations to teachers and researchers who wish to use CPA/PDEV. Three corpora were used to identify uses of SUGGEST: ICLE-Chinese, LOCNESS and the BNC. Based on 300 concordance lines taken from the BNC, four patterns of SUGGEST were identified. Uses of SUGGEST were identified in the learners' corpora, following a built PDEV entry for SUGGEST. The results show that the Chinese learners produced correct patterns of SUGGEST most of the time, and when they produced divergent patterns they still produced patterns that were recognisable and are used by several learners. A similar phenomenon found in Chapter 6, where some Swedish learners avoided the use of non-congruent ALLOW and nominalization, was shown by the Chinese learners with the use of SUGGEST. The chapter concluded with some recommendations to teachers and researchers about the use of CPA/PDEV, and suggested some pedagogical implications.

In the discussion chapter, Chapter 8, the major implications of the research were re- addressed as a whole. Some ideas concerning various possible applications in pedagogy were put forward, such as the use of PDEV as DDL-supported classroom activities for the Chinese and Swedish learners. The underlying research question this research intends to answer (i.e. how useful is PDEV to learners?) was discussed in detail in this chapter.

Chapter 9 discussed the pedagogical implications of this thesis and answered research question 3. Various ways of presenting the outcomes as errors and clarify how to get to correct alternatives were shown.

10.6 Limitations of PDEV/CPA

Although CPA/PDEV is a useful tool for examining learners' language and can be used effectively by both teachers and learners, as demonstrated in the previous chapters, it has some limitations that need to be acknowledged.

First, CPA is labour-intensive. For instance, analysing one verb following the CPA approach requires performing a number of steps, as explained in Chapter 7. Besides, CPA is done manually, which needs considerable resources and the amount of ground that can be covered is limited.

This limitation also relates to the pedagogical implications detailed in Chapter 9. Though such implications would prove useful for students, the efforts to achieve them would be time consuming for teachers. To address the divergent patterns produced by students, the teacher could prepare soft exercises with which most teachers are familiar. These include fill-in-the-blanks with the correct verb or multiple-choice exercises, in which students choose the verb that makes a pattern correct.

Second, the CPA approach is difficult to generalise. Unlike the Pattern Grammar approach, where all verbs that have similar meanings when used in a particular pattern are grouped together, for instance, the CPA approach goes verb by verb.

Finally, the PDEV is incomplete. At the time of writing this thesis, among the 5,392 verb entries that have been set to be completed by the PDEV team, 3,556 verb entries have yet to be worked on, 240 verb entries are still under progress, and 1,416 verb entries are complete. Since this project is incomplete, changes to the verb entries and ontology may take place;

information may be added, changed or even deleted. For example, there are still more semantic types to be added to the current PDEV ontology.

10.7 Some limitations of the research

Though it has been demonstrated in the previous chapters that CPA/PDEV is a useful way of looking at learner language, and can be usefully used by both teachers and learners, the research is inevitably limited in scope. It looks at some verbs in detail, but there are certainly some limitations of this thesis that need to be acknowledged.

First, although the CPA/PDEV is sufficient for interpreting the usages of verbs, this thesis only covers a sub-set of verbs in English. It would be worthwhile to extend the investigation to as many other verbs as possible, because they may provide more richness of data. In addition, due to time limitations this study focused on only two learner corpora, ICLE-Chinese and ICLE-Swedish, as well as LOCNESS. It would be worthwhile to extend the investigation to as many learner corpora as possible, because they may reveal new phenomena regarding learner-specific use of verbs.

It would have been beneficial for me if I could have included a written corpus of data from Saudi learners of English, since I come from the same background; however, no such corpus exits, and time was not sufficient to travel back to Saudi Arabia for data collection.

10.8 Directions for future research

This study has focused on identifying uses of verbs in written corpora. As reported in Chapter 5, the investigation of the 16 target verbs in the learner corpora and LOCNESS reveals that the learners produce correct patterns most of the time. In other words, the proportion of non-

divergent patterns in each of the target corpora outnumbers that of divergent patterns. More varieties of divergent usages of verbs may be identified in spoken corpora; thus, extending the investigation to include spoken corpora is worth doing since several researchers identify different linguistic features in spoken and written academic registers (cf. Biber et al. 1999). Suggested spoken learner corpora include a spoken counterpart to ICLE: the Louvain International Database of Spoken English Interlanguage (LINDSEI), comparable to the native-speaker the LOCNEC (Louvain Corpus of Native English Conversation. These two corpora, as described by Sylviane Granger, include transcribed interviews produced by advanced learners of English from various mother-tongue backgrounds.

