Use of lexical bundles in academic writing in English

by expert writers, native students, and non-native students

in Applied Linguistics

Department of Language and Linguistics

University of Essex

First submission

October 2021

Revised submission

21 May 2022

Table of Contents

List of Tables

List of Figures

Abstract

This study compared the use of lexical bundles in academic writing in Applied Linguistics across three corpora: expert writers, native students and non-native students. The expert corpus consisted of articles published in Applied Linguistics journals; the native student corpus consisted of MA dissertations of native English students who did an Applied Linguistics Masters in English universities. The non-native student data consisted of Applied Linguistics MPhil dissertations of Pakistani students who did their MPhil in Pakistani universities. The size of the three corpora were as follows: native student corpus:312981, non-native student corpus:502945, expert writers' corpus:505958. The highly frequent bundles used in the three corpora were categorized into structural and functional categories (Hyland, 2008). These bundles were analyzed quantitatively as well as qualitatively.

The findings revealed that the expert writers were different from native and non-native students in their use of structural and functional bundles. The expert writers used more Phrasal bundles and more bundles for organizing the text than the two student groups. The expert writers also showed better control of bundles for hedging. The students, on the other hand, used more bundles for describing research. Occasionally, they used vague and informal bundles,

especially for quantifying. The non-native Pakistani students used far more bundles for describing the procedures of research and used far more bundle tokens than the other two groups. This might be due to the larger size of their dissertations.

Interestingly, most of the differences between expert and student writers in their use of bundles applied to both sets of students. This suggests that the main challenge for all students is learning the conventions of academic writing, rather than any problems linked to non-nativeness. Therefore, the appropriate use of bundles in academic writing might need to be taught more explicitly to both native and non-native students.

Acknowledgements

I want to thank Almighty Allah who gave me knowledge and strength. I am indebted to my supervisor and mentor, Professor Florence Myles, for her unfailing support, for the countless hours she spent in helping me to finish my thesis, but most of all for her friendship and mentoring which made me believe in myself.

I want to thank Professor Dr.Saiqa Imtiaz Asif Bahauddin Zakariya University Multan and Dr Muhammad Faisal NCB&E Multan for their help in the data collection. I want to thank the faculty members of the Department of language and linguistics, University of Essex: Dr Sophia Skoufaki, Professor Monika Schmid, Dr Laurel Lawer, for taking time to help me with the statistical analysis. I want to also thank Professor Paul Baker (Lancaster University), Professor Paul Thompson (Birmingham University), Professor Viviana Cortes (Georgia State University) and Professor Yves Bestgen (Université Catholique de Louvain-la-Neuve), for answering my emails and helping me in exploring methodological issues.

I want to thank my friends Dr Zia Ullah Khan (University of Essex), Dr Muhammad Ghufran, Dr Zardad Khan, and Dr Nosheen Khan (Abdul Wail Khan University) for helping me in statistical issues.

I am grateful to my parents, brothers and sister for their prayers and support. I thank my cousin Hamza Khan for helping me in editing my thesis. Finally, I want to thank my wife, Rabia, my son, Muhammad Haris Khan, and my daughter Hikmah Khan whose love and delightful disruptions enabled me to keep calm and complete my thesis.

Chapter 1

Introduction

1.1 Background

The use of appropriate formulaic language in academic writing is challenging for native and non-native students (Chen & Baker, 2010; Shin, 2019). In academic discourse, there are 'multi-word sequences that recur most frequently and are distributed widely across different texts' (Biber, 2010, p.170). These multi-word sequences are known as lexical bundles. Chen and Baker (2010) define lexical bundles as follows:

> …continuous word sequences retrieved by taking a corpus-driven approach with specified frequency and distribution criteria. The retrieved recurrent sequences are fixed multi-word units that have customary pragmatic and/or discourse functions, used and recognized by the speakers of a language within certain contexts. (p.30)

The lexical bundles have been investigated in terms of their structure and function. In terms of their structure, lexical bundles have been categorized into the following categories: Noun-based bundles, Preposition-based bundles, Verb-based bundles, and Other structures (Biber et al., 1999). In terms of their discourse functions, the lexical bundles have been categorized into

the following categories: Research-oriented bundles, Text-oriented bundles, and Participant-oriented bundles (Hyland, 2008a). Research has shown that the structural characteristics and discourse functions of lexical bundles vary in different registers, e.g., speech and academic writing (Biber et.al., 1999; 2004). The other studies on lexical bundles have shown that the use of bundles varies in different disciplines, therefore the uses of lexical bundles are known as discipline specific (Cortes, 2004; Hyland, 2008b). To master the bundles used in academic writing is important because the use of these genre-specific bundles is valued by academic writers and demonstrates the writer's competence as an academic writer. Hence, the effective use of these bundles is a sign of appurtenance to the community (Cortes, 2004; Hyland, 2008a). My interest in this topic stems from my own experience while writing an MPhil thesis at Bahauddin Zakariya University Multan and having difficulties in using appropriate language. It also came to my notice that Pakistani students generally face difficulties in using appropriate English language for writing in research setting (Fareed et al., 2016).

Against this backdrop, I decided to undertake this research focusing on the use of lexical bundles in non-native student dissertations and to compare their use of bundles with native students and expert writers. Expert writers are taken to be authors of research articles published in top tier journals, as students are expected to follow the language of research articles. The

research article is a new academic genre that is equally important for native and non-native students to master. As Hyland (2008a) notes,

> The research article is not only the principal site of disciplinary knowledge-making but, as Montgomery (1996) has it, "the master narrative of our time". One reason for this pre-eminence is the value attached to the processes of peer review as a control mechanism for transforming beliefs into knowledge, while another is the prestige attached to a genre which restructures the processes of thought and research it describes to establish a discourse for scientific fact creation. Consequently, the article is often presented to students as a model of good academic writing and as an ideal to be emulated as far as possible. (p.47)

So, both the research articles and student dissertations are the two important parts of academic discourse. Academic discourse that represents the conventions of the academia and determines the knowledge itself, is the central theme of this study. In the next section, I will define the term academic discourse and will discuss its importance.

1.2 Academic discourse

Hyland (2009) defines academic discourse in the following words:

> Academic discourse refers to the ways of thinking and using language which exist in the academy. Its significance, in large part, lies in the fact that complex social activities like educating students, demonstrating learning, disseminating ideas and constructing knowledge, rely on language to accomplish. Textbooks, essays, conference presentations, dissertations, lectures and research articles are central to the academic enterprise and are the very stuff of education and knowledge creation. (p.1)

The significant role played by academic discourse is more than the conventions, rather it involves the socialization of the students through assigning new social roles and identities to the students. These new social roles in the academic discourse poses problems for the newcomers as they have to assume new roles and develop new identities. The problems are more complicated for the second langue learners who might have learned different conventions in their first language academic writing. Learning the conventions of academic discourse put pressure on second language learners. Hyland (2009, p.6) further explains the nature of this pressure of the learners and notes ‘These frequently demand that students are more explicit about the structure and purposes of their texts, more cautious in making claims, clearer in signposting connections, and generally that they take more responsibility for coherence and clarity in their writing.’

Considering the importance of academic discourse for the second language learners, this research aims at exploring the use of lexical bundles in the written academic discourse of expert writers, native students and non-native students, in the field of Applied Linguistics. For this purpose, the distribution and use of lexical bundles is compared across the three corpora collected from these three cohorts.

1.3 Research on lexical bundles

Research on lexical bundles has shown differences in the use of bundles across different registers. Biber et al. (1999; 2004) found that the structural and functional characteristics of bundles are different in speech and (academic) writing. The lexical bundles used in speech tend to be verb-based and are used for presenting speakers' stance, whereas the bundles used in academic writing tend to be phrasal and are used for organizing the text. The research on disciplinary variation showed that different bundle types are used in different disciplines (Cortes, 2004; Hyland, 2008a). Studies that investigated the use of lexical bundles in native expert, native student and non-native student corpora showed differences between native expert writers and novice students. Research comparing the use of lexical bundles has yielded conflicting results. Some studies have reported differences (Adel & Erman, 2012; Bychkovska & Lee, 2017; Lu & Deng, 2019) while others have found striking similarities (Chen & Baker,

2010; Shin, 2019). The use of lexical bundles in native and non-native expert writers has also been investigated, and significant differences were found between the two groups of writers. (Pan et.al., 2016). So, previous research on lexical bundles has highlighted their important role in academic writing, but no clear picture has emerged about similarities and differences in their use across different populations. The current study further explores the role of bundles in academic writing and any differences and similarities between expert writers, native and non-native students. In the next section, I will briefly explain the methodology of the current research.

1.4 Current research

This study is a corpus-driven study for which a specialized corpus was developed. Two genres were selected: research-articles and Masters student dissertations. The data for research articles was collected from prestigious online journals in the field of Applied Linguistics. The native student data was collected from native English students who passed their Masters from Universities in England, and the non-native student data was collected from non-native English students who passed their MPhil degrees from different institutes of higher education in Pakistan. The expert corpus consisted of 67 texts and 505,958 words, the data for native students consisted of 20 texts and 312,981 words, and the non-native student data consisted of

19 texts and 502,945 words. Efforts were made to make the three corpora representative of the population that is being studied. Although it was nearly impossible to match the three corpora exactly, some measures were taken to make the three corpora as close to each other as possible. These steps were taken regarding the size of the three corpora, selection of sub-disciplines; data collection for the sub-disciplines across the three corpora, setting the frequency and dispersion criteria based on the size of each corpus. Making the size of the three corpora equal was important as the difference in size could affect the results of the three corpora. The steps were taken to make the size of the three corpora closer to each other. Therefore, non-native and the expert corpora are almost similar in size, representing 502945 words and 505958 words respectively. But native student corpus is smaller in size as it was difficult to make the size closer to the other two corpora because of time constraints. It was decided to neutralize the impact of small size corpora through setting a higher raw frequency for the native student corpus. Then, for the selection of the sub-disciplines in each corpus, steps were taken to ensure that each of the three corpora represents the data from similar sub-disciplines with nearly similar number of words. For example, steps were taken to ensure that the three corpora represent similar sub-disciplines of Applied Linguistics (see Section 3.3). Following previous research on lexical bundles (Adel & Erman, 2012; Biber et.al., 1999; 2004; Bychkovska & Lee,

2017; Chen & Baker, 2010; Cortes, 2004; Hyland, 2008a; Lu & Deng, 2019; Pan et.al., 2016; Shin, 2019; Staples et al., 2013), only 4-word lexical bundles were selected for investigation, and different frequency criteria and dispersion criteria were set for extracting lexical bundles from the three corpora in order to make statistical analysis possible (Adel & Erman, 2012; Bychkovska & Lee, 2017; Lu & Deng, 2019; Pan et.al., 2016; Shin, 2019). For the purpose of extracting lexical bundles, AntConc was used (Anthony, 2020). The loglikelihood test determined the significance of differences in the frequencies of lexical bundles across the three corpora. Finally, a comparison of the use of bundles was conducted across the three corpora.

1.5 Significance of research on lexical bundles

Research on lexical bundles is an important area of research that has helped researchers, linguists and second language teachers in understanding the role of formulaic sequences in native and non-native student writing. It has highlighted the important role of register variation and has helped understand the role of expertness and nativeness in the use of lexical bundles. For example, the use of bundles for hedging has been one of the areas of difficulties for non-native students (Adel & Erman, 2012; Bychkovska & Lee, 2017; Chen & Baker, 2010; Cortes, 2004; Hyland, 2008a; Lu & Deng, 2019). In the following section, I will define and discuss hedging and its importance in writing.

1.5.1 Hedging in academic writing

Hedging 'refers to any linguistic means used to indicate either a) a lack of complete commitment to the truth value of an accompanying proposition, or b) a desire not to express that commitment categorically (Hyland, 1998, p.1.). The use of hedging is an important technique in academic writing (Hyland, 1998). It is an essential part of the conventions of the academic discourse which demands the use of cautious and precise language while presenting claims or evaluation of a proposition (Hyland, 2009). Describing hedging as an important feature of academic discourse, Hyland (1996, p. 440) notes that 'Almost all academic discourse is a balance of fact and evaluation as writers try to present information as fully, objectively, and accurately as possible.' Hedging fulfils this purpose of maintaining balance between facts and writers' evaluation in academic writing.

It is important to mention that hedging devices also serve the purpose of establishing a relationship between the writers and their audience. These devices present writers' evaluation on a proposition in a way that they do not fully commit themselves to a proposition. Therefore, there is a space for the readers to criticise the claim presented by the writers. Emphasizing the importance of hedging devices, Hyland (1996) further explains the purpose of hedging devices as follows:

> They imply that a claim is based on plausible reasoning rather than certain knowledge and so both indicate the degree of confidence it might be wise to attribute to a claim while allowing writers to open a discursive space for readers to dispute interpretations. (p.440)

Hyland (1996, p.440) also described the types and functions of hedging devices as follows: Content-motivated hedges and Reader-motivated hedges. 'Content-motivated hedges mitigate the relationship between what a writer says about the world and what the world is thought to be like.'

Content-based hedges perform two functions: a. *accuracy-based hedges*: these hedges precisely present the claim with a degree of accuracy of a proposition. e.g., the use of adverbials *possibly, almost completely, generally*. b. *writer-based hedges*: these devices allow the writers to limit their commitment to a proposition, e.g., *It seems that, the current data indicates that* (Biber, 2006; Hyland, 1998). Reader-motivated hedges are used to build a relationship with the reader and to allow a room for error in presenting the evaluation of a claim or proposition. They allow the writers not to impose their evaluation rather they leave room for the alternative evaluation, e.g., *I believe that, I suggest that, my analogy is*.

The use of bundles for organizing text and the use of informal and vague words have been other areas of difficulty for non-native students. Non-native students might therefore need to be trained in certain areas of bundle use, such as in the use of hedging devices or in using more formal and precise language. Moreover, there are some sub-types of formulaic bundles that are not usually acquired spontaneously by non-native English students, such as some idiomatic expressions (Hinkel, 2017). Research on lexical bundles has thus enabled the researchers to uncover these areas of academic writing that might need to be taught explicitly to non-native students (Adel & Erman, 2012; Bychkovska & Lee, 2017; Chen & Baker, 2010; Pan et.al., 2016; Lu & Deng, 2019; Shin, 2019; Staples et al., 2013). It might therefore have implications for language teachers and syllabus designers of courses likes English for Academic purposes, and English for specific purposes. To conclude, I will provide an outline of this dissertation in the following section.

1.6 Outline of the thesis

Chapter 1 is the introductory chapter. It presents the background and motivation behind this research. It presents a brief overview of some important findings on the topic of this research, and also of the methodology followed in this study.

Chapter 2 is the literature review. It begins with an introduction of approaches to research on formulaic language follows, focusing on its frequency-based dimension. Lastly, lexical bundles, i.e., frequency-driven formulaic sequences, are defined and discussed in detail, along with a critical review of previous studies on the use of lexical bundles in academic writing.

Chapter 3 presents the methods of data collection and the methodology adopted for analysis in this study. The procedures adopted for collecting data and compiling the three corpora are described, and details about the corpus software, AntConc, adopted in this study, are discussed in detail. Moreover, the details of the statistical analysis test, Loglikelihood test, will also be discussed at the end of this chapter.

Chapter 4 presents the results of the analysis of the three corpora. The chapter contains an analysis of the most frequent bundles, and of the structural and functional characteristics of bundles in each corpus. A comparative analysis of the three corpora is then carried out, outlining similarities and differences between the three corpora.

Chapter 5 is the discussion chapter. It discusses the results of this study in the light of previous research and highlights the main features of bundle use in the expert, native and non-native students' corpora.

Chapter 6 is the Conclusion, which summarises the study and presents its limitations, making some recommendations for future research.

Chapter 2

Literature Review

2.1 Introduction

There are three main sections of this chapter. The first section will introduce formulaic language and the two approaches relevant to the current study: the phraseological approach and the frequency-based approach. The second section deals with lexical bundles, their definition and classification, both structural and functional. In this section, I will also review previous studies on the use of lexical bundles in academic writing, including studies on the use of lexical bundles of non-native English Pakistani students. In the final section, I will take an overview of the corpus research implications for English language teaching and learning.

2.2 Formulaic language

Formulaic language has been an important area of research in corpus studies. Gablasova et al. (2017, p.156) note that 'Corpora represent a rich source of information about the regularity, frequency, and distribution of formulaic patterns in language'. The seminal work by Sinclair (1991) presented a new aspect of language, the idiom principle, based on the idea that language is a co-selection of strings of words that constitute single choices. Meanings are based on these

strings of words, not on individual words. Partington (1998) also worked on collocational associations which showed that even synonyms like *pure*, *complete*, and *absolute* are not interchangeable because their frequent collocates are different.

Another important area of research on formulaic language has been the use o formulaic language in various registers of academic discourse. Altenberg (1998) investigated the use of recurrent word sequences in spoken English. Biber et al. (1999) compared the use of lexical bundles in conversation and academic writing. In a later study, Biber et.al. (2004), compared two more registers: university classroom lectures and textbooks with conversation and academic prose. Cortes (2004) compared published texts with university student writing in the fields of history and biology (Biber & Barbieri, 2007). These studies found that bundle types, their structural characteristics and functions were very different in spoken and written registers. In speech, more bundles were generally used than in writing. The structural analysis of the bundles showed that the majority of bundles in speech consisted of Verb-based bundles, whereas the majority of bundles in writing consisted of Noun-based and Preposition-based bundles. There were differences in the functions of lexical bundles as well. In speech, lexical bundles were used primarily for stance functions and discourse organizing functions. The Stance bundles are used for presenting the writers' or speakers' evaluation, whereas the

discourse organizers are based on bundles that function for elaborating the topic and describing research. On the other hand, in writing, bundles were used for referential functions. The referential functions include the functions for organizing new information in text and establishing coherence in the text.

One of the issues regarding formulaic language is that it is a slippery term (Paquot, 2008; Paquot & Granger, 2012; Wray, 2013) because it has many meanings. Describing the problematic nature of the term formulaic language, Siyanova-Chanturia (2015) notes:

> Formulaic language can be of many different kinds, such as, collocations (fast food), binomials (black and white), multi-word verbs (rely on), idioms (tie the knot), speech formulae (what's up?), discourse markers (by the way), lexical bundles (as well as), expletives (damn it!), grammatical constructions. (p.286)

Addressing the issue of ambiguity with formulaic language, Myles and Cordier (2017) state:

> The term formulaic sequence has been used with a multiplicity of meanings, including in the SLA literature, some overlapping but others not, and researchers have often been unclear in defining precisely what they are investigating, or in limiting the implicational domain of their findings to the type of formulaicity they have focused on. (p.4)

Different definitions of formulaic language are based on different approaches to formulaicity, which I will discuss in the next section.

2.2.1 Phraseological approach

In the phraseological approach to formulaic language, the analysis of the word sequences is guided by the syntactic and semantic analysis of word sequences that are non-compositional. Moon (2015) defines non-compositionality as follows:

> [Non-compositionality] refers to the extent to which a string of words has a unitary meaning that cannot be derived by decoding, literally, each component word. Sometimes unitary meanings are obscure and impossible to retrieve synchronically (as with *rain cats and dogs* = 'rain heavily'. Others are more amenable to interpretation (*alarm bells ring*) or have specialized pragmatic functions (*happy birthday*). (p. 3)

However, at times non-compositionality depends on the context. Sometimes, the meanings of an idiom can be interpreted. For example, idiom like *a piece of cake*. 'Thus non-compositionality is subjective, depending on individuals' linguistic and metaphorical competence and their decoding of component words' (Moon, 2015, p. 6). Similarly, phrasal verbs are semantically non-compositional but some of them have particles that are repeatedly

used in a particular sense and convey typical meanings. For example, the particles *up, out and away* convey the meaning of fulness or completeness (Moon, 2015). Therefore, there are idiomatic units that might be considered fully non-compositional, whereas there are others whose might be taken as partially non-compositional.

The most idiomatic units, whose meanings cannot be derived from the meanings of their constituents, are often presented as the defining components of formulaicity (Cowie, 1998). This approach draws a clear line between formulaic and non-formulaic sequences on the basis of their non-compositionality (Allerton, 1984).

Paquot and Granger (2012) also believe that opaque word sequences are considered the core of the phraseological approach. To explain formulaicity they describe the difference between compositionality and non-compositionality in the following words:

> …the non-compositionality of certain expressions, defining formulaicity in terms of either the degree to which the meaning of a word combination is predictable from the meaning of its parts or the degree to which words with similar meanings can be substituted into the phrase. Non-compositional phrases include idioms (e.g., kick the bucket, spill the beans) and certain collocations (e.g., curry favour, French window). The 'formal idioms' (Fillmore, Kay, & O'Connor, 1988) of construction grammar (e.g.,

> what's NP doing Y; the ADJ-er the ADJ-er) can also be included int this category as items which cannot be easily understood and/or produced without specific learning. (p.4)

Cowie's (1981) continuum goes from free combinations to pure idioms through restricted collocations, e.g., *blow a fuse,* and figurative idioms, e.g., *blow your own trumpet*. These items are defined as formulaic only on the basis of syntactic and semantic restrictions. On the one end of this continuum, there are word sequences that are variable, whereas on the other end there are fixed and opaque words sequences such as idioms. Idioms are considered to be the archetypical formulaic sequence under the Phraseological approach.

2.2.2 Frequency-based approach

The Frequency-based approach centres around the analysis of word sequences on the basis of frequency and statistics (Cortes, 2015) and it takes a different view of formulaicity. "It refers to statistically significant word co-occurrences, that is, lexical items occurring within a certain distance of the search item "with a greater frequency than the law of averages would lead you to expect" (Sinclair, 1987, p. 70). This approach introduced the notion of collocational associations, for example adjectives that frequently collocate with a noun form a close association. This approach is inductive. "Instead of adopting a top-down approach which

identifies phraseological units on the basis of linguistic criteria, it uses a bottom-up corpus-driven approach to identify lexical co-occurrences" (Cortes, 2015, p.199).

Under this approach, rather than linguistic criteria, statistical methods such as MI scores, are used to extract formulaic sequences. MI score is used as a measure of collocational strength. It 'measures the amount of non-randomness present when two words occur' (Hunston, 2002, p.71). In this approach, formulaic sequences that were considered free combinations, and therefore not formulaic, in the Phraseological approach, such as *drink coffee* might be considered collocations and therefore formulaic.

The frequency-based approach has made it possible to include many word sequences in the realm of formulaic language that were considered free combinations in the phraseological approach (Cortes, 2015).

In this approach the semantic and syntactic boundaries are not restricted. Instead of semantic restriction, the frequency-based approach takes the notions of semantic prosody: "the positive or negative connotations shared by the set of collocates that co-occur with a word" (Biber, 2010, p.5). For example, the verb *commit* has a negative semantic prosody as it occurs with nouns like *crime*, *offenses*, *suicide* etc. Similarly, Sinclair (1991, p.74) notes that the nouns that co-occur as the subject of *set in* are mostly unpleasant states of affairs, such as *rot*, *decay*,

malaise, *despair*, *infection*, *disillusion*, and so on. It has opened up a "huge area of syntagmatic prospection" (Sinclair 2004, p.19) encompassing sequences like frames and collocational frameworks, consisting of sequences containing one or more free slots. Examples include 'a + ? + of', 'an + ? + of' (e.g. *a kind of*; *an example of*), 'be + ? + to', and 'too + ? + to' (e.g. *be nice to; too lazy to*) (Renouf & Sinclair, 1991, p.128). Stubbs (2007) has referred to these multi-word sequences as phrase-frames.

A typical example of a frequency-based approach is the extraction and analysis of recurrent word sequences known as lexical bundles. Biber et al. (1999, p. 990ff) defined lexical bundles as "simple sequences of word forms that commonly go together in natural discourse". These formulaic sequences are extracted from a corpus on the basis of a frequency criterion. They are mostly syntactically and semantically regular, e.g., *the use of the*, *in the form of* etc., and these bundles perform important discourse functions, such as hedging, organizing, etc.

For the corpus analysis, two frequency-based approaches, Corpus-based approach and Corpus-driven approach, have emerged (Tognini-Bonelli, 2001). In the next section, I will discuss these two frequency-based approaches and some corpus studies that have followed these Corpus approaches.

2.2.2.1 Corpus-based approach

The corpus-based approach used for the identification of multiword sequences is based on the analysis of word sequences that are defined as formulaic, e.g., fixed expressions, idioms, collocations, phrasal verbs (Cortes, 2015; Moon, 1998). One of the major contributions of corpus-based lexical studies has been the insight that collocational associations are an important consideration for describing the meaning of a word. Considering the example of three copular verbs: *turn*, *come*, and *go* which have a similar dictionary meaning: *to become*, or *to change to another state* (Biber, 2010). However, these three copular verbs have completely different collocational associations. For example, common adjectives that collocate with *turn* are colours like *black*, *brown*, *white* etc. The most common adjectives following *come* denote a process of change to a more dynamic condition, such as *alive*, *awake*, *clean*, *loose*, and *unstuck*. And in contrast to both other verbs, the most common adjectives following *go* are all negative: *crazy*, *mad*, and *wrong* (Biber et al., 1999). Sinclair (1991) has provided a thorough description of some of the collocations of *decline, yield*, and *set in*. Partington (1998) discusses the word *sheer* and its supposed synonyms *pure*, *complete*, and *absolute*, showing how these words are not at all interchangeable when considered from the perspective of their frequent collocates. Mahlberg (2005) examines common nouns in English (e.g., *time, day, man,*

woman, people, thing, way), describing their meanings and use with respect to their collocational associations.

Biber et.al. (1998) show how the near-synonyms *big*, *large*, and *great* co-occur with very different sets of collocates (e.g., *big enough* versus *large number* versus *great deal*), and further shows how the collocational associations are very different in fiction versus academic writing.

Other collocational studies taking a register perspective include those by Gledhill (2000) and Marco (2000), which both describe the functions of collocations in academic research writing.

2.2.2.2 Corpus-driven approach

This type of approach includes studies of lexical collocations in which a corpus is used to discover the collocations of a target word. This type of studies is considered corpus-driven, even though a preliminary step is based on the analysts' selection of interesting target words for analysis (Cortes, 2015). The studies that are based on a corpus-driven approach have identified the most common sequences of words in spoken and written registers by investigating the corpus. These sequences of words, 'often referred to as lexical bundles, are usually not idiomatic and are not complete structures, but they are important building blocks of discourse' (Biber, 2010, P.55).

A number of studies have examined the use of word sequences in different genres: for example, Altenberg (1998) in spoken English, Biber et al. (1999) in speech and academic writing. Following Biber et al. (1999), various studies have examined lexical bundles in terms of their structural characteristics and functions in different registers. For example, some studies have investigated the use of lexical bundles in university lectures and textbooks (Biber, et.al., 2004; Nesi & Basturkmen, 2006), others have compared published writing with the writing of university students (Cortes, 2004). Biber and Barbieri (2007) compared university institutional and advising registers and Partington and Morley (2004) investigated the use of lexical bundles in political debates.

Corpus-driven studies have shown that the use of lexical bundles is very different in spoken and written registers. The frequency, structure and functions of lexical bundles are different in different registers. For example, more lexical bundles are generally used in spoken registers than academic registers. Spoken registers are dominated by Verb-based bundles, whereas Noun-based and Preposition-based bundles are more common in academic registers. In speech, bundles are mainly used for stance and discourse organizing functions, whereas in writing, bundles are used for referential functions (Biber, 2010). In the next section, I will discuss define and discuss different features of lexical bundles in detail.

2.3 Lexical bundles

The term ‘lexical bundles’ was first used by Biber et al. (1999) in a study of Longman Grammar of Spoken and Written English. Biber (2010) defines the term lexical bundle as follows:

> Lexical bundles are defined as the multi-word sequences that recur most frequently and are distributed widely across different texts. Lexical bundles in English conversation are word sequences like *I don't know if* or *I just wanted to.* They are usually neither structurally complete nor idiomatic in meaning. (p.12)

Chen and Baker (2010) define lexical bundles as:

> …continuous word sequences retrieved by taking a corpus-driven approach with specified frequency and distribution criteria. The retrieved recurrent sequences are fixed multi-word units that have customary pragmatic and/or discourse functions, used and recognized by the speakers of a language within certain contexts. (p.30)

Describing the characteristics of lexical bundles, Biber (2010) enlists the following three features of lexical bundles:

- Lexical bundles are extremely common in speech and writing

- Lexical bundles are usually not idiomatic in meaning and not perceptually salient, e.g., *do you want to*, *I don't know what*. The meanings of these bundles are transparent.
- Lexical bundles are usually structurally incomplete units. 15% in conversation and 5% in academic writing are complete structural units. Most of the bundles bridge two structural units, e.g., *I want to know, well that's what I, in the case of, the base of the* etc. (p.5)

Although lexical bundles usually are structurally incomplete and usually non-idiomatic, they have strong grammatical correlates. Based on these structural correlates, they can be divided into two types: clausal and phrasal. The clausal structure usually consists of a verb phrase and sometimes a dependent clause. On the other hand, phrasal structures consist of noun phrase components. The clausal and phrasal distinction is important to understand register differences as it was found that almost 90% of 4-word lexical bundles in conversation were clausal fragments, whereas almost 70% of the bundles in academic writing happened to be phrasal with embedded prepositional phrase fragments (Biber et al.,1999).

Iin the same way as structural characteristics of lexical bundles vary according to register; their functional characteristics also do so. For example, stance bundles, used for presenting the

speaker's or writer's evaluation, are mostly found in conversation and referential bundles are mainly used in academic writing (Biber et al., 2004; 2007).

There is also a strong association between structural and functional types of lexical bundles, e.g., most stance bundles employ verb or clause fragments, while most referential bundles are composed of noun phrases or prepositional phrases (Biber, 2010; Pan et.al., 2016).

2.3.1 Lexical Bundles: Frequency, dispersion, and size

Lexical bundles are identified using a corpus-driven approach, based solely on distributional criteria (rate of occurrence of word sequences and their distribution across texts). Frequency is the most important characteristic of lexical bundles (Cortes, 2004). The frequency cut-off that is used for extracting lexical bundles is arbitrary (Conrad & Biber, 2005), and different studies set different frequency criteria based on a range of factors such as corpus size, size of bundles etc. I will discuss these factors in detail in the methodology chapter. Biber et al. (1999), in his pioneer study, set the frequency criterion at 10 occurrences per million words for 3–4-word lexical bundles, though for longer sequences he set a lower frequency, 5 occurrences per million words, because longer sequences occur less frequently as compared to 3-4 word lexical bundles. Dispersion is another important criterion for extracting lexical bundles. Dispersion ensures that the extracted lexical bundles are spread widely across the corpus and represent

most of the corpus. A corpus may regroup many different corpora (for example, a corpus of university lectures may contain 100 individual lectures), and each one of those might be of a different size. So, there is a possibility of a bundle type being idiosyncratic, i.e., all the tokens of this particular bundle are found only in one or two corpora. Dispersion rate is set to ensure that a lexical bundle occurs across the range of corpora. Another feature of lexical bundles is the size of word sequences, which is determined based on the scope and nature of the study. Most of the studies have investigated 4-word lexical bundles (Adel & Erman, 2012; Bychkovska & Lee, 2017; Chen & Baker, 2010; Cortes, 2004; Hyland, 2008a; 2008b; Pan et.al., 2016; Lu & Deng, 2019; Shin, 2019), as 5-word bundles are very rare and 3-word bundles are very frequent. Moreover, the majority of 3-word bundles are included in the 4-word bundles. Therefore, 4-word bundles are usually considered to be the most appropriate size (Hyland, 2008a).

2.3.2 Fixedness

Lexical bundles are fixed sequences of words that cannot be substituted. However, Cortes (2004), noticed that the fixedness of lexical bundles is due to the frequency cut-off, due to which only those word sequences are extracted which meet the frequency criteria. Salazar (2011) gave the example of the lexical bundle, *are expressed as* that might meet the frequency

criteria due to high frequency, and its variant *is expressed as* might not meet the frequency due to low frequency, thus the bundle, *are expressed as* would emerge as a fixed word sequence.

Biber (2009) examined the fixedness of 4-word lexical bundles and identified that some 4-word lexical bundles tend to have variable slots. So, he came out with the concept of frames and fixed bundles. The bundle frames contain one or two slots might be filled with different words, e.g., *the/?/of/the*, is a frame in which different words can be used in the second slot (e.g. *the top of the; the meaning of the*). In contrast, a lexical bundle such as *on the other hand* remains unchanged in 50% of its occurrences; is therefore treated as a fixed bundle.