Moreover, the use of the CPA/PDEV to identify patterns of verbs at different proficiency levels is worth doing, since several researchers reported that phraseology can distinguish between intermediate and advanced learners (cf. Granger and Bestgen 2014).

Finally, the use of the CPA/PDEV to identify patterns of verb in different genres is highly encouraged. A study done by Jane Bradbury (2016) uses the CPA/PDEV to compare adults' and children's use of *blow* in two different genres: a general corpus of non-fiction written by adults (BNC) and creative writing by children aged 13 and under (BBC2013) (ibid.). Lexical items that populate subject slot of one pattern of *blow* were found in the BBC2013 sample but not found in the BNC. These items are *whistle* in *whistle blew*; *dragon* in *dragon blew fire*. Such a difference is argued to be related to genre difference; extending the investigation to different genres or different specialized corpora may provide new related phenomena.

Rees (2018) examined differences in meaning and use of vocabulary across academic disciplines employing the CPA as a method. One key findings in this regard is that the

semantic types that populate the subject slot of *assemble* in the Management sub-corpus are different from in Microbiology, where Physical Objects and Microbiological Processes *assemble* Objects, while in Management People or Organizations *assemble* Objects and Groups.

10.9 Final remarks

Having spent considerable time on studying PDEV and applying the concepts in it on some sub-corpora of the learner corpora, I have come to appreciate the wealth of information that is provided in PDEV. In addition, doing this research has increased my awareness of the semantic restrictions on patterns of English verbs.

It has inspired me to work further on this topic in the future, and my next step will be to encourage teachers to use PDEV-produced materials that will help them to employ PDEV. Finally, if copyright issues can be resolved, I would like to produce a version of PDEV for use in the Arab-speaking world that would focus on those areas that Arabic learners are likely to find the most difficult.

References

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. 132-141). London/New York: Longman.

Abdulmajeed, H. M. (2017). An integrated approach to achievement: measuring the development of writing skills in Kurdish learners of English as a foreign language (EFL). PhD thesis. University of Birmingham.

Aijmer, K. (2002). Modality in advanced Swedish learners' written interlanguage. In S. Granger, J. Hung, & S. Petch-Tyson (Eds.), Computer Learner Corpora, Second Language Acquisition and Foreign Language Teaching (pp. 55-76). Amsterdam/Philadelphia: John Benjamins Publishing Company.

Altenberg, B. (1998). On the phraseology of spoken English: the evidence of recurrent word-combinations. In A. P. Cowie, (Ed.), Phraseology: Theory, Analysis, and Applications (pp. 101-122). Oxford: Clarendon 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.

Andersen, R. W. (1977). The impoverished state of cross-sectional morpheme acquisition/accuracy methodology (or: The leftovers are more nourishing than the main course.) Working Papers on Bilingualism, No. 14.

Baisa, V., El Maarouf, I., Rychlý, P., & Rambousek, A. (2015). Software and data for corpus pattern analysis. In Horák, A., Rychlý, P., and Rambousek, A. (Eds.), Ninth Workshop on Recent Advances in Slavonic Natural Language Processing. Brno. Tribun EU. 75–86.

Baisa, V., Može, S., & Renau, I. (2016). Multilingual CPA: Linking Verb Patterns across Languages. Retrieved from http://euralex.org/wp-content/themes/euralex/proceedings/Euralex%202016/euralex_2016_044_p410.pdf.

Benson, M., Benson, E., & Ilson, R. (2009). The BBI Combinatory Dictionary of English: Your Guide to Collocations and Grammar. Amsterdam/Philadelphia: John Benjamins.

Benson, M., Benson, E., & Ilson, R. F. (2010). The BBI Combinatory Dictionary of English: Your Guide to Collocations and Grammar. Revised by R. Ilson. Amsterdam/Philadelphia John Benjamins.

Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (1999). Longman Grammar of Spoken and Written English. London: Longman.

Boulton, A. (2012). Hands-on/hands-off: varying approaches in data-driven learning. Paper presented at the 9th Teaching and Language Corpora Conference. Brno, Czech Republic: Masaryk University, 30 June–3 July.