Chen and Baker (2010) found that native expert writers used significantly more types and tokens of Noun-based bundle frames 'the + Noun + of + the/a' and Preposition-based bundle frames 'In + the + Noun + of'. Based on these findings, they concluded that native expert writers used more lexical bundles with variable slots in academic writings than were used by novice and non-native students. Staples et al. (2013) also examined the fixedness of lexical bundles used by non-native students at different proficiency levels. They found that there was no significant difference in the use of fixed and variable slots across proficiency levels. Lexical bundles have strong grammatical correlates, and also function as discourse markers (Biber,

2010). Therefore, they have usually been categorised according to structural and functional information.

2.3.3 Lexical bundles: Structure and function

Biber et. al. (1999) presented a broad categorization of the structural or grammatical correlates of lexical bundles. Biber et. al. (1999) investigated the use of multiword sequences which they called lexical bundles through a comparison of two registers: academic prose and conversation. About the structural characteristics of these bundles, Biber et. al. (1999) noted that lexical bundles mostly represent incomplete units and very few of them were based on complete structural units (15% in conversation, 5% in academic prose). However, Biber et. al. (1999) found that lexical bundles have strong grammatical correlates. Furthermore, they also found that the grammatical correlates of the bundles vary in different registers. For example, in conversation, the common structure of the bundles was found to be clausal, Pronoun + verb + complement, e.g., *I want you to, it's going to be*. On the other hand, in academic prose, the common structure was found to be phrasal, parts of noun phrases or preposition phrases, e.g., *the nature of the, as a result of*. Keeping in view these characteristics of lexical bundles found in different registers, a structural taxonomy was built by Biber et al. (1999). The following structural taxonomy of lexical bundles represents the bundle structures typical of academic

prose. Table 2.1 presents the main and sub-categories of lexical bundles according to their structure.

Table 2.1 Structural classification of lexical bundles in academic prose

Structures	**Examples**
Noun phrase with of- phrase fragment	*the beginning of the, the shape of the*
Noun phrase with other post-modifier fragments	*the way in which, the extent to which*
Prepositional phrase with embedded of-phrase	*as a result of, in the case of*
Other prepositional phrase fragment	*at the same time, on the other hand*
Anticipatory it + verb / adjective phrase	*it is possible to, it should be noted that*
Passive verb + prepositional phrase fragment	*is shown in figure, is based on the*
Copula be + noun / adjective phrase	*is one of the, is part of the*
(Verb phrase) + that- clause fragment	*has been shown that, that there is no*
(Verb/ adjective) + to-clause fragment	*are likely to be, has been shown to*

Adverbial clause fragment	*as we have seen, if there is a*
Pronoun/ noun phrase+ be (+…)	*this is not the, there was no significant*
Other expressions	*as well as the, than that of the*

In the Longman Spoken and Written English Corpus (LSWE), the twelve categories of lexical bundles were grouped in academic prose. In the current study, these twelve structural categories were carried out on the extracted lexical bundles from the expert corpus, the native student corpus, and the non-native student corpus. There was a complete match in the bundle structures found in the current study.

Biber et. al (2004) presented a functional taxonomy of lexical bundles as well. They classified bundles into three main functional categories: Stance expressions, Discourse organizers, and Referential expressions. These functional categories of lexical bundles reflect their meaning and the discourse function they perform in the text (Salazar, 2011).

Stance bundles present writers' evaluation on a proposition, e.g., *are more likely to, seems to have been*. These bundles are also used to show the obligation by the writer, e.g., *it is necessary*

to, it is important to, and to represent the writer's assessment about the ability of something, e.g., *it is difficult to, to be able to.*

Stance bundles also express attitudes or assessments of certainty that frame some other proposition. Discourse organizers reflect relationships between prior and coming discourse. Referential bundles make direct reference to physical or abstract entities, or to the textual context itself, either to identify the entity or to single out some particular attribute of the entity as especially important.

Table 2.2 ***Functional classification of lexical bundles (Biber et.al., 2004)***

Functions	**Examples**
Stance bundles	are often used to express a writer's evaluation of a proposition in terms of certainty or uncertainty (epistemic)
Epistemic	*are more likely to, it can be argued, the fact that the*
Obligatory/directive	*it is necessary to, that need to be, it has to be*
Ability	*it is difficult to, to be able to*
description	*the structure of the, the size of the*);

topic	*in the Hong Kong, the currency board system*)
Referential expressions	are characterized by the function of attribute specification
Framing	in the context of, the nature of the, the existence of
Quantifying:	a wide range of, the extent to which, in a number of
Place/time/text-deictic:	are shown in fig, at the same time
framing signals	*in the case of, with respect to the*
Discourse organizers	are used to structure texts
Topic introduction	essay is going to, last but not least, in this essay I
Topic elaboration	in more detail in, on the other hand, can be used to
Inferential	as a result of, in view of the, this is due to
Identification/focusing	one of the most, there would be no, we can see that

Hyland (2008a) presented a functional taxonomy that represents the functional categories used in academic writing. He based this functional taxonomy on Halliday's (1994) macro functions:

ideational, textual, and interpersonal and classified the functional categories of lexical bundles into three main categories: Research-oriented bundles, Text-oriented bundles, and Participant-oriented bundles. Describing this classification Hyland (2008a) briefly explains the purpose of bundle functions:

> The clusters in the corpus fall into three broad categories, which are loosely based on Halliday's (1994) linguistic macro functions: research, or real-world clusters, serve an ideational function, text-oriented clusters are combinations concerned with textual functions, and participant-oriented bundles express interpersonal meanings. (p.49)

The detail of these three categories and sub-categories is given in Table 2.3.

Table 2.3 ***Functional classification of lexical bundles***

Functions	**Examples**
Research-oriented bundles	Help writers to structure their activities. These activities refer to physical aspect of research activity, e.g., time, place, procedures of research, measurements etc.)
location bundles	*at the beginning of, at the same time*

procedure bundles	*the use of the, the role of the, the purpose of the*
quantification bundles	*the magnitude of the, a wide range of*
description	*the structure of the, the size of the*);
topic	*in the Hong Kong, the currency board system*)
Text-oriented bundles	These bundles are concerned with the organisation of the text (The organization of the text consists of bundles that show contrast or addition; explain the results and the outcome; refer to the study itself; describe the information in context)
transition signals	*on the other hand, in addition to the*
resultative signals	*as a result of, it was found that*
structuring signals	*in the present study, in the next section*
framing signals	*in the case of, with respect to the*

Participant-oriented bundles	These are focused on the writer or reader of the text
stance features	*are likely to be, it is possible that*); − address readers directly
engagement features	*it should be noted that, as can be seen*

So, both the taxonomies of functions of lexical bundles presented by Biber et.al. (2004) and Hyland (2008a) are helpful in understanding and analysing the discourse functions of lexical bundles used in academic discourse. However, for the current study, the functional taxonomy of bundles by Hyland (2008a) was carried out to the extracted bundles because this taxonomy was also based on the student writing and publish research articles that makes a complete match of the corpus of the current study. So, the taxonomy proposed by Hyland (2008a) was carried out to the retrieved bundles in the expert corpus, the native student corpus and the non-native student corpus. In the next section, previous studies on lexical bundles in academic writing will be evaluated.

2.4 Previous studies of lexical bundles in academic discourse

The use of lexical bundles in different register has attracted much interest, and particularly in academic writing, which tends to be constrained by many conventions.

2.4.1 Use of lexical bundles in different registers

Previous studies on lexical bundles have shown important register differences in the use of lexical bundles (Biber et.al., 1999; 2004; Cortes, 2004, Hyland, 2008a; 2008b).

Biber et al. (1999) in a pioneering study on lexical bundles, highlighted the features of lexical bundles in speech and writing. Their study was based on the Longman Spoken and Written English Corpus. The spoken section of the corpus consists of conversations, whereas the written section of the corpus consists of formal academic writing, each section containing around 5 million words. The study extracted lexical bundles in both the spoken and the written corpora. The frequency criterion was set at 10 bundles/million words, and the dispersion criterion was set at 5 texts.

The study found that lexical bundles were used more frequently in conversation (43 types) than the academic prose (19 types). It also found that more clausal bundles were used in conversation, e.g., *I want you to, it's going to be* etc., whereas in academic prose, more Noun-based and Preposition-based bundles were used, e.g., *the use of the, on the other hand* etc.

Their study also found that lexical bundles have strong structural correlates, that is they contain either a noun, a preposition or a verb, on the basis of which their structure can be classified. Two structures are dominant in academic prose: Noun phrases, e.g., *the end of the* and Prepositional phrases, e.g., *as a result of*. Furthermore, some frames of lexical bundles, e.g., 'the___ of the' were found to be highly productive. These frames with variable slots have nominal or prepositional elements which cooccur together and are more frequent in academic prose as compared to speech. Biber's later research also substantiated the highly frequent use of these formulaic frames in academic discourse (Biber & Barbieri, 2007).

In another study, Biber et al. (2004) compared the use of lexical bundles in four different registers: conversation, classroom teaching, textbooks and academic prose. The sample texts for the study were taken from the TOEFL 2000 Spoken and Written Academic Language Corpus (T2K SWAL), and the Longman Spoken and Written English Corpus. For the classroom lectures, six different disciplines (Business, Education, Engineering, Humanities, Natural science, Social science) were chosen and students at three different levels, lower division undergraduate, upper division graduate, and graduate, were selected. The texts were collected from following academic institutes: North Arizona university, Iowa State University, California State University at Sacramanto, Georgia State University. The corpus of Classroom

lectures consisted of 176 texts, and 124,800 words, whereas the corpus of Textbooks consisted of 87 texts and 760,600 words.

The study found that the use of lexical bundles was different in the four registers. The highest number of bundle types were used in classroom teaching (84 bundle types) and the least in academic prose (19). 43 bundle types were used in conversation and 27 in textbooks. The classroom lectures showed characteristics of both oral and written registers, and therefore used a larger number of bundle types.

The structural characteristics of bundles in the four registers were quite different from each other. In conversation, 90% of the bundles were Verb-based bundles, with 50% of the total bundles beginning with a personal pronoun, e.g., *I was going to, I thought that was* etc. In academic prose, 70% of the total bundles consisted of Noun-based and Preposition-based bundles, e.g., *the nature of the, a result of the* etc.

In the use of bundle functions, Stance bundles were the most common bundles in conversation with 29/43 types. In classroom lectures, both the Stance bundles and the Referential bundles were used equally frequently with 33/84 and 32/84 types respectively. In textbooks and in academic prose, bundles were primarily used for Referential functions (20/27 types in textbooks and 15/19 types in academic prose).

These differences in the use of lexical bundles in different registers also suggest that lexical bundles reflect the formal and informal use of language in different settings. The formal use of language reflects more distant stance, personal detachment and focused on the content, whereas the informal use of language is the use of colloquial language with more involved and personal stance (Hyland & Jiang 2017). Describing these differences in the formal and informal use of language reflected through different bundle structures, Conrad and Biber (2005) note the following:

> The structures typical of conversation are used for more personal expressions, particularly expressions of attitudes and desires, with bundles such as *I don't know what or you want me to*. The structures typical of academic prose are useful for specifying aspects of information with bundles such as *the nature of the, the extent to which, and as a result of*. These functional differences provide greater insight into lexical bundles' role in building discourse. (p.64)

For example, the use of stance bundles with first personal pronoun and second person pronoun are highly frequent in conversation and classroom lectures. The structure of these bundles with personal pronoun is particular to spoken registers, and it reflects the informal use of language. The structure of stance bundles in the written registers is different where 'Anticipatory it'

structures, e.g., *it is possible to* and 'unattended this', e.g., *this suggests that*, is used for presenting stance. In formal communication, the writers try to distant themselves from the claim or a proposition.

Cortes (2004) compared the use of 4-word lexical bundles in published research papers in the field of history and biology with unpublished students' term papers at three different study levels (undergraduate lower division, undergraduate upper division, and graduate students). They termed target bundles the lexical bundles found in the published research papers. Comparing the disciplinary use of lexical bundles, the study found some similarities and some differences: 54 types were found in history and 109 types were extracted in biology. Structurally, in history, Noun-based and Preposition-based bundles dominated, however, in biology more VP-based bundles were used. Similarly, more epistemic bundles, *are likely to be, is likely to be* etc., were used in biology for hedging but these bundles did not occur frequently in history.

Comparing the published and unpublished writing, the study found that very few target bundles occurred in the students' corpora. In the corpus of history students, 29/54 bundle types were found in only five or a smaller number of student papers at all levels. This showed that the student writers made a less frequent use of bundles for organizing the text.

The student writers very often repeated and overused a limited number of types of bundles, and the target bundles that they used most frequently were time markers. Some target bundles like *at the same time* were not used appropriately by the students, however; for example, they used *at the same time* for addition rather than for simultaneity.

Some target bundles were rarely or never used by the student writers, such as '*on the evolution of*', '*an order of magnitude*'.

So, the study suggests that lexical bundles are characteristic of specific disciplines and each discipline has distinct features of lexical bundle usage. The study did not find any developmental changes in the use of lexical bundles across three levels of students. But the comparison of published and unpublished writing reveals that the published writers use lexical bundles differently from the unpublished writers. The student writers show repetition and redundancy in their use of lexical bundles.

Hyland (2008a) also focused on disciplinary variation and compared the use of 4-word lexical bundles in three different genres: published research articles, PhD dissertations, and masters theses in pure and social sciences. The research articles consisted of 120,30 words in each of the four disciplines. The corpus of PhD and MA students consisted of 20 texts in each discipline, accounting for 1.9 million words, and 825,000 words respectively. The student data

was gathered from Cantonese-speaking students at 5 different universities in Hong Kong. In terms of disciplinary differences, the results of this study were similar to those of Cortes (2004). Hyland (2008a) found important differences in frequency and use of lexical bundles across the four disciplines investigated: Electrical engineering, Business studies, Applied linguistics, and Microbiology. The MA students used the greatest number of bundles with 149 types, whereas PhD students used 95 and the least number of bundles were found in research articles, 71. In the list of the 50 most frequent bundles, only about half were from the student corpora. Moreover, the frequencies of individual bundles were also higher in the student corpora. For example, the most frequent bundle, *on the other hand*, was twice more frequent in the MA texts, and three times more frequent in the PhD texts than in the research articles. Similarly, common bundles such as *at the same time* and *is one of the* were also significantly more frequent in both the student corpora than in the research articles. The overall frequent use of bundles in the student corpora is characteristic of speech rather than academic writing. (Biber et al, 1999; 2004)

In terms of structural distribution, the corpus of research articles was dominated by Noun-based and Preposition-based bundles. Previous research suggests that the majority of bundles used in academic discourse are phrasal in nature, whereas more clausal bundles are used in

conversation, mainly VP-based bundles (Biber et al., 1999; 2004). In terms of functional distribution, significantly more Text-oriented bundles and participant-oriented bundles were found in the corpus of research articles. Especially, more stance markers e.g., *it is possible to, it is obvious that, are more likely to, may be due to* etc. were used by the expert writers. These bundles are important as they work as hedging devices, and academic writers use them to detach themselves from the information. In the Text-oriented bundles, two sub-categories were significantly more frequent in the research articles: framing bundles and resultative bundles (the bundles that describe the cause of something, e.g., *as a result of,* or precede the results, e.g., *the result show that* etc). The study found that the PhD students also used more Phrasal bundles (45% tokens) and showed some awareness of the conventions of academic discourse. In terms of functional categories, they frequently used framing devices e.g., *in the case of, in relation to the* etc. These devices are used to focus the reader on a particular situation or to specify the conditions under which a statement can be accepted. In Participant-oriented bundles, the PhD students used more engagement features (70% of total Participant-oriented bundles) especially with 'anticipatory it' structure e.g., *it should be noted, it can be seen, it is important to* etc. The lack of stance features in the PhD dissertations showed these mature students' reluctance to maintain a personal voice. However, these students used more Text-

oriented bundles than Research-oriented bundles, which rendered their academic writing well organized and reader friendly.

In the Masters dissertations, Research-oriented bundles were the most frequently used bundles representing 48.6% types of the total bundles, followed by Text-oriented bundles (42.5% of the total bundles), and Participant-oriented bundles were the least frequently used category (8.9% of the total bundles).

The frequent use of Research-oriented bundles made student writing more research-focused. In Research-oriented bundles, almost 25% of the total tokens were procedure bundles used to describe procedures and focus on research objectives or context rather than focusing on the structure of the text. The less frequent use of Participant-oriented bundles, as compared to PhD students and expert writers, made students maintain authorial anonymity.

2.4.2 Use of lexical bundles in native and non-native students

As mentioned before, formulaic sequences are an integral part of language use in general (see Section 2.2). Thus, proficiency in using academic formulaic sequences is essential for becoming efficient academic writers, but EFL learners find it difficult to learn and use these academic formulaic sequences in their writing (Gilquin & Paquot, 2008; Peters & Pauwels, 2015).

To register themselves as insiders of the discourse community, it is incumbent on L2 learners to display familiarity with academic writing by using formulaic sequences appropriately and confidently. A lack of familiarity with the formulaic sequences used in the academic discourse of that community has serious consequences for learners (Hyland, 2008a). It not only affects their overall academic progress, but also undermines their ability to become part of their academic community (Li & Schmitt, 2009). Describing the importance of formulaic language in academic writing, Cortes (2013) notes:

> The importance and omnipresence of FS in academic writing means that mastering academic FS becomes a prerequisite for any FL learner who wants to be successful in their academic writing. FL learners should not only know how a text is organized in terms of functional units but also how these units are realized linguistically and lexically. (p.35)

This section reviews studies which have examined the use of lexical bundles in non-native students and compared them to native students. Some studies have found similarities between native English and the non-native English students in the distribution and use of lexical bundles (Chen & Baker, 2010; Shin,2019).

Chen and Baker (2010) compared native English expert writers with native English students, and with non-native English students. It is the only study that compares native expert, native and non-native students, as in the current study, so I will present this study in detail. The native English expert writers' data is based on published research articles taken from the Freiburg-Lancaster-Oslo/Bergen (FLOB) corpus. The FLOB-J used in this study is a section of this corpus that contains 80 excerpts, each 2000 words, from academic texts retrieved from published journals and book sections. For the native and non-native English student data, the corpus of British Academic written English (BAWE) was used. BAWE was released in 2008 and contains approximately 3,000 pieces (appx. 6.7 million words) of assessed student writing from British universities. BAWE-CH is the part of this corpus which contains corpora of Chinese students, whereas BAWE-EN is a comparable dataset contributed by native English students. The size of each of the three corpora in this study is 150,000 words, and the corpora cover a range of disciplines: Arts and Humanities, Life Sciences, Physical Sciences, and Social Sciences.

The results of the study showed that native English expert writers used lexical bundles differently from native and non-native students. The L1 English students used the highest number of bundle types and tokens (120 types, 757 tokens), followed very closely by the L1

English expert writers (118 types and 749 tokens). The L1 Chinese students used the least number of bundle types and tokens (90 types and 554 tokens).

The study found differences between the native expert corpus and the student corpus (native and non-native alike). The findings show that the native expert writers used far more Phrasal bundles (Noun-based, and Preposition-based bundles) than Clausal bundles (Verb-based bundles). In contrast, the native students and the non-native students used far less Phrasal bundles representing 44% of the total bundle in the native student corpus and 48% of the total bundles in the non-native student corpus. The large number of Phrasal bundles in the expert writing is in keeping with formal academic discourse which typically contains more Phrasal bundles than Clausal bundles (Biber et al., 2004). On the other hand, the greater number of Clausal bundles in the student corpus (native and non-native alike) makes their writing closer to informal speech in which the majority of the bundles are Clausal (Biber et al., 1999; 2004).

Similar findings were found in the use of bundle frames across the three corpora. The native expert writers used significantly more types and tokens of Noun-based and Prepositional-bundle frames than the native and non- native students.

Previous research (Biber et al., 1999) has shown that there are two most frequent bundle frames in academic discourse: the Noun-based bundle frame, ‘the Noun of the/a’ and the preposition-

based bundle frame, ‘in the Noun of’. In both these frames, the expert writers used significantly more types and tokens than the native and the non-native students. The expert writers in Chen and Baker (2010) used 16 bundle types and 100 tokens, whereas native English students used 9 types and 69 tokens, and non-native English students used 8 types and 37 tokens. The other frame that is most frequent in academic writing is ‘in the + noun + of the’ (Biber et al., 1999). The expert writers again used more varied bundles and more tokens in this frame as compared to student group: experts: 10 types, 87 tokens; L1 English: 3 types, 35 tokens; L1 Chinese: 3 types, 19 tokens. These features of lexical bundle usage mark the difference between formal academic writing and informal speech, as research shows that formal academic writing is characterized by variable frames of NP and PP-based bundles (Biber et al., 2004). The 68.5% of total bundle types in the expert corpus are NP-and PP-based bundles, which shows the dominance of phrasal bundles in the expert corpus. On the other hand, VP-based bundles represent more than 50% of types in the L1 English and L1 Chinese student corpora. In other words, Clausal bundles are more frequently used by both the native and the non-native students. One important finding of the Chen and Baker (2010) study is that the non-native English students did not use a single bundle in the category Noun-based bundles with other phrase fragment, e.g., *the extent to which, the degree to which*. These bundles are used for

contextualizing information in the text. The absence of these important bundles might suggest that students struggle with contextualisation.

Both groups of students used more VP lexical bundles as compared to expert writers, especially VP bundles with 'to-clause fragment'. Chen and Baker (2010) observed that the non-native English students showed special tendency to use the bundle frame *in order to* + verb frame. They used six different verbs in that frame: *achieve, avoid, be, maintain, make and understand*, while native English students used two verbs in this frame: *make and minimize*. However, in the use of passive verbs, the L1 English students used 11 types, 55 tokens, the experts used 7types, 34 tokens, whereas the Chinese students only used 4 types and 19 tokens. So, in general, the native and the non-native English students used more types and tokens of bundle frames of Clausal bundles, whereas the native English expert writers used more Phrasal bundles than Clausal bundles.

Another difference between the native English writers (expert writers and native English students alike) and the non-native English students was the use of Verb-based bundles for hedging. The native English experts and the native English students used a variety of techniques for hedging. They used the frame "Copula be + likely to" e.g., *is likely to be*, hedging nouns, e.g., *there is no evidence*, anticipatory it + adjective fragment, e.g., *it is clear that, it is*

possible to, hedging verbs, e.g., *seems to have been*, modal verbs, e.g., *would have to be, would need to be*. to mitigate a proposition.

In contrast, the non-native students only used four bundle types for hedging purposes: *are more likely to, is considered to be, it has been suggested that, it is believed that*. Their limited use of hedging devices might be linked with non-native students' general inhibition to present their evaluation in academic writing, as also shown by Hyland (2008) and Salazar (2011).

Chen and Baker (2010) also find differences in bundle functions across the three corpora. The analysis of distribution of functional categories shows that the expert writers frequently used referential expressions, with 60% of types being referential expressions that include framing bundles, quantifying bundles and place/time deictics. In contrast, L1 English students used 37% and L1 Chinese 41% of types in this category. These bundles make writing more precise, contextualized and organized. In the use of quantifying bundles, there were differences across the three corpora. The L1 English students and expert writers used some quantifying bundles (*the degree to which, the extent to which, to a large extent*) that were not used by the L1 Chinese students. The Chinese students used quantifying bundles like *in the long run, in the recent years, and all over the world.* This shows L1 Chinese students' tendency to overgeneralize and

be categorical. These types of quantifying bundles are normally used more frequently in speech than academic writing.

On the other hand, both group of students used Discourse organizers more frequently than the expert writers (L1 English: 39%, L1 Chinese: 42%, L1 English expert;21%), which include topic introduction and topic elaboration functions. The findings show that the native and the non-native English students used significantly more bundles for topic elaboration, primarily Verb-based, such as Passive verb+ Prepositional phrase fragment, e.g., *can be regarded as, be included in the* etc., Verb+ to-clause fragment, e.g., *in order to make* etc., and Subject + verb, e.g., *that is to say* etc. The overuse of discourse organizers and underuse of referential expressions makes students' writing more focused on the research topic, describing the piece and giving procedural details rather than making the text more coherent and well-structured.

Stance bundles were used differently, especially epistemic bundles used to convey certainty or uncertainty of the writers' evaluation, e.g., *is likely to be*. The L1 Chinese writers used the smallest range of bundles compared to the L1 English students and expert writers. So, in some features of bundle use, like hedging the native English students have used bundles like the native expert writers, however, in the distribution and overall use of bundles the native expert writers are different from both the native and the non-native students.

Shin (2019) also finds similarities in the use of lexical bundles between native English and non-native students. The study investigates the use of lexical bundles in argumentative essays written by undergraduate students. The native English corpus was built from writing samples produced by native English first year students in a public university in the US. The learner corpus was based on writing samples from entering course at a highly ranked university in Korea using the Criterion Online writing Evaluation Service developed by Educational Testing Service (ETS), which provides students with a holistic score (1-6) on their essays. The students had to write an essay as a part of a placement test for first year English courses. They were instructed to write an essay in response to a given writing prompt in 50 minutes in a computer lab. The following is an example of a statement given in the prompt: 'Do you agree or disagree with the following statement: Is it better to be a member of a group than to be the leader of a group? Use specific reasons and examples to support your answer.' (Shin, 2019, p.25)

The study found that the structural distribution shows that Verb-based bundles are the most common bundles in both corpora representing 65% of the total bundles in each corpus.

There were some small differences in structural characteristics of bundles found in the two corpora. The native English students used significantly more Noun-based bundles with embedded of-phrase fragment, whereas the non-native English students used significantly

more Noun-based bundles with other phrase fragments e.g., *the person who are, the reason why I.* In the category, Preposition-based bundles with other post-modifier fragment, three bundles were used in both corpora: *a lot of people, a lot of thing, the most important thing.* These bundles were shared by the native English students in this study, although one of these bundles, *the most important thing,* was used significantly more frequently by the non-native English students. Overall, the native English students used significantly more Preposition-based bundles (native 21 types, non-native 18 types), however the native students also used idiomatic bundles, e.g., *in the long run* (20 tokens), *in the real world* (18 tokens).

In the use of Verb-based bundles both groups mostly used bundles with personal pronouns, Native corpus (NC): 26, Learner corpus (LC): 32. Very often they used the first person e.g., *I think it is (NC), so I want to (LC).* However, non-native students mostly used the first-person pronoun with the word 'think' e.g., *I think that it* etc., whereas the native English students did not use this bundle.

The non-native English students used significantly more that-clause fragment bundles (NC: 2.1%, LC: 3.9%) e.g., *first reason is that, the problem is that* etc., as well as bundles with initial 'there' (existential there; NCL: 8 types (134 tokens) LC: 16 types (472 tokens)). However, they mostly used these bundles with informal quantity expressions like '*a lot of*'e.g., *there is a lot*

of and the determiner ‘many’ *there are so many*. This contradicts previous research which had showed that native English students used significantly more bundles with initial *there* than the non-native English students (Adel & Erman, 2012).

No significant difference was found between the two corpora in the use of Verb-based bundles with anticipatory it structure, and few bundles were used in the most important NP frames ‘*the__ noun + of the* : NC: 3 (86), 2 (41)’ and the PP-frame ‘*in the +noun + of* : NC: 3 (41), 2 (27)’ (Biber et al., 2004).

In terms of unique frames, the native English students used the frame ‘be able to’ e.g., *will be able to, I was able to.* The native English students used 11 types of these bundles with 238 tokens (8.6% of total tokens).

The non-native English students used bundles with the expression ‘think’ e.g., *I think it is.* This type of bundle was not used by the native English students. Biber (2010) claims that verbs like *think, know* and *say* are features of native English conversation, and the use of these verbs by non-native English students shows that they might lack awareness of the academic writing register.

In terms of functional distribution of bundles, this study found that stance expressions are the most common bundles in both corpora, representing over 45% of bundles in each corpus: NC: 70, LC: 70, followed by referential expressions with 60 types in each corpus. Discourse organisers are the least used bundles in both corpora: NC: 16 types, 10.9%; LC 15.4%, 24 types. Biber et al. (1999; 2004) showed that the majority of bundles in conversation consist of stance and discourse organisers, whereas the majority of bundles in academic discourse are referential expressions. The overall distribution of the bundles in this study shows that both native and non-native English students employ features of bundles that are characteristic of both conversation and academic discourse.

Epistemic bundles were used significantly more frequently by the non-native English students with 15 unique types, 4 of which were used with the word 'think'. These bundles are used for presenting authors' certainty or uncertainty about a proposition.

The non-native English students used more topic elaboration/clarification bundles e.g., *as because of these reasons, as I mentioned above.* Similarly, they used significantly more that-clause bundles to state their ideas e.g., *as first reason is that, the problem is that* etc. The native English students, on the other hand, used these bundles in small numbers and they used

informal topic elaboration/clarification markers such as, *with that being said* (20 tokens), *don't get me wrong* (18 tokens), neither of which were found in the non-native student corpus.

In terms of referential expressions, the native English students used 59 types, 40.4% of the total bundles, and non-native English students used 60 types, 38.5% of the total bundles. Place/time/text-deictics bundles were the most frequent bundles in both corpora: NC: 28.8%, LC: 24.8%. However, native English students used more place related bundles e.g., *in the real world, the world around us, the world we live in* etc., and non-native English students used more time related bundles e.g., *from now on I, as time goes by, as soon as possible* etc.

So, Chen and Baker (2010), and Shin (2019) found similarities in the distribution, structural characteristics and the functions of lexical bundles. But there are other studies that found several differences in the distribution and use of lexical bundles in native and non-native English students (Adel & Erman, 2012; Bychkovska & Lee, 2017).

Adel and Erman (2012) examined the use of lexical bundles in native and non-native English students. The corpus for this study was taken from Stockholm University Student English Corpus (SUSEC) and includes 325 essays, with over 1 million words. The corpus of non-native students consists of 243 texts and 863,207 words. These texts were based on the writing of the students of department of English in the Stockholm university and were collected from students

in their 1st term to 4th term. The native English students' corpus consisted of 82 texts with 247,435 words and included data from students in the 2nd and 3rd years in the department of Linguistics at King's College London. There were some important differences in the two corpora: the size and the number of texts was 4 times larger in the non-native corpus.

4-word bundles were selected to be extracted from the corpus, and the frequency criteria were set at 25/million words, with a dispersion rate of 3 for native texts, and 9 for the non-native texts. The higher dispersion rate for the non-native texts was because the number of the non-native texts was roughly 3 times more than the native texts.

The study focused on the comparison of the features of bundles shared by both corpora and of the bundles unique to each corpus. The study found that native English students used more lexical bundles, 130, as compared to non-native students, 60, and that 22% of the bundles were used in both corpora. The number of tokens was not included in the analysis in this study, which might have presented a clearer picture of the greater use of bundles in the native English corpus. There were 4 bundles that were significantly overused in the native English corpus e.g., *as a result of, at the beginning of, to look at the, can be used to.* On the other hand, 3 bundles were significantly more frequently used by the non-native writers: *as well as, the aim of this, the results from the.*

The difference in frequency of use of bundles in the two corpora might be due to differences in the content of the two corpora: the non-native texts are based on the analysis of empirical data, whereas the native English students' texts are based on the discussion of published writing. The study found four structural characteristics of bundles that are used for hedging in academic writing, exhibiting differences in native and the non-native writing, as follows:

'Unattended this', e.g., *this can be seen*

'Existential there', e.g., *there is no evidence*

'Anticipatory it', e.g., *it is likely to,*

'Passive fragment', e.g., *can be used as*

The use of 'unattended this' is found to be frequent in published academic writing. The bundles with 'this' are used for showing varying degree of certainty and for adopting a neutral tone in academic writing (Pan et.al., 2016), for example, *this can be seen, this may be because* etc. The native English students used 9 different strings of this type of bundles whereas the non-native English students did not use any of these bundles. Instead, the non-native students used bundles with the 'attended this' where this is followed by a noun, e.g., *In this study the* etc. Adel and

Erman (2012) found that the ‘attended this’ was rare in native student writing, however the non-native students used the bundles with ‘attended this’ more frequently.

Bundles with ‘there’ make academic writing appear impartial and are used for hedging (Chen & Baker, 2010). Adel and Erman (2012) found that the native students used 7 types of these bundles, whereas the non-native students used 3 types. The use of ‘there’ in this type of bundles is also known as ‘existential there’ and is used as ‘a springboard in developing the text’ (Biber et al., 1999, p.52).

Verb-based bundles with ‘anticipatory it’ are used to mitigate a proposition. The native students used 20 types of hedging with a variety of lexical verbs such as *‘seem’ ‘appear’ ‘would’ ‘could’ ‘may’*, whereas the non-native students only used four bundle types for hedging, two of them involving *‘can’* and two *‘seem’*.

Adel and Erman (2012) found that both the native and the non-native students used ‘Anticipatory it’ bundles for hedging, however, there was one feature unique to the non-native students. They used informal words like, ‘*hard’* and ‘*easy’*, in these bundles. These expressions are usually used in conversation rather than in academic writing, making non-native writing more like informal speech.