Bradbury, J. (2006). The use of verb patterns by children and adults: a comparative corpus study. Master's thesis. University of Birmingham.

Castro, M. B., & Faber, P. (2014). Collocation dictionaries: a comparative analysis. MonTi: Monografías de Traducción e Interpretación, 6, 203-235.

Charles, M. (2011). Using hands-on concordancing to teach rhetorical functions: evaluation and implications for EAP writing. In: Frankenburg-Garcia, A., Flowerdew, L. and Aston, G. (Eds.), New trends in corpora and language learning (pp. 26-43). London: Continuum.

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), 418-442.

Chujo, K., Oghigian, K., & Akasegawa, S. (2015). A corpus and grammatical browsing system for remedial EFL learners. In A. Leńko-Szymańska & A. Boulton (Eds.), Multiple Affordances of Language Corpora for Data-driven Learning (pp. 109–128). Amsterdam: John Benjamins.

Cobb, T. (2003). Analyzing late interlanguage with learner corpora: Quebec replications of three

European studies. Canadian Modern Language Review, 59(3), 393-424.

Coffey, S. (2010). A new pedagogical dictionary of English collocations. International Journal of Lexicography, 24(3), 328-341.

Corder, S. P. (1967). The significance of learner's errors. IRAL: International Review of Applied Linguistics in Language Teaching, 5(1-4), 161-170.

Cowie, A. P. (1998). Phraseology: Theory, Analysis, and Applications. Oxford: Clarendon Press.

De Cock, S. (2004). Preferred sequences of words in NS and NNS speech. Belgian Journal of English Language and Literatures (BELL), New Series, 2(2), 225-246.

De Cock, S., & Granger, S. (2005). Computer learner corpora and monolingual learners' dictionaries: the perfect match. Lexicographica, 20, 72-86.

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/New York: Longman.

Deshors, S. C., Götz, S., & Laporte, S. (2016). Linguistic innovations in EFL and ESL: rethinking the linguistic creativity of non-native English speakers. International Journal of Learner Corpus Research, 2(2), 131-150.

Edwards, A., & Lange, R.-J. (2016). In case of innovation. International Journal of Learner Corpus Research, 2(2), 252-277.

Edwards, A., & Laporte, S. (2015). Outer and Expanding Circle Englishes: the competing roles of norm orientation and proficiency levels. English World-Wide, 36(2), 135-169.

El Maarouf, I. (2013). Methodological Aspects of Corpus Pattern Analysis. ICAME Journal, 37, 119–148.

El Maarouf, I., Bradbury, J., & Hanks, P. (2014a). PDEV-lemon: a Linked Data implementation of the Pattern Dictionary of English Verbs based on the Lemon model.

Retrieved from https://www.researchgate.net/publication/275714820_PDEV-lemon_a_Linked_Data_implementation_of_the_Pattern_Dictionary_of_English_Verbs_based_on_the_Lemon_model.

El Maarouf, I., Bradbury, J., Baisa, V., & Hanks, P. (2014b). Disambiguating verbs by collocation: corpus lexicography meets natural language processing (pp. 1001-1006). In *LREC*.

Ellis, R. (1994). The study of second language acquisition. Oxford: Oxford University Press.

Ellis, R., & Barkhuizen, G. P. (2005). Analysing Learner Language: Oxford: Oxford University Press.

Ellis, N. C., Römer, U., & O'Donnell, M. B. (2016). Usage-Based Approaches to Language Acquisition and Processing: Cognitive and Corpus Investigations of Construction Grammar. Language Learning Monograph Series. West Sussex: Wiley-Blackwell.

Firth, J. R. (1951). General linguistics and descriptive grammar. TPhS 69–87. (Reprinted in Papers in Linguistics 1934–1951. London: Oxford U.P., 1957. 216–228.)

Fillmore, C. J., Johnson, C. R. and Petruck, M. R. L. (2003). Background to FrameNet. International Journal of Lexicography 16: 235–251.

Francis, G. (1993). A corpus-driven approach to grammar: Principles, methods and examples. In Baker, M., Francis, G., and Tognini-Bonelli, E. (Eds.), Text and technology: in honour of John Sinclair (pp.137-156). Benjamins, The Netherlands.