The use of passives in bundles also shows differences across the two groups of writers. 5 types of passives were shared by both groups, but overall native students used more passives (25 types) compared to non-natives who used only 8 types. The native English students used a wide variety of verbs with the passive structure, such as *see, refer to, find, say, note, attribute to, relate to, define, support, suggest, assume, and use*, whereas only 5 of these verbs were used by the non-native English students. Passive voice is considered to be highly characteristic of academic writing (Biber et al., 1999), and the relative underuse of passives by the non-native students renders their texts less academic-like.

In terms of pragmatic functions, Adel and Erman (2012) found that both native and non-native students used an almost similar proportion of referential bundles, however there were differences in terms of their use of stance bundles and discourse organisers. Adel and Erman (2012) argue that the native-English students showed better understanding of academic writing because they made more frequent and varied use of hedging devices than the non-native students, in line with previous findings (Chen & Baker, 2010). Contrary to Chen and Baker (2010) who concluded that native and non-native students were generally similar in their use of bundles, and both deviated from the norms of academic writing, Adel and Erman (2012) found differences between native and non-native English students. There are various possible

reasons for the differences in both the studies, such as the size of the corpora for example. The three corpora in Chen and Baker (2010) are similar in size, whereas in Adel and Erman (2012), the native students' corpus is much smaller than the non-native student corpus. A smaller corpus generates more bundle types when compared with a bigger corpus, following the same frequency criteria, which might have led to those differences. Additionally, the native and the non-native student corpora in Chen and Baker (2010) are multidisciplinary, whereas both corpora in Adel and Erman (2012) are discipline specific. These differences might have given rise to the conflicting findings in both studies.

Bychkovska and Lee (2017) also found significant differences between the native English and the non-native English students in their use of lexical bundles. Their study compares the use of lexical bundles in native English and non-native English undergraduate student argumentative essays. Their corpus consists of assessed argumentative essays written by US-based native English and Chinese ESL undergraduate students: the Michigan Corpus of upper-level student papers (MICUSP), an approximately 2.6-million-word corpus of various high rated (A-graded) academic papers produced by native English and non-native English students. The native English students' data consisted of 101 texts (220,233 words), mostly from humanities and social sciences. The non-native English students' data is taken from the Corpus of Learner and

Teacher English (COLTE) from Ohio University, consisting of 105 texts and 105,043 words. The COLTE corpus contains argumentative essays written by Chinese ESL students. The essays are on general themes e.g., education, economy, environment, health etc. between 900-1200 words. 4-word bundles were selected for analysis in this study. The frequency criterion was set at 40/million words, with dispersion rate at 5 texts, and the concordance tools AntConc (Anthony, 2014) were used.

The findings show that the native English students used 23 types and 337 tokens, whereas non-native English students used significantly more types and tokens (52 types and 404 tokens). The non-native students used bundles more frequently, which could be linked to a less formal register, as claimed by previous research which has shown that lexical bundles are more frequent in conversation than in academic writing (Biber et.al, 1999; 2004).

In terms of structural characteristics, the native English students used significantly more Phrasal bundles (78.3% types, 77.1% tokens) than the non-native English students (51% types and 57.3% tokens), as is characteristic of academic prose (Biber et al., 1999; Chen & Baker, 2010; Hyland, 2008a). The native English students used significantly more PP-bundles with of-phrase fragment, whereas the non-native English students used significantly more PP-bundles with other post-modifier fragments.

In terms of functional distribution, referential expression bundles are the most frequent in both corpora: MICUSP (types 69.6%, tokens 68.3%), COLTE (types 58.5%, tokens 59.5%). In all the sub-categories of Referential expression bundles, the non-native English students used significantly more bundles, except for framing bundles. Framing bundles are important part of academic discourse as they are used for organizing the text, and previous research has shown that non-native students use framing signals significantly less frequently than native English students (Chen & Baker, 2010; Shin, 2019). Non-native English students used significantly more Quantifying bundles (types and tokens) than native students, however, their bundles tended to contain vague words like e.g. people, *people who do not, people do not have*,; more and more e.g., *more and more people, nowadays more and more*; a lot of e.g., *a lot of people.* This feature of using informal and vague quantifying bundles has also been reported in previous research (Chen & Baker, 2010; Shin, 2019).

In their use of discourse organizing bundles, the non-native students used more topic elaboration bundles than the native students, who produced no topic introduction bundles. Non-native students used bundles that are mostly used in speech e.g. *is a good choice, a huge amount of, a lot of time* etc. Bychkovska and Lee's (2017) study shows that the use of lexical bundles in non-native student corpus is characteristic of informal and conversational speech rather than

academic writing (Biber et al., 1999; Cortes, 2004; Hyland, 2008a; 2008b). The non-native student writing tends to be clausal in nature and they used significantly less bundles for contextualizing new information. In contrast, the native English student writing is predominantly phrasal in nature, and they use significantly more bundles for organizing the text, especially for contextualizing new information.

They also showed that the Chinese students made grammatical mistakes in their use of bundles, especially in the use of articles and prepositions (over 50%). For example, mistakes in the use of articles included missing articles (*on other hand*), misplaced articles (*the one of most*), misused articles (*the large amount of*), missing prepositions (*according the articles*), and misused prepositions (*in the same time*). The authors suggest that the influence of Chinese language, that is articles less language, might be a possible reason for the misuse of article in L1 Chinese students. Overall, the Chinese students misused 95 lexical bundles, 85 of which were grammatical mistakes and 10 functional mistakes.

The findings of this study are different from Chen and Baker (2010) who found that both native and non-native students used more clausal bundles than phrasal bundles, and more discourse organizers than Referential expression. Chen and Baker (2010) also found that the non-native students used a smaller number of Stance bundles. There can be many possible reasons for the

difference in findings in the two studies. One possible reason might be difference in the corpora used. Chen and Baker's (2010) native and non-native student corpora are based on student assignments, whereas Bychkovska and Lee's (2017) native and non-native student corpora are based on the argumentative essays. The different sizes of the corpora might be another possible source of differences. In Chen and Baker (2010), the size of the native and the non-native student corpora is similar, however, in Bychkovska and Lee (2017), the native student corpus is double the size of the non-native student corpus. This might be one of the reasons that more types and tokens were found in the non-native student corpus because the smaller corpus is more likely to contain more types and tokens when similar normalized frequencies are set for extracting bundles.

2.4.3 Use of lexical bundles in native and non-native English expert writers

Pan et al. (2016) examined the frequency, structural characteristics and functions of lexical bundles in published articles in telecommunication journals. The study compares research articles published between 2007-2014 by native English writers (TELE-EN) and by non-native Chinese (TELE-CH) writers. The native English journals were selected on the basis of their high impact factor, and the TELE-CH journals were published by top universities in China. The frequency criteria were set as 40/million words with 5 texts dispersion for the TELE-EN

and 10 texts for TELE-CH as the number of texts was higher in the TELE-CH corpus. The corpus tools Wordsmith 4.0 were used for the analysis.

This study revealed important differences in the frequency, structures, and functions of lexical bundles used by native and non-native expert writers. In terms of frequency, the non-native expert writers used a higher number of bundle types, 71, as compared to native expert writers who used 53 types. Out of these total 124 types extracted, 24 bundle types were used in both corpora. The non-native expert writers also used bundle tokens more frequently than the native expert writers.

In terms of structural distribution, the native English expert writers have used significantly more phrasal bundles (69% types and 67% tokens) than the non-native expert writers. On the other hand, non-native expert writers used significantly more clausal bundles (VP-based bundles 58% types, 56% tokens) than the native expert writers. This made the non-native expert texts less academic sounding, because previous research has shown that phrasal bundles are more prevalent in academic writing, and more clausal bundles are used in speech (Biber et al., 1999). Phrasal bundles, consisting of Noun-based and Preposition-based bundles, are known for their high information focus. They are used for contextualizing information, establishing

coherence, referring to the results, and are therefore frequently used in published academic writing (Biber et al., 2004; Hyland, 2008a).

In terms of Noun-based and Preposition-based bundle frames, i.e., 'the Noun of the/a', and 'in the Noun of', Pan et al. (2016) found no significant difference between the native and the non-native expert writers. In the Prepositional bundle frame 'in Noun of the/a' the native expert writers used 3 types and 128 tokens whereas the non-native expert writers used 2 types and 68 tokens, a non-significant difference. This indicates that the non-native expert writers have mastered these important bundle frames typical of academic writing.

In their use of Verb-based bundles, the non-native expert writers used significantly more bundle types and tokens of passive verb + Prepositional phrase fragment than native experts. They displayed a strong preference for the Passive verb frame 'can be + passive verb + complement' (e.g., *can be expressed as*), using six different verbs (describe, divided, obtain, express, use, write) in this frame. The passive verb is an important part of academic discourse as it presents the authors' evaluation in a more mitigating way and helps engaging with the readers (Chen & Baker, 2010). So, the non-native experts seem to have mastered its use.

In terms of functional distribution, both groups used Text-oriented bundles the most frequently, representing 49 % types and tokens of the total bundles in the native corpus, and 45% types

and 49% tokens of the total bundles in the non-native corpus. The non-native experts used significantly more bundle tokens in all the sub-categories of Text-oriented bundles except framing bundles, but both groups were similar in their use of Text-oriented bundle types. However, structuring signals used to introduce the organization of the paper were used frequently by native expert writers, e.g., in *the next section, in the previous section, in this section we*, but not by the non-native expert writers. This indicates that the native expert writers' texts were more reader friendly as they used these bundles for signposting.

The native expert writers used a variety of framing bundles (*in the case of, in terms of the, in the context of*), whereas the non-native expert writers only used a limited range (*in the case of, with respect to the*). The bundle, *in the context of,* was particularly neglected by the non-native expert writers.

The non-native expert writers used Research-oriented bundles more frequently in all the sub-categories except Quantification bundles for which the native expert writers used more types and significantly more tokens than the non-native expert writers. The native expert writers used bundles like *a large number of, a wide range of,* whereas the non-native students used only Quantification bundles with the noun '*number*', e.g., *is the number of, the average number of.*

Participant-oriented stance features were used significantly more frequently by non-native than native English. The non-native expert writers used a variety of stance bundles, such as certainty bundles used for showing strong commitments and assertation, e.g., *it is obvious that, it is clear that*, without necessarily providing sound evidence. They also used evaluative bundles (e.g., *it is difficult to, it is easy to*) to present their evaluation in a proposition. However, these types of bundles weaken the writer's credibility as they contain subjective adjectives (Pan et al., 2016). The non-native experts used some tentative bundles as well, e.g., *is assumed to be, we assume that the*.

In terms of participant-oriented bundles, the non-native writers have more and a wider range of stance bundles than the native writers. This finding is different from previous studies (Chen & Baker, 2010; Cortes, 2004; Salazar, 2011) that showed that non-native writers used a small number of participant-oriented bundles, especially stance bundles. This makes the non-native writing less reader friendly and less neutral (Adel & Erman, 2012).

Phrasal bundles, that are used to present information with high information focus, are used significantly more by the native expert writers, such as framing signals are used for contextualizing new information. These bundles make native expert writing more cohesive and reader friendly (Pan et.al, 2016). On the other hand, the non-native expert writers used

significantly more clausal bundles. In their use of hedging also, the non-native expert writers use more types and tokens than the native expert writers, however they used some evaluative bundles, e.g., *it is difficult to, it is easy to,* and certainty bundles, e.g., *it is obvious that, it is clear that* etc. These types of bundles affect the writers' credibility because they contain subjective adjectives. These are some of the differences that make non-native expert writing less effective and different from the formal academic discourse that is the hallmark of native expert writing (Adel & Erman, 2012; Chen & Baker, 2010; Pan et al., 2016).

In this section, I have reviewed the studies on the use of lexical bundles in different registers; in native and non-native student writing; and in the native and non-native expert writing. As the current study is based on the use of lexical bundles in Pakistani context, it is important to review studies on the use of lexical bundles in Pakistani postgraduate students' academic writing. Hence, in the following section, I will review the studies on the use of lexical bundles in Pakistani postgraduate students' academic writing.

2.4.4 Use of lexical bundles in the academic writing of Pakistani students

Pakistani postgraduate students have been shown to rely heavily on lexical bundles (Fazal et. al., 2019; Yousaf and Shehzad, 2018). Fazal et al. (2019) studied the use of lexical bundles in PhD theses in five different social sciences disciplines (Education, English, History, Political

Sciences, Psychology) written by native English and non-native Pakistani postgraduate students. For the non-native students, 100 PhD theses were collected from Pakistan's Research Repository maintained by Higher Education Commission of Pakistan (HEC) 6,350,130 words in total. The native students' data comprised 100 PhD theses from British library, from the same disciplines (13,026,919 words). The study compared the discourse functions and the distribution of lexical bundles in native and non-native students' academic writing. 4-word bundles were selected for the analysis with AntConc 3.3.5. Biber et al.'s (2004) functional categorization was used which divides the bundles into the following three functions: Stance, Referential, Discourse.

The results of the study revealed that the Non-native writers use more bundle types (558) made more varied use of bundles than the native writers (327 types). Moreover, the comparison of the top 20 bundles between the two corpora showed that the non-native Pakistani students used bundles more frequently in all the categories. In both the corpora, the Referential function is the most used function, representing 75% bundles (246) in the native student corpus and 64% (358) in the non-native student corpus. Stance function is the second highest,20% (65) in native and 22% (124) in non-native students. Discourse bundles comprise only 5% (16) in the native and 13% (76) in the non-native corpus.

The analysis of the sub-categories showed that in both corpora, Framing bundles were the most frequent bundles, however the non-native students used three times more 35% of the total bundles, than the native students (13% of the total bundles). Similarly, Epistemic stance bundles were more frequent in the non-native corpus (10% of the total bundles) than in the native corpus (2%). So, although Pakistani students made more varied and more frequent use of lexical bundles, their distribution in terms of discourse functions were somewhat similar to the native students. Both the native and the non-native students clearly preferred using the bundles for Referential functions, however, in the use of Discourse bundles the non-native students used more than double the bundles used by the native students.

Yousaf and Shehzad (2018) also found out that Pakistani postgraduate students rely on lexical bundles in their writing. The data of the study comprised PhD theses from 9 disciplines: English studies: Linguistics, Literature, ELT; Social sciences: Political science, Education, Psychology; Biological sciences: Biotechnology, Botanical sciences, Zoological science. The PhD theses were taken from Pakistan's Research Repository. The total size of the corpus was 4.7 million words. In order to extract 4-word bundles the frequency criteria was set at frequency/million words: 10, Range: 5. AntConc 3.4.4 was used for the quantitative analysis.

Biber et al. (1999) Structural taxonomy was used to categorize the structural characteristics of the lexical bundles.

The results also showed that Pakistani students rely heavily on the use of lexical bundles, and that the structural distribution of the use of bundles varies across different disciplines. For example, in Linguistics and Literature, PP-based bundles were the most frequent whereas NP-based bundles were the second most frequent. In ELT, NP-based bundles were the most frequent, whereas PP-based bundles were the second most frequent.

More variation in the structural characteristics of bundles across overarching fields (English studies, Social sciences, Biological sciences) was found than was found within different sub-disciplines. English studies and Social sciences were found to be predominantly PP-based bundles whereas in Biological sciences VP-based based bundles dominated. These results suggests that the Pakistani students from the English studies and the Social sciences have more frequently used lexical bundles that are characteristic of academic writing than the students from the biological sciences.

2.4.5 Conclusion: Lexical bundles in academic writing

Previous research on use of lexical bundles in academic writing can be summarized as follows:

- Lexical bundles used in speech and academic writing are different in structure and discourse functions. The majority of the bundles used in academic discourse are NP+PP bundles and referential bundles, whereas more VP based bundles and more stance bundles are used in conversation. Similarly, bundles used in conversation are primarily clausal, whereas more phrasal bundles are used in academic writing.

- Lexical bundles are discipline specific and genre specific; therefore, the role of discipline specific bundles is important in academic writing.

- Some studies (Chen & Baker, 2010; Shin, 2019) have found similarities in the distribution and use of lexical bundles in native and non-native English student writing. They showed that both native and non-native students used more clausal bundles than phrasal bundles, and that the distribution and use of bundle functions was also similar. Both native and non-native students used significantly fewer framing and stance bundles than referential bundles. Other studies (Adel & Erman, 2012; Bychkovksa & Lee, 2017) found differences in distribution and use of lexical bundles in native and non-native students. For example, these studies found that non-native students used bundles more frequently than native students, and that clausal bundles were the most frequent in non-native writing, whereas phrasal bundles were the most frequent in

native student writing. On the basis of these findings, these two studies found native student writing to be more organized and mature than non-native student writing.

- Previous studies (Adel & Erman, 2012; Bychkovksa & Lee, 2017; Chen & Baker, 2010; Hyland, 2008a) found that non-native English student and non-native experts do not use appropriate bundles for hedging. In addition, non-native students used evaluative bundles that are used for emphasis and making forceful statements. The same trend was found in the writing of non-native expert writers (Pan et.al., 2016).

- Previous studies (Adel & Erman, 2012; Bychkovksa & Lee, 2017; Chen & Baker, 2010; Hyland, 2008a) show that non-native students and non-native expert writers use significantly fewer and less varied framing bundles, used for organizing text and new information, than native students and native expert writers. These bundles are important for textual cohesion.

- Studies on the use of lexical bundles by Pakistani postgraduate students found that Pakistani students rely heavily on lexical bundles (Fazal et al., 2019; Yousaf & Shahzad, 2018). Fazal et al. (2019) found that the non-native English Pakistan students used epistemic stance bundles 5 times more frequently and Discourse bundles 2 times

more frequently than the native English students. Yousaf and Shahzad (2018) studied the structural characteristics of lexical bundles and found that the students from English studies and Social Sciences used PP-based bundles more variedly and more frequently, whereas the students of biological sciences used VP-based bundles more variedly and frequently.

The research on native and non-native English students has therefore provided diverging results, and more research needs to be carried out in this area to find more evidence of the ways different populations use lexical bundles (expert/student/native/non-native). This study aims to contribute to this debate.

The review of previous literature on the use of lexical bundles leads us to the implications of this type of corpus research on English language learning and teaching. The corpus research has brought new insights in the field of English language learning and teaching (Romer, 2011). In the following section, I will discuss the implications of corpus research for English language learning and teaching. I will discuss the role of corpus research in three important areas of English language teaching: Learner dictionaries, Learner corpora, and Data driven learning.

2.5 Implications of corpus research for ELT

Corpus research has implications for English language learning and teaching. Romer (2011, p.205) notes, 'over the past few decades, corpora, corpus, corpus tools, and corpus evidence have not only revolutionized linguistic research but have also had an impact on second language learning and teaching.' This impact has been in ELT areas like syllabus design, teaching materials, data driven learning (DDL) in L2 classroom through corpora, learner dictionaries, textbooks etc. (Anthony, 2016; Granger, 2002; Huang, 2017; Romer, 2011). The pedagogical applications of corpus research can be classified into direct applications and indirect applications (Romer, 2011). The indirect applications are related to the decision-making process that involves the use of corpus research for syllabus designing, or the content of the materials is selected based on the results of corpus research. The direct applications of learner corpus are related to direct involvement of the students in the language learning process using corpus. It can be the use of already available learner corpora for teaching various lexical and discoursal features to English language learners (Anthony, 2016). It can also be the use of leaners' own individual corpus through which they learn language and discourse functions (Charles, 2014). In this section, the implications of the corpus research will be discussed in three areas: learner dictionaries, learner corpora, and data driven learning.

2.5.1 Learner dictionaries

One of the most important contributions of corpus-based research has been the production of learner dictionaries (Granger, 2002; Huang, 2017). Learner dictionaries are important as they are based on the frequency of words. These dictionaries provide information related to frequency, collocations, grammar, usage guides, concordance samples. These dictionaries provide authentic information to the teachers and the learners. One of the most important aspects of these dictionaries is that they are based on evidence-based findings and provide genuine instead of invented examples (Romer, 2011). Some of the most important learner dictionaries are the Longman Grammar of Spoken and Written English, Longman Dictionary of Common Errors, Collins COBUILD Advanced Learner's English Dictionary, Collins COBUILD Intermediate English Grammar and Practice, The Cambridge Advanced Learner's Dictionary, Longman Dictionary of Contemporary English, Oxford Advanced Learner's Dictionary, Oxford Collocations Dictionary for Students of English, Oxford's Practical English Usage, and Macmillan English Dictionary (Huang, 2015). Although these learner dictionaries have played a very important role in various aspects of language learning and teaching, the research has shown these dictionaries still lack in coverage and accessibility of lexical bundles (Chen & Zhao, 2022). Therefore, it is important that the multiword units like lexical bundles

are presented more prominently in learner dictionaries. For example, lexical bundles that are used more frequently in academic writing can be presented as headwords in learner dictionaries (Granger & Lefer, 2016). This will help the learners in learning the use of lexical bundles in academic writing. The next section deals with the importance of learner corpora in English language teaching.

2.5.2 Learner corpora

The use of learner corpora has been significant in ELT, for example for understanding the nature of second language learning (Myles, 2005). Learner corpora have not only helped in documenting developmental patterns in second language learning, but also in understanding the effect of the first language on second or foreign language learning (Granger, 2002). They have enabled the comparison of native and non-native language production, to document deviations, underuse and overuse of linguistic items in learner productions (Granger, 2002; Huang, 2017). Huang (2017, p.386) notes that 'The insights that learner corpora can provide through analysing the language production of certain groups of leaners from particular language backgrounds with respect to the difficulties they face are invaluable in language teaching and learning.'

Some of the largest important leaner corpora are Longman Learners' Corpus (10 million words of written language), the Cambridge International Corpus (20 million words from Cambridge exam scripts written by learners of English), the International Corpus of Learner English (3 million words in the form of written essays by learners with various first languages) (Huang, 2015). With regard to the teaching of lexical bundles, the leaner corpora have enabled the researchers to bring new insights into ELT. The research on leaner corpora has highlighted the important differences between the native and the non-native students and expert writers. For example, it has been shown that the leaners use some bundles more frequently, whereas some of the lexical bundles have been used only by the expert writers and native English students. These types of important insights from learner corpora have implications for English language teachers and syllabus designers. The second language learning should be based on the findings of these types of research. It also emphasizes the importance of using corpora inside the classroom. The method that is named as Data Driven Learning (DDL). The following section will explain the importance of DDL for English language teaching.

2.5.3 Data Driven Learning (DDL)

'Data-driven learning (DDL) can be defined in broad terms as any use of a language corpus (i.e., a representative sample of target language) by second or foreign language users' (Boulton,

2012, p.263). Anthony (2016) notes that DDL is a learner centred approach in which the learners play an active role. He further explains the learners' role as follows:

> [Learners] analyse the corpus with software tools, and through the observation of bottom-up, lexico-grammatical features and top-down rhetorical or discourse features in the corpus, they identify patterns, deduce rules and form hypotheses about the target language, which they can then apply in future receptive or productive tasks.' (p.163)

The use of corpus tools like concordance, KWIC (Key Word In Context), File view are important in using different corpora in the classroom to the native and non-native students of English language.

The corpus-based research has given importance to the most frequent words (i.e., lexical bundles), in contrast to the learning and teaching of the infrequent words. The corpus tools have made it possible to identify the most frequent words in different corpora. By generating the list of most frequent words, the teachers can make well informed decisions to focus on the most frequent words (Huang, 2017). Leaners can be asked in the classroom to use the concordance and key-word-in-context (KWIC) tools to explore the meaning and usage of the most frequent words. The same tools can be used for learning the use of sentences in context.

The learner corpus can also be used for teaching discourse features, like the use of genre-specific vocabulary, linguistic markers used for presenting writers' stance, and the use of context specific vocabulary. For example, the COCA corpus which contains samples from different genres, e.g., speech, fiction, magazine, newspapers, academic discourse etc. The COCA corpus interface provides information about the frequency profile of vocabulary items, parts of speech etc., across different genres. By gaining this information, students can learn the use of different vocabulary items in their relevant discourse. Through this practice students can learn to understand the discourse specific vocabulary and grammatical patterns (Anthony, 2016)

The corpus can be used for identifying different linguistic markers in a discourse, e.g., model verbs that are used for presenting writers' stance. For example, students might be asked to identify different linguistic markers like *might, would, may be, possible,* etc. to see how these linguistic markers are used by the writers to make the claim stronger or weaker in text. The context of communication is another aspect of discourse that can be taught through learner corpus. The corpus like MICASE (Michigan Corpus of Academic Spoken English) can be used for this purpose by using different contexts or variables, e.g., type of speaker (teacher/ student), first language background, type of speech event (seminar/defence) (Huang, 2017).

2.6 Conclusion and Research Questions

This chapter has reviewed studies exploring the use of lexical bundles in academic writing, including studies in Pakistani context. It has focused more centrally on reviewing similarities and differences found in lexical bundle use in expert writers (both native and non-native) and in student writers (native and non-native), which is the focus of this study. At the end, I have also presented some pedagogical implications of corpus research in the field of English language learning and teaching. This has led me to setting the following research questions for my investigation of the use of lexical bundles in my study.

1 (a) What are the most frequent bundle structural categories in the expert writers' corpus, the native students' corpus, and the non-native students' corpus?

(b) How does frequency of structural categories compare across the three corpora?

2 (a) What are the most frequent bundle functional categories in the expert writers' corpus, the native students' corpus, and the non-native students' corpus?

(b) How does frequency of functional categories compare across the three corpora?

Chapter 3

Methodology

3.1 Introduction

This chapter will explain the process of data collection for each corpus (section 3.2). As published research articles and student dissertations make the corpus of this study, I will highlight and discuss the differences between published research articles and student dissertations in this section. This section will also focus on issues of comparability across the three corpora, and the process used for preparing the three corpora. Sections 3.3, 3.4, and 3.5 will discuss the selection of the three criteria, frequency, dispersion, and size of lexical bundles, used for the extracting lexical bundles. Section 3.6 will present the procedure followed for refining the extracted lexical bundles. Section 3.7 will present the software tools of AntConc used in the present study for corpus analysis. Section 3.8 will briefly explain the statistical methods used in the study. Finally, I will describe the details and process followed for the qualitative analysis in this study.

3.2 Selection of the Corpus data: research articles vs dissertations

Academic writing tends to be highly conventionalized and formulaicity plays an important role in it (Hyland, 2008; Simpson-Vlach & Ellis, 2010). Published research articles can be seen as a model for this genre. Masters students can be considered as being in training in writing in this genre. It therefore makes it interesting to compare student and expert writers, to assess how far they have acquired the specificity of this genre.

The two genres, research articles and student dissertations were selected because the comparison between the experts' academic writing with students' academic writing can be enlightening in their use of lexical bundles. Both genres are also comparable in many ways. For example, both genres are considered formal and aim to follow the norms of academic writing. Published research articles are one of the most important sources of knowledge. Describing the centrality of research articles as a genre, Hyland (2009) notes the following:

> Beginning life in the form of the letters published in The Philosophical Transactions of the Royal Society in the mid-seventeenth century, the RA is now not only the principal site of disciplinary knowledge-making but, as Montgomery (1996) has it, 'the master narrative of our time. (p.67)

Hyland (2009) enlists three main features of research articles:

Review and revision

The process of review and revision is a challenging task for the novice academic writers, native English and non-native English alike. It is used for ensuring quality and also as a mechanism for community control by the process of peer review and editorial revisions.

Novelty and relevance

These are the two important elements that are highlighted by the writers in research articles. For example, the writers in hard sciences tend to emphasise the novelty and benefit of their work, whereas in social sciences the importance of the contribution is highlighted by the writers. Following the general structure of the research articles and the placement of the work through its relevance to the existing literature is displayed by the researchers.

Stance and engagement

Taking a personal stance while presenting a claim and engagement with the potential audience are the two essential elements of writing in research articles. The stance includes the use of hedging devices, e.g., *We propose several <u>possible</u> reasons for this:* boosters, e.g., *On the contrary, the role of contingencies <u>should be</u> stressed*. attitude markers, e.g., *<u>It is interesting</u>*

right off the bat to notice that . . . self-mention, e.g., *We also asked them about their attitudes toward writing at work.*

The engagement features include the use of Reader pronouns, e.g., directives, e.g., refer to table, personal asides, e.g., And – as I believe many TESOL professionals will readily acknowledge – critical thinking has now begun to make its mark, particularly in the area of L2 composition. appeals to shared knowledge, e.g., It is, of course, possible to realize capacitors using the inter-metal, linearmetal-poly, metal-diffusion,, questions, e.g., Is it, in fact, necessary to choose between nurture and nature? So, these are the important elements of research articles. The student dissertation is also an important genre that has its own features. Describing the importance of student dissertations, Hyland (2008b, p.47) discusses the challenge faced by master's students in dissertation writing and observes that 'The problem for master's students is to demonstrate a suitable degree of intellectual autonomy while recognising readers' greater experience and knowledge of the field.

Andrews (2007, p.13) describes the following essential features of a student dissertation: 'scholarship, independent critical thought, an original contribution to knowledge, argumentative coherence, conventions of presentation.' So, both the research article and student dissertations are similar in aiming to embody scholarship, and in the organization and

structure of their writing, as well as in adhering to discourse conventions. However, there are some differences between the two genres. For example, the size of dissertations is generally much larger than the research articles. The other difference between the article and dissertation is that the articles are written for the creation of knowledge while the dissertations are written for gaining training in the subject. So, articles demonstrate expertise in the field whereas dissertations are a type of apprenticeship.

Students write dissertations for passing their degrees, imitating published research in so doing; therefore, we have used them for comparison in this study. Moreover, the research papers not only function as a reference text of research work for student dissertations, but students also learn the conventions of academic writing through these research papers. Hence, the expert writers' corpus has been selected as a reference corpus in this study to see how expert writers use lexical bundles in research papers and how students' use of lexical bundles is different from them. Three corpora were used in this study. The detail of these corpora is given in the following sections.

3.2.1 Non-native students' Corpus

The data for the non-native student's corpus was collected from Applied Linguistics MPhil dissertations of Pakistani students from 4 different universities in Pakistan. These 4 universities

are located in 4 different cities of Pakistan and have degree awarding status recognised by the Higher Education Commission of Pakistan. The data was collected through emails to faculty in the departments of Language and Linguistics of four universities in Pakistan. The size of non-native students' corpus is 502945 words. There are 19 dissertations, with an average size of 26490 words.

3.2.1.1 Distribution of sub disciplines in non-native students' corpus

The distribution of sub-disciplines in the dissertations of non-native students is presented in Table 3.1.

Table 3.1 Distribution of sub-disciplines in the non-native students' corpus

Sub discipline	No. of texts	No. of words	% in Corpus
Second language acquisition	7	167,496	33%
Discourse Analysis	4	118,145	24%
English Language Teaching	4	104,933	21%
Critical Discourse Analysis	3	80,555	16%
Sociolinguistics	1	31,816	6%
Total: 5	19	502945	100%

3.2.2 Native students' corpus

Data collection for native students' corpus was not straightforward, especially because of differences in Education standards of Pakistan and the UK. Therefore, for the non-native student corpus, MPhil students in Applied Linguistics were chosen while for the native students' corpus, I collected data from the Masters' students (MA) in Applied Linguistics. This makes the students' data comparable at the level of degree qualification.

The native students' corpus consists of 312,981 words in total and 20 MA dissertations, with the average size of each dissertation 15,649 words. These dissertations were collected from10 research intensive universities in the UK. These universities ranked among the top 330 universities according to QS ranking 2017. To collect dissertations for this corpus, the following sources were used:

Twitter

Facebook

The Linguist List

Email

The detail of the sub-disciplines in native students' corpus is given in Table 3.2.

Table 3.2 Distribution of the sub-disciplines in the native students' corpus

Sub discipline	No. of texts	No. of words	%in Corpus
Second language acquisition	7	117,260	38%
Sociolinguistics	4	57,473	18%
psycholinguistics	3	47,449	15%
Discourse Analysis	2	33,404	11%
Critical Discourse Analysis	1	19,690	6%
English Language Teaching	1	16,679	5%
Phonology	1	15,995	5%
Literacy Studies	1	5,031	2%
Total: 8	20	312981	100%

3.2.3 Expert writers' Corpus

The expert writers selected for this study, are not necessarily native writers, but the research publication in highly ranked journals ensures strict adherence to English academic writing conventions. The expert writers' corpus is based on research articles published in 6 journals of

Applied Linguistics, as shown in Table 3.3. The expert writers' corpus will be used as a reference corpus in the study because native and non-native students can refer to experts' writings for research purposes; and also, to learn the features of academic writing, e.g., vocabulary, sentence structure, discourse function etc., adopted by expert writers.

The expert writers' corpus contains 510,829 words and 68 texts, with the average size of 7,194 words. The detail of sub disciplines and the number of articles included in the expert writers' corpus is given in Table 3.3.