Francis, G. (1995). Corpus-driven grammar and its relevance to the learning of English in a cross-cultural situation. English in Education: Multicultural Perspectives. Singapore: Unipress.

Francis, G. (2015). Hugh's across all that': Some changing uses of prepositions. In N. Groom, M. Charles and S. John (eds.) Corpora, Grammar and Discourse. Amsterdam: Benjamins. 73-96.

Francis G., Hunston S. & Manning E. (1996). Collins COBUILD Grammar Patterns 1: Verbs.

London: HarperCollins.

Francis G., Hunston S. & Manning E. (1998). Collins COBUILD Grammar patterns 2: Nouns and adjectives. London: Harper Collins.

Frankenberg-Garcia, A. (2014). The use of corpus examples for language comprehension and production. ReCALL, 26(2), 128-146.

Gilquin, G. (2010). Corpus, cognition and causative constructions. Amsterdam: John Benjamins.

Gilquin, G. (2011). Corpus linguistics to bridge the gap between World Englishes and Learner Englishes. Communicación en el Siglo, 21(2), 638-642.

Gilquin, G. (2012). Lexical infelicity in English causative constructions. Comparing native and learner collostructions. Retrieved from https://www.researchgate.net/publication/267426912_Lexical_infelicity_in_causative _constructions_Comparing_native_and_learner_collostructions.

Gilquin, G. (2015). Contrastive collostructional analysis: causative constructions in English and French Z. Angl. Am., 63 (2015), pp. 253-272.

Gilquin, G., & Granger, S. (2011). From EFL to ESL: evidence from the International Corpus of Learner English. In Hundt, M.& Mukherjee, D. (Eds.), Exploring second-language varieties of English and learner Englishes: Bridging a paradigm gap (pp. 55-78). Amsterdam, the Netherlands: John Benjamins.

Gläser, R. (1998). The stylistic potential of phraseological units in the light of genre analysis. In A. P. Cowie (Ed.), Phraseology: Theory, Analysis, and Applications (pp. 125-143). New York: Oxford University Press.

Goldberg, A. E. (1995). Constructions: A Construction Grammar Approach to Argument Structure. Chicago: University of Chicago Press.

Goldberg, A. E. (2003). Constructions: a new theoretical approach to language. Trends in Cognitive Sciences, 7(5), 219-224.

Goldberg, A. E. (2006). Constructions at Work: The Nature of Generalization in Language. Oxford University Press.

González Fernández, B., & Schmitt, N. (2015). How much collocation knowledge do L2 learners have? ITL: International Journal of Applied Linguistics, 166(1), 94-126.

Götz, S., & Schilk, M. (2011). Formulaic sequences in spoken ENL, ESL and EFL: focus on British English, Indian English and learner English of advanced German learners. In J. Mukherjee & M. Hundt (Eds.), Exploring Second-Language Varieties of English and Learner Englishes: Bridging a Paradigm Gap (pp. 79-100). Amsterdam/Philadelphia: John Benjamins.

Granger, S. (1993). International Corpus of Learner English. In J. Aarts, P. de Haan, & N. Oostdijk (Eds.), English Language Corpora: Design, Analysis and Exploitation. Amsterdam: Rodopi, pp. 57-71.

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.

Granger, S. (Ed.). (1998a). Learner English on Computer. London/New York: Longman.

Granger, S. (1998b). Prefabricated patterns in advanced EFL writing: collocations and formulae. In A. P. Cowie (Ed.), Phraseology: Theory, Analysis, and Applications (pp. 145-160). New York: Oxford University Press.

Granger, S. 1998c. 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/New York: Longman.

Granger, S. (2002). A bird's-eye view of learners corpus research In Granger, S., Hung, J., & Petch-Tyson, S. (Eds.), Computer learner corpora, second language acquisition and foreign language teaching (pp. 5-33). Amsterdam: John Benjamins.

Granger, S. (2015). Contrastive interlanguage analysis: a reappraisal. International Journal of Learner

Corpus Research, 1(1), 7-24.

Granger, S., & Bestgen, Y. (2014). The use of collocations by intermediate vs. advanced non-native writers: A bigram-based study. International Review of Applied Linguistics in Language Teaching, 52(3), 229-252.

Granger, S., & Rayson, P. (1998). Automatic profiling of learner texts. In S. Granger (Ed.), Learner English on Computer (pp. 119-131). London/New York: Longman.