Table 3.3 List of Journals for Expert writers' corpus

Name of Journal	No. of research articles
Studies in Second Language Acquisition	25
Language teaching research Quarterly	16
Discourse Studies	9
Journal of Sociolinguistics	7
Applied Linguistics	7
Critical Discourse Studies	7
Total: 6	71

The distribution of the expert writers' corpus in terms of sub-disciplines is based on the combined distribution of sub-disciplines in the native and non-native students' corpora. Although all writings in all three corpora belonged to the discipline of applied linguistics, it is recognised that there might be variations in academic discourse in the sub-disciplines of applied linguistics. Therefore, it was considered important to match as far as possible the sub-disciplines across the three corpora.

For this purpose, the median of the accumulative proportion (e.g., the proportion of Sociolinguistics in native corpus was 11% while 24 % in non-native students' corpus; the accumulative proportion would be 24+11=36, and the median of 36 is 18. The proportion of Sociolinguistics in expert corpus would be 18) of native and non-native student corpus was adopted to set a criterion for the selection of sub disciplines in the expert writers' corpus. The highest proportion, 35%, of the expert writers' corpus consists of the sub- discipline second language acquisition. The distribution of sub disciplines in the expert writers' corpus is presented in Table 3.4.

Table 3.4 Distribution of sub disciplines in Expert writers' corpus

Sub disciplines	No. of texts	No. of words	% in corpus

Second language acquisition	25	177500	35%
Discourse Analysis	9	91329	18%
English Language Teaching	16	64401	13%
Sociolinguistics	7	63933	13%
Critical Discourse Analysis	7	58465	11%
Psycholinguistics	4	35027	7%
Phonology	2	13213	2%
Literacy Studies	1	6961	1%
Total: 8	71	510829	100%

3.2.4 Comparability of the three corpora

As mentioned earlier, for comparability in terms of sub disciplines, I combined the proportion of each sub discipline in native and non-native students' corpora and calculated the median of the accumulative percentage of both the corpora. The median number was adopted as the criteria for proportion of sub discipline in the expert writers' corpus. The comparison of the proportions of each sub disciplines across the three corpora is given in Table 3.5:

Table 3.5 Distribution of sub disciplines across the three corpora

Sub disciplines	**native students**	**non-native students'**	**Expert writers**
Second language acquisition	33%	37 %	35%
Sociolinguistics	24%	11%	18%
English Language Teaching	21%	5%	13
Sociolinguistics	6%	18%	12%
Critical Discourse Analysis	16%	6%	11%
Psycholinguistics	0%	15%	7.5%
Phonology	0%	5%	2.5%
Literacy Studies	0%	2%	1%
Total: 8	100%	100%	100%

Table 3.5 shows the details of the percentage of sub discipline across the three corpora.

In order to minimize the impact of some unavoidable differences between the corpora, I took some measures. For example, I set dynamic raw frequency and dispersion criteria to deal with the impact of size differences, and I considered the combined proportion of sub disciplines in

native and non-native students' corpus for the selection of sub disciplines in the expert writers' corpus.

3.2.5 Preparing the corpora

The preparation of the expert corpus included the process of downloading the online versions of research articles from the selected journals. The selected pdf files were converted into plain text files. For the smooth processing of the files in the corpus software, the files were cleaned of references, graphs, charts, headers, footers, captions, and appendices. The native and non-native student dissertations were collected through emails in pdf files. These files were converted into plain text files. These files were cleaned of title pages, acknowledgements, dissertation outlines, lists of figures, list of tables, references, graphs, charts, headers, footers, captions, and appendices.

In the next section, I will describe the procedures followed for generating a final list of lexical bundles from the three corpora. Before the corpus analysis, there are three important criteria that need to be set: frequency, dispersion, and size of the lexical bundles. These three criteria define the lexical bundles that will be extracted and compared across the three corpora.

3.3 Setting the frequency threshold

In this section I will discuss the two frequency criteria, normalized and dynamic frequency, used in the literature for extracting lexical bundles. The frequency criterion is an important criterion for extracting lexical bundles, because this criterion defines the lexical bundles in a corpus. Hence, any lexical bundle that does not fulfil the set frequency criterion will not be considered a lexical bundle. In most of the studies on lexical bundles, the normalized frequency criterion has been set between 20-40 lexical bundles/one million words; therefore, this range, 20-40, is known as the standard frequency criterion for extracting lexical bundles. However, this criterion is arbitrary, and some studies have used much higher frequency criterion, especially for comparing different size corpora. Setting a frequency criterion while comparing different size corpora can be difficult to set for different size corpora because different size corpora can generate different number of lexical bundle types and tokens, based on a normalized frequency criterion. The large size corpora will normally have more types and tokens of lexical bundles as compared to small corpus because large size corpora will have more words, so it will generate more lexical bundles. Moreover, if we set a normalized frequency criterion, the converted raw frequency, used for extracting lexical bundles, will be lower for the small size corpus. As a result of low raw frequency, it will be easier for many

lexical bundle types to meet the frequency criterion; therefore, we will get relatively more lexical bundle types from the small size corpora as compared to large size corpora. In reality, these types would not be truly representative of that corpus because they were extracted due to low raw frequency. To deal with this problem, some studies have adopted a dynamic frequency criterion for comparing the use of lexical bundles in different size corpora (Chen & Baker, 2010; 2016). Taking a dynamic frequency criterion means setting an appropriate criterion, that might not be standard and identical for different size corpora. In the coming sections, we will discuss these two frequency criteria: normalized and dynamic, along with their advantages and disadvantages.

3.3.1 Normalized frequency threshold

Setting a Normalized frequency threshold is to take a normalized frequency threshold between (20-40 lexical bundles/one million words, as used in most of the studies on lexical bundles) and to convert it into a raw frequency as per size of the corpora. The underlying principle in this approach is to adopt an identical normalized frequency criterion (e.g., 20/30/40 lexical bundles per one million words) for each corpus, so that we may set an equally proportional frequency criterion for each corpus. However, setting a normalized frequency criterion does not work when we compare different size corpora. There are two reasons for this. First, even if

we set an identical normalized frequency criterion, and convert that into raw frequencies for each corpus, according to corpus size, the normalized rates will be different for each corpus. In other words, the normalized frequency does not provide equally proportional frequency criterion when different size corpora are being compared. To illustrate this problem, I compared the raw frequencies and normalized rates of two corpora of different sizes (80000 and 40000 words) respectively.

For comparison of the two corpora, I have set an identical normalized frequency (40 lexical bundle/one million words) for each corpus:

1. $\frac{80{,}000\times 40}{10{,}00000} = 3.2$ Round down= 3

2. $\frac{40{,}000\times 40}{10{,}00000} = 1.6$ Round up= 2

As a results of conversion process, we have got two converted raw frequencies, 3.2 and 1.6. These frequencies with decimal numbers need be operationalized because for extracting lexical bundles from a corpus we need rounded numbers. Therefore, we use round figures, 3 and 2. In

the following example, these rounded raw frequencies, 3 and 2 have been converted into normalized rates to see if they correspond to set normalized frequency, 40:

1. $\frac{3\times10,00000}{80,000} = \mathbf{37.5}$

2. $\frac{2\times10,00000}{40,000} = \mathbf{50}$

As a result of this conversion, we got different normalized rates from the set normalized frequency criterion, 40. For Corpus 1, we got the normalized rate, 37.5, but for corpus 2, normalized rate is 50, which is much higher. So, not only we have got different normalized rate for each corpus from the set normalized frequency criterion, 40, but also different normalized rates for each corpus. Hence, the basic principle of using normalized frequency criterion, to provide equally proportional frequency criterion for each corpus, is broken when we apply it to different size corpora.

Secondly, if we take normalized frequency criterion and convert that into raw frequencies according to different sizes of each corpus; the converted raw frequency in small size corpora will be lower as compared to large size corpora, because raw frequencies correspond to the size

of the corpus. This difference between raw frequencies of different size corpora affects the number of lexical bundle types generated in each corpus. The small size corpora will generate more lexical bundles because its frequency criterion is low, and the large size corpora will generate less lexical bundle types as the raw frequency criterion will be higher/ stricter in larger corpora. I will illustrate this issue with the example below.

1. $\frac{80{,}000 \times 20}{10{,}00000} = 1.6$ Round down= 2

2. $\frac{40{,}000 \times 20}{10{,}00000} = 0.8$ Round up= 1

In this example, I have used a normalized frequency criterion, 20 for two different size corpora. The converted raw frequencies, on the basis of which we will extract lexical bundle types, are 2 and 1. In corpus 2, the raw frequency is as low as 1, which means any lexical bundle type

occurring only once will be considered a lexical bundle in this corpus. So, the normalized frequency is not appropriate for comparing two different size corpora.

The same problem will occur if we take another normalized frequency criterion, 30 lexical bundles/one million words. The example below illustrates the raw and rounded frequencies:

1. $\frac{80{,}000 \times 30}{10{,}00000} = 2.4$ Round down= 2

2. $\frac{40{,}000 \times 30}{10{,}00000} = 1.2$ Round down= 1

Like previous example, we have got very low raw frequency rates, 2 and 1, that pose the same problems of affecting the number of lexical bundles that will be generated based on these raw frequencies. Therefore, taking normalized frequency criterion is not appropriate for different size corpora. It is better to set a dynamic frequency criterion while comparing different size corpora. In the next section, I will describe the dynamic frequency criterion.

3.3.2 Dynamic frequency threshold

The Dynamic frequency threshold is a frequency criterion through which identical or nearly identical raw frequencies, rather than normalized frequencies, are set for different size corpora. Following this criterion, it is not essential to adopt the standard frequency threshold, 20-40 lexical bundles/one million words, because normalized standard frequency thresholds prove inappropriate while comparing different size corpora. The dynamic criterion is based on the principle that sets such frequency criterion for each corpus that would not affect the number of frequencies of lexical bundles generated in different size corpora. In other words, under this criterion, setting an appropriate raw frequency for each corpus is more important than setting identical normalized frequencies. It is also important that the set raw frequencies of all the corpora should be high enough, that they generate only high frequency lexical bundles from each corpus because lexical bundles are defined as high frequency words.

Chen and Baker (2010) set an identical raw frequency, 4, for each of the three corpora being compared. Later, they checked the normalized rates of these raw frequencies, that were somewhat different (24.3, 25.7, and 27.2) for each corpus. However, Chen and Baker (2010) believe that based on this frequency criteria, they were able to generate almost identical frequencies of lexical bundle types and tokens that were representative of each corpus. If there

is big difference between the generated lexical bundle types and tokens of different corpora, without any solid reason, the results would be less reliable. The detail of frequency criterion and the generated number of lexical bundle types and tokens in Chen and Baker (2010) is given in Table 3.6.

Table 3.6 Frequency criteria followed in Chen and Baker (2010)

Corpus	Size	Set raw frequency	Corresponding normalized frequency	Types	Tokens
1	164,742	4	24.3	118	749
2	155,781	4	25.7	120	757
3	146,872	4	27.2	90	554

Chen and Baker (2016) also took a dynamic frequency approach for comparing different size corpora. They set a criterion that 4-word bundles must occur 4 times in a corpus of 88000 words (normalized rate= 45 occurrences per million words) and 3 times in a corpus of 26000 words (normalized rate= 114 occurrences per million words). In this way, they set nearly identical raw frequencies that have very different normalized rates. However, these raw

frequencies do not affect the number of generated numbers of lexical bundles from each corpus (see Table 3.7).

Table 3.7 Frequency criteria followed in Chen and Baker (2016)

Corpus	Size	Set raw frequency	Converted normalized rates
1	26,356	3	113.8
2	87,970	4	45
3	87,828	4	45

3.3.3 Frequency criterion in the current study

For this study, I have adopted a dynamic raw frequency threshold, 10 lexical bundles for each corpus (Biber & Barbieri, 2007; Chen & Baker, 2010; 2016). The dynamic raw frequency criterion was set in this way because I had to compare three different size corpora. Considering different size of the three corpora, the raw frequency, 10 lexical bundles was considered high enough to generate high frequency lexical bundles across the three corpora. The raw frequency 10 was considered high enough on the grounds that a higher frequency criterion would be inappropriate for the small average size of the texts in expert corpus (7512 words) while

comparing the larger average size of the texts in native student corpus (15,649 words) and non-native student corpus (26,470 words).

The raw frequency threshold, 10 lexical bundles, produced the following converted normalized rates in three corpora:

$\frac{10\times10,00000}{510829}$ = **19.5**.

$\frac{10\times10,00000}{312981}$ = **31.9**

$\frac{10\times10,00000}{502945}$ = **19.8**

As can be seen above that the normalized rates are different in each corpus, though in the expert writers' corpus and in the non-native students' corpus, they are very close, 19.5 and 19.8. However, in the native students' corpus, the normalized rate is quite high, 31.9. As discussed earlier, while comparing different size corpora, it is more important to set an identical raw

frequency, because in this way we can get a close range of lexical bundle types and tokens from each corpus. For example, in this study, the lexical bundles generated, on the basis of dynamic raw frequency, 10 lexical bundles/one million words, are almost similar (95, 92). However, in non-native students' corpora, the types and tokens are much higher. To verify if the higher number of lexical bundle types and tokens in non-native students' corpus generated lexical bundles and the other two corpora is due to set raw frequency, I tried various frequencies and compared the three corpora. In all experiments, I found that the non-native students used much higher number of lexical bundles as compared to expert writers and native students. On the basis of these experiments, I can say that the generated lexical bundles through raw frequency criterion, 10 lexical bundles, are representative of the non-native students' corpus and were not affected by the set frequency. They are there because non-native students used much more lexical bundle types and tokens in their writing.

The details of extracted lexical bundle (after refinement) types and tokens have been shown across the three corpora:

Table 3.8 ***Details of extracted lexical bundle (types and tokens) after refinement***

Corpus	**Size**	**Types**	**Tokens**	**Types**	**Tokens**
Expert writers	502,945	142	2616	95	1844

Native students	312,981	120	1944	92	1553
Non-native students	510,829	382	10164	242	6720

In the next section, I will discuss another important criterion, dispersion.

3.4 The dispersion criterion

The dispersion criterion is the criterion by which a lexical bundle has to occur in several sub corpora in a corpus. This ensures that a generated lexical bundle is not idiosyncratic. As mentioned earlier, the frequency criterion ensures that all lexical bundles meet the minimum frequency criterion in a corpus. However, the frequency criterion does not ensure that the required frequency of a lexical bundle is spread across a corpus. There is a chance that some of the lexical bundles fulfilling the minimum frequency threshold might be found in one or two texts only. For example, in the non-native students' corpus used in this study, *on the other hand*, was used in 16 texts, while, *in the analysis of*, was used 5 times. Similarly, the lexical bundle, *within the relevant framework*, occurred 10 times but it was found only in two texts. There are many such lexical bundles that are found in 2 texts only. These types of bundles

cannot be considered the representative of the whole corpus because they occur only in two texts.

There are two issues related to setting dispersion criterion: the first issue is how to set the dispersion criterion for different size corpora. As in this study, we have three corpora with following sizes:

Corpus 1, 312981 words, Corpus 2, 502945 words, corpus 3, 510829 words.

The second issue is how to set dispersion criterion for corpora comprising a different number of texts. For example, in this study, corpus 1 contains 68 texts, corpus 2, 20 texts, and corpus 3, 19 texts. In the next section, I will discuss these two factors that are important for setting the dispersion criterion.

3.4.1 Setting dispersion criterion

According to the dispersion criterion adopted by Hyland (2008a), a fix dispersion criterion, 10% of the texts is set for each corpus. The detail of dispersion criterion followed by Hyland (2008a) is presented in Table 3.9.

Table 3.9 **Details of dispersion criterion followed by Hyland (2008a)**

Corpus 1: 730,000 words, 120 texts	Dispersion= 12 texts
Corpus 2: 190,0000 words, 20 texts	Dispersion=2 texts
Corpus 3: 825,000 words, 20 texts	Dispersion=2 texts

The problem with this criterion is that when we take this criterion, the corpus with fewer texts results into very small dispersion frequency. As in the above example, for corpus 2 and corpus 3, the dispersion criterion is 2 texts. Due to this low dispersion frequency relatively, more bundles will be generated from corpus 2 and corpus 3 as compared to corpus 1.

In another study, Pan et al. (2016) compared two corpora of different sizes and set a higher dispersion criterion for the corpus having more texts than the corpus having fewer texts,

The detail of dispersion criterion followed by Pan et. al. (2016) is presented in Table 3.10.