Granger, S., & Tribble, C. 1998. Learner corpus data in the foreign language classroom: form-focused instruction and data-driven learning. In S. Granger (Ed.), Learner English on Computer (pp. 199-209). London/New York: Longman.

Granger, S., Dagneaux, E., Meunier, F., & Paquot, M. (2009). International Corpus of Learner English: UCL, Presses Univ. de Louvain.

Granger, S., Hung, J., & Petch-Tyson, S. (2002). Computer Learner Corpora, Second Language Acquisition, and Foreign Language teaching (Vol. 6). Amsterdam/Philadelphia:: John Benjamins Publishing.

Granger, S., Gilquin, G., & Meunier, F. (Eds) (2015). The Cambridge handbook of learner corpus research. Cambridge: Cambridge University Press.

Grice, H. P. (1968). Utterer's meaning, sentence meaning, and word-meaning. Foundations of Language 4 (3).

Grice, H. P. (1975). Logic and conversation. In Peter Cole and Jerry L. Morgan (eds.), Syntax and Semantics. Vol. 3, Speech Acts. Academic Press.

Guo, X. (2006). Verbs in the Written English of Chinese Learners: A Corpus-Based Comparison between Non-Native Speakers and Native Speakers. PhD thesis. The University of Birmingham.

Halliday, M. (1993). On the language of physical science. In M. Halliday & J. Martin (Eds.), Writing Science: Literacy and Discursive Power. London: Taylor and Francis.

Hanks, P. (2000). Do word meanings exist? Computers and the Humanities, 34(1-2), 205-215.

Hanks, P. (2004). Corpus pattern analysis. Retrieved from

http://www.euralex.org/elx_proceedings/Euralex2004/009_2004_V1_Patrick%20HANKS_C

orpus%20pattern%20analysis.pdf

Hanks, P. (2012). How people use words to make meanings: semantic types meet valencies.

In A. Boulton and J. Thomas (Eds.), Input, Process and Product: Developments in

Teaching and Language Corpora (pp. 54–69). Brno.

Hanks, P. (2013). Lexical Analysis: Norms and Exploitations. Massachusetts: The MIT Press.

Hanks, P., & Ježek, E. (2008). Shimmering lexical sets In E. Bernal & J. De Cesaris (Eds.),

Proceedings of the XIII EURALEX International Congress (pp. 391-402). Barcelona:

Universitet Pompeu Fabra.

Hanks, P., El Maarouf, I., & Oakes, M. (2018). Flexibility of multiword expressions and

Corpus Pattern Analysis. Multiword expressions, 93.

Hasselgren, A. (1994). Lexical teddy bears and advanced learners: a study into the ways Norwegian

students cope with English vocabulary. International Journal of Applied Linguistics, 4(2),

237-258.

Herbst, T., Heath, D., & Roe, I. F. (2004). A valency dictionary of english : A corpus-based

analysis of the complementation patterns of english verbs, nouns and adjectives.

Retrieved from https://ebookcentral.proquest.com

Hill, J., & Lewis, M. (1997). LTP dictionary of selected collocations. Hove, UK: Language Teaching.

Hoey, M. (2003). Lexical priming and the qualities of text. Retrieved July 27, 2017

http://www.monabaker.com/tsresources/LexicalPrimingandthePropertiesofText.htm.

Hoey, M. (2004). Textual colligation: a special kind of lexical priming. Language and Computers,

49, 171-194.

Hoey, M. (2005). Lexical Priming: A New Theory of Words and Language.
London: Routledge.

Hoey, M. (2009). Corpus-driven approaches to grammar. In U. Römer & R. Schulze (Eds.).
Exploring the Lexis–Grammar Interface. Amsterdam/Philadelphia: John Benjamins.

Hoey, M., Mahlberg, M., Stubbs, M. and Teubert, W. (2007). Text, Discourse and Corpora: Theory
and Analysis. London: Continuum.

Hornby, A. S. (1954). A Guide to Patterns and Usage in English. Oxford: Oxford University
Press.

Hunston, S. (2002). Pattern grammar, language teaching, and linguistic variation: applications
of a corpus-driven grammar' in Reppen, Biber and Fitzmaurice (eds.) Using Corpora
to Explore Linguistic Variation. Benjamins. 167-186.

Hunston, S. (2004). The corpus, grammar patterns, and lexicography. Lexicographica 20: 99-
112.

Hunston, S. (2014). Pattern Grammar in Context. In T. Herbst et al (eds.) Constructions
Collocations Patterns Berlin/Boston: de Gruyter Mouton. 91-111.

Hunston, S. forthcoming. Patterns and constructions: finding an alignment between models.
International Journal of Corpus Linguistics.

Hunston, S. and Su, H. (2018). Pattern, construction and local grammar: the case of
evaluation. Applied Linguistics accepted, in press.

Hunston, S. (2002). Corpora in Applied Linguistics. Cambridge: Cambridge
University Press.

Hunston, S., & Francis, G. (2000). Pattern grammar: a corpus-driven approach to the lexical grammar
of English. Amsterdam: Benjamins.

Hunston, S., & Francis, G. (1998). Verbs observed: A corpus-driven pedagogic grammar. Applied

Linguistics, 19(1), 45-72.

Johns, T. (1986). Micro-concord: a language learner's research tool. System, 14(2), 151-162.

Johns, T. (1988). Whence and whither classroom concordancing. In T. Bongaerts, P. de

Haan,S. Lobbe, and H. Wekker (Eds.), Computer Applications in Language Learning

(pp. 9-27). Dordrecht, The Netherlands: Foris.

Johns, T. (1991). Should you be persuaded: two samples of data-driven learning materials.

Retrieved from

https://s3.amazonaws.com/academia.edu.documents/31165100/Tim_Johns_and_DDL.

pdf?AWSAccessKeyId=AKIAIWOWYYGZ2Y53UL3A&Expires=1541595446&Sig

nature=wFrtdaM1aVEx%2Fs%2BnePnWftlU1b8%3D&response-content-

disposition=inline%3B%20filename%3DShould_you_be_persuaded_Two_samples_o

f_d.pdf

Johns, T. (1994). From printout to handout: grammar and vocabulary teaching in the context

of data-driven learning. In T. Odlin (Ed.), Perspectives on Pedagogical Grammar (pp.

293-313). Cambridge: Cambridge University Press.

Johns, T. (2002). Data-driven learning: the perpetual challenge. In B. Kettemann & G. Marko

(Eds.) Teaching and Learning by Doing Corpus Analysis. Proceedings of the Fourth

International Conference on Teaching and Language Corpora (pp. 107-117).

Amsterdam: Rodopi.

Kachru, B. B. (1985). Standards, codification and sociolinguistic realism: the English language in the

outer circle. In R. Quirk & H. G. Widdowson (Eds.), English in the World: Teaching and

Learning the Language and Literatures. Cambridge: Cambridge University Press.

Kaszubski, P. (1998). Learner corpora: The cross-roads of linguistic norm. In TALC98 Proceedings

(pp. 24-27). Oxford: Humanities Computing Unit, Oxford University.

Kennedy, G. D. (1998). An Introduction to Corpus Linguistics. London: Longman.

Kleinmann, H. H. (1977). Avoidance behavior in adult second language acquisition 1. Language Learning, 27(1), 93-107.

Laporte, S. (2012). Mind the gap! Bridge between World Englishes and Learner Englishes in the making. English Text Construction, 5(2), 264-291.

Lea, D., & Runcie, M. (2002). Blunt instruments and fine distinctions: a collocations dictionary for students of English. In Braasch A. & C. Povlsen (Eds.), Proceedings of the Tenth EURALEX International Congress (pp. 819-829). Copenhagen: Center for Sprogteknologi.

Leech, G. (2001). The role of frequency in ELT: new corpus evidence brings a re-appraisal. Foreign Language Teaching and Research, 33(5), 328-339.

Little, D. (2007). The Common European Framework of Reference for Languages: perspectives on the making of supranational language education policy. Retrieved January 10, 2018 from https://onlinelibrary.wiley.com/doi/full/10.1111/j.1540-4781.2007.00627_2.x.

Lorenz, G. (1998). Overstatement in advanced learners' writing: stylistic aspects of adjective intensification. In S. Granger (Ed.), Learner English on computer. London/New York: Longman.

McGee, I. (2012). Collocation dictionaries as inductive learning resources in data-driven learning— an analysis and evaluation. International Journal of Lexicography, 25(3), 319-361.

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

Neff, J., Dafouz, E., Herrera, H., Martínez, F., & Rica, J. P. (2003). Contrasting learner corpora: the use of modal and reporting verbs in the expression of writer stance. In S. Granger & S. Petch-Tyson (Eds.), Extending the Scope of Corpus-Based Research: New Applications, New Challenges (pp. 211-230). Amsterdam: Rodopi.

Nelson, G., Wallis, S., & Aarts, B. (2002). Exploring Natural Language: Working with the British

Component of the International Corpus of English (Vol. 29). Amsterdam/Philadelphia: John Benjamins.

Nesselhauf, N. (2004). Learner corpora and their potential for language teaching. In J. M. Sinclair (Ed.), How to Use Corpora in Language Teaching (pp. 125-152). Amsterdam/Philadelphia: John Benjamins.

Nesselhauf, N. (2009). Co-selection phenomena across New Englishes: parallels (and differences) to foreign learner varieties. English World-Wide, 30, 1-26.

O'Keeffe, A., & Mark, G. (2017). The English Grammar Profile of learner competence. International Journal of Corpus Linguistics, 22(4), 457-489.

O'Keeffe, A., & McCarthy, M. (2010). The Routledge Handbook of Corpus Linguistics. London: Routledge.

O'Keeffe, A., McCarthy, M., & Carter, R. (2007). From Corpus to Classroom: Language Use and Language Teaching. Cambridge: Cambridge University Press.

Partington, A. (1998). Patterns and Meanings: Using Corpora for English Language Research and Teaching. Amsterdam/Philadelphia: John Benjamins.

Pawley, A., & Syder, F. (1983). Two puzzles for linguistic theory. In Richards, J. C. & Schmidt, R. W. (Eds.), Language and Communication (pp. 191-227). London: Longman.

Pu, J. (2000). A survey of Chinese learners' use of English verbs in grammatical and lexical patterns. Modern Foreign Languages, 23(1), 24-44.

Pravec, N. A. (2002). Survey of learner corpora. ICAME Journal, 26(1), 8-14.

Quirk, R. (1972). A Grammar of Contemporary English. London: Longman Group.

Rees, G. P. (2018). A phraseological multi-discipline approach to vocabulary selection for English for academic purposes. PhD thesis. Universitat Pompeu Fabra.

Renouf, A., & Sinclair, J. (1991). Collocational frameworks in English. English Corpus Linguistics, 128-143.

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/New

York: Longman.

Rundell, M. (2010). Macmillan Collocations Dictionary for Learners of English. Oxford: Macmillan

Publishers Ltd.

Schachter, J., & Celce-Murcia, M. (1977). Some reservations concerning error analysis. TESOL

Quarterly, 11(4), 441-451.

Schneider, E. W. (2007). Postcolonial English: Varieties around the World: Cambridge: Cambridge

University Press.

Schneider, G., & Gilquin, G. (2016). Detecting innovations in a parsed corpus of learner English.

International Journal of Learner Corpus Research, 2(2), 177-204.

Sinclair, J. (1991). Corpus, Concordance, Collocation. Oxford: Oxford University Press.

Sinclair, J. (1996). The search for units of meaning. Textus, 9, 75-106.

Sinclair, J. (2004). Trust the text: language, corpus and discourse. London/New York:

Routledge.

Sinclair, J. (2007). Collocation reviewed. Manuscript, Tuscan Word Centre, Certosa di Pontignano,

Italy.

Spolsky, B. (1979). Contrastive analysis, error analysis, interlanguage, and other useful fads. The

Modern Language Journal, 63(5/6), 250-257.

Stefanowitsch, Anatol and Stefan Th. Gries. (2003). Collostructions: investigating the

interaction of words and constructions. International Journal of Corpus Linguistics

8.2:209-243.

Stubbs, M. (2001). Words and Phrases: Corpus Studies of Lexical Semantics. Oxford:

Blackwell Publishing.

Stubbs, M. (2007). An example of frequent English phraseology: Distribution, structures and

functions. In R. Facchinetti (Ed.), Corpus Linguistics 25 years on (pp. 89–105). Amsterdam: Radopi.

Virtanen, T. (1998). Direct questions in argumentative student writing. In S. Granger (Ed.), Learner English on Computer. London/New York: Longman.