Table 3.10 **Details of dispersion criterion followed by Pan et.al., (2016)**

Corpus 1 505,373 words, 87 texts	average size: 5808	Dispersion= 5 texts
Corpus 2 473,912 words, 179 texts	average size: 2647	Dispersion=10 texts

The problem with setting the dispersion criterion according to number of texts occurs when one of the corpora has more texts but the average size of those texts is smaller than the other corpora. As in example above, the corpus 1 contain 87 texts, with an average size, 5808 words, while corpus 2 contains 179 texts with an average size, 2647 words. In principle, 2647 words text, due to its small size would generate fewer bundle types than the text with 5808 words, because more words result into more sequences of words. If a higher dispersion is set for these small size texts, as Pan et al. (2016) did in their study, this would further reduce the number of lexical bundle types in such small size texts. Therefore, setting a much higher dispersion criterion for corpora having more texts affects the number of lexical bundle types in the corpora. In short, the above two studies set the dispersion criterion that is set in proportion to the number of files in a corpus. If there are more files, the dispersion criterion will be higher; if there are fewer files, the dispersion criterion will be smaller.

Chen and Baker (2010) have set a dynamic dispersion criterion of 3 texts for three different size corpora. They compared three corpora of different sizes (see Table 3.11).

Table 3.11 **Details of dispersion criterion followed by Chen and Baker (2010)**

Corpus 1: 164,742 words, 80 texts	2059	Dispersion= 3 texts
Corpus 2: 155,781 words, 60 texts	2596	Dispersion= 3 texts
Corpus 3: 146,872 words, 53 texts	2771	Dispersion= 3 texts

According to this approach, the dispersion is not set in proportion to the number of texts. Instead, the same raw dispersion criterion was set because there is not much difference in the size of the three corpora, as well as in the average size of texts in the three corpora. It is important to consider all these factors because all these, can affect the number of lexical bundle types generated from each corpus. For example, if a higher dispersion is set for corpus 1, as it has the highest number of words as well as texts, it will produce fewer lexical bundle types because of small size of its files.

In another study, Chen and Baker (2016) adopted the dynamic dispersion criterion for comparing three different size corpora. They used three corpora for comparison. The details of the three corpora and the dispersion criteria are as follows:

Table 3.12 **Details of dispersion criterion followed by Chen and Baker (2016)**

Corpus 1: 26,356 words, 189 texts	139 words	Dispersion= 3 texts
Corpus 2: 87,970 words, 239 texts	368 words	Dispersion=4 texts
Corpus 3: 87,828 words, 157 texts	559 words	Dispersion=4 texts

So, according to dynamic dispersion criterion, the dispersion criterion is set by considering the size of the corpus, number of texts and the average size of texts in the corpus. Finally, one

needs to be careful while setting dispersion criteria for different size corpora, because if set much higher dispersion criterion for large size corpora, it will reduce the number of generated lexical bundles in that corpus. Therefore, in the example above, Chen and Baker (2016), set a different dispersion frequency for Corpus 1 and 2 because their size and number of texts are different. However, for corpus 2 they set a dispersion 4 texts which is not much higher than 3 texts.

3.4.2 Dispersion criterion in the current study

For this study, I have set a dynamic dispersion criterion. According to this criterion, a lexical bundle type needs to occur 5 times in each of the three corpora being compared in this study. This criterion was selected considering the size of each corpus, the number of texts and the average size of those texts.

The details of the dispersion criterion in this study are presented in Table 3.13.

Table 3.13 **Details of dispersion criterion followed in the current study**

Corpus 1: 510829 words,	68 texts,	average size :7512 words	5 texts
Corpus 2: 312981 words,	20 texts,	average size: 15,649 words	5 texts
Corpus 3: 502945 words,	19 texts,	average size: 26470 words	5 texts

The dispersion criterion of 5 texts was chosen for corpus 1 and 3 because they are almost similar in size, though there is a big difference in the number of texts in corpus 1 (19 texts) and corpus 2, 68 texts. However, the average size of texts in corpus 2 is much smaller than that of corpus 1. If we set higher dispersion for corpus 1, it will reduce the number of lexical bundle types in it, due to much smaller size of its texts. Hence, 5 texts were considered appropriate for both corpora. In corpus 3, the number of words is much smaller than corpus 1 and 3, and the number of texts is much smaller than corpus 1. On the basis of these two factors, it was thought to set a smaller dispersion for corpus 2. However, considering the fact the number of texts in corpus 2 is greater than corpus 3, and the average size of its texts is greater than corpus 1, I decided not to set a lower dispersion frequency for corpus 2. Considering all these factors, it was decided that 5 texts is an appropriate dispersion criterion for the native students' corpus. Based on the dispersion criteria, 5 texts, the following frequencies of lexical bundle types in each of the three corpora were obtained (see Table 3.14).

Table 3.14 **Details of extracted bundles (types and tokens) in the current study**

Corpus	Size	Types	Tokens	Types	Tokens (after refinement)
Expert writers	502,945	142	2616	95	1844
Native students	312,981	120	1944	92	1553
Non-native students	510,829	382	10164	242	6720

On the basis of this dispersion criteria, we can see that the frequency of refined lexical bundle types is almost similar in expert writers' corpus and native students' corpus. Though the native students' corpus produced much higher types and tokens of lexical bundles, but these frequencies are not affected by dispersion or frequency criterion.

In the next section, I will discuss the third criterion used for the extraction of lexical bundles, the size of lexical bundles.

3.5 Size of lexical bundles

A third criterion for extracting lexical bundles is to set the size of word sequences in lexical bundles. Comparing 4-word lexical bundles with 3-word and 5-word bundles, Hyland (2008a, p. 8.) argues that 4-word bundles provide a 'clearer range of structure and functions than 3-word bundles', and 'they are far more common than 5-word strings.' In this study, I had to compare the use of lexical bundle types and tokens by structural and functional categories. To examine the use of lexical bundles, I needed a size of lexical bundle that could be clearly categorized into structural and functional categories for quantitative and qualitative analysis. The size of 4-words in a lexical bundle provides a much clearer sense for their structural and functional categorization. For example, the lexical bundles, *it should be noted,* and *it can be argued* , can be easily categorized into structural category, Verb based bundles (it + verb

phrase/adverb) and, functional category, Participant oriented bundles (Engagement feature). But if we take 3-word lexical bundles, the bundle would have been *should be noted*, and *can be* argued which do not give a clear sense for their categorization and qualitative analysis. Another option was to take 5-word lexical bundles, but they are very few in a corpus, which makes it difficult to gather a sufficient number of bundles for the type of structural and functional analysis envisaged in this thesis (Cortes, 2004). That is why 4-word bundles are the most suitable choice in terms of size of lexical bundles. For these reasons, I opted to take 4-words bundles for analysis in this study.

3.6 Refinement of extracted lexical bundles

After selecting frequency, dispersion and size of lexical bundles, the last important process is refinement of lexical bundles. For refinement, I decided to take two kinds of lexical bundles from the generated list of bundles.

First, context-dependent bundles, that are bundles related to the topic of the text. For example, lexical bundles like *in second language learning*, *of English language teaching* are related to the field of Applied Linguistics. The reason for taking out context-dependent bundles is that these types of lexical bundles do not present the distinct structural or functional features of

native/non-native/experts writing. Thus, they do not fulfil the purpose of this research; they were excluded from the three corpora.

Second, I excluded overlapping bundles from the final list of extracted lexical bundles. These are 4-word lexical bundles that are part of bigger size, 5 or 6-word lexical bundles, but due to automatic retrieval process, they are separated into 2 or three distinct lexical bundles. For example, the 4-word lexical bundles *it should be noted*, and *should be noted that* are part of a 5-word lexical bundle, *it should be noted that.* The usage of these two lexical bundles was checked through Concordance function, and I found out that the lexical bundle, *Should be noted that*, was followed by '*it'* at all the instances, which confirmed that they are part of 5-word lexical bundle, *it should be noted that.* There were some lexical bundles, that partially overlapped. For example, the lexical bundle *as a result of* occurred 37 times and *a result of the* occurred 12 times. There were 11 instances where both these lexical bundles occurred together. Therefore, *a result of the*, which is a low frequency lexical bundle, was merged into high frequency lexical bundle, *as a result of.* These overlapping lexical bundles were merged because they produce inflated rates of lexical bundle frequencies. Table 3.15 presents the detail of lexical bundle types and tokens in each corpus before and after refinement

Table 3.15 **Detail of lexical bundles (types and tokens) before and after refinement in the current study**

Corpus	Types	Tokens	Types	Tokens
Experts	142	2616	95	1844
Native students	120	1944	92	1553
Non-native Students	382	10164	242	6720

3.7 Corpus Analysis tools

For the corpus analysis in this study, I had to get the lists of lexical bundles along with the frequency of tokens and dispersion from three corpora. Then for the qualitative analysis of those extracted lexical bundles, I needed the concordance lines through which I could analyze the extracted lexical bundles. Finally, for the analysis of discourse functions of the extracted lexical bundles, I needed the detailed view of the texts in each corpus where those lexical bundles occurred. In short, I needed a corpus software that could provide frequency of occurrences of lexical bundles, concordance lines of those lexical bundles and the detailed file view. Considering these requirements of the current study, I decided to use AntConc (Anthony, 2019). It has 7 functions for doing different types of lexical analysis. These functions are

Concordance, Concordance Plot, File View, Clusters/N-Grams, Collocates, Word List, Keyword List. For analyzing lexical bundles, it provides three tools: Clusters/N-Grams, Concordance, and File View. The tool, Clusters/N-Grams, is an inbuilt tool in AntConc for generating lexical bundles. For generating lexical bundles, one can directly select all the corpus files while using this tool. After setting the size of lexical bundles, frequency and dispersion criteria, start button generates a list of lexical bundles in a corpus along with their frequency and dispersion in the corpus. Due to an inbuilt tool of generating list of lexical bundles, the process is very quick and easy. The second tool relevant to lexical bundles is Concordance. The list of lexical bundles can be used to examine the concordance lines of each lexical bundles. If we click on any lexical bundle, all the concordance lines of that lexical bundle appear on the screen. The lexical bundle in each concordance line is coloured that makes it easy to spot the lexical bundle in each line and examine it. The third function in AntConc is File View, that provides a detailed text of any selected lexical bundle. For this function, one needs to enter into the function File View and select a lexical bundle, that opens the location of the file where that lexical bundle occurred. So, AntConc provides all the tools required for the corpus analysis in this study. Therefore, I decided to use AntConc for this study. In this study, I have used

AntConc (3.5.8) for extracting lexical bundles, and the following three tools for analysis of the three corpora:Clusters/N-Grams, Concordance, File View

3.7.1 Clusters/N-Grams

At first step, a frequency list of lexical bundle types and tokens of each bundle was needed. The function ‘n-grams/clusters’ in Antconc provides a frequency list of lexical bundles. For extracting these frequencies, I used the in-built program (N-gram/Cluster) in AntConc and obtained a list of 4-word lexical bundles based on the dispersion criterion, and frequency threshold set for this study. Figure 1 is a screenshot of native students’ corpus in AntConc window, which shows some (rank 101-119) of the lexical bundles generated in the native students’ corpus. These bundles have been extracted from all the 20 files in native students’ corpus.

Figure 3.1 Screeshot of Native students' Corpus Analysis in AntConc Window

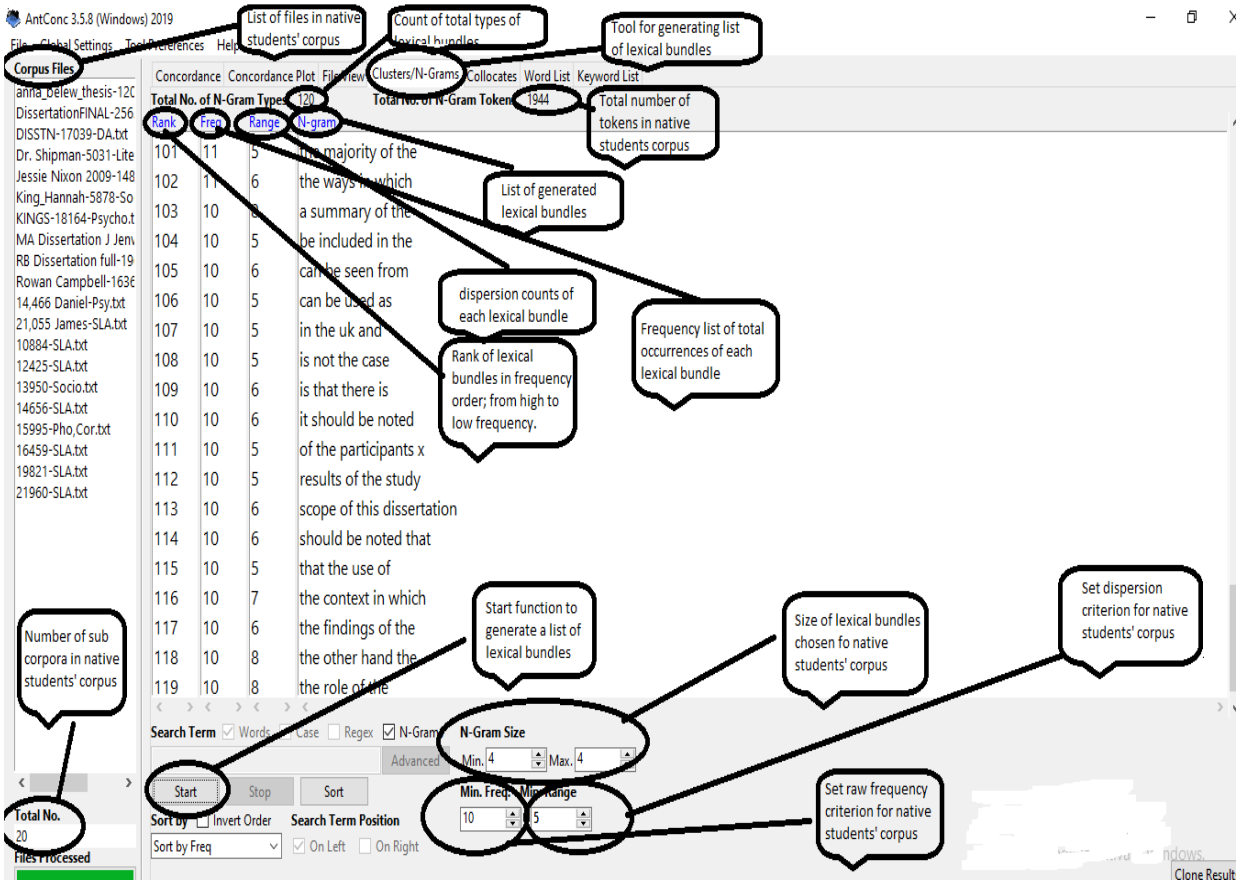


Figure 3.1 shows the output of the corpus analysis of the native students' corpus based on set frequency, 10, and dispersion, 5. The analysis provides a list of lexical bundle types,120, with total number of N-gram tokens, 1944, as well as number of tokens for each lexical bundle in

the second column, frequency. In the output window, there are four columns. The first column, 'rank' is the order of lexical bundle types in terms of frequency, from the most frequent to the least frequent lexical bundles. The second column, 'frequency', provides a list of total occurrences of each lexical bundle type. The third column, 'range', is the number of texts in which lexical bundle types were found. The last column is the list of lexical bundles found in the corpus.

3.7.2 Concordance

Concordance is a tool which shows the lexical bundles in context by using the Concordance tool of AntConc. This tool provides all the occurrences of a lexical bundle type along with its context, enabling the analysis of the usage of different lexical bundles in the context in which they were used. The concordance lines usually include 4 to 5 words on either side of the selected word or phrase. For example, in this study, the concordance lines include 7 words before and 4 words after the selected lexical bundles. The number of words before and after the lexical bundles can be changed as per requirement. In this study, I used the concordance tools to examine and refine the overlapping bundles (see Section 3.6) The screenshot below shows the concordance lines of lexical bundle, *should be noted that*, used in native students' corpus. There were two lexical bundles, *it should be noted, should be noted that*, that were

generated in the list of native students' corpus. When I checked through concordance function of AntConc, I found out that at all the occasions, the lexical bundle, *should be noted that*, was followed by It. This shows that both the bundles, *it should be noted, and should be noted that*, are part of a 5-word lexical bundle, *it should be noted that* (see Figures 3.2 & 3.3). So, we had to filter one of the bundles so that we do not get the inflated frequencies from these overlapping bundles. So, the same function was used for other instances where there was partial overlap of lexical bundles (see Section 3.6) Without Concordance tool, it would have been very difficult to find and filter out the overlapping lexical bundle types.

Figure 3.2 Screenshot of the Concordance lines of lexical bundle 'should be noted that'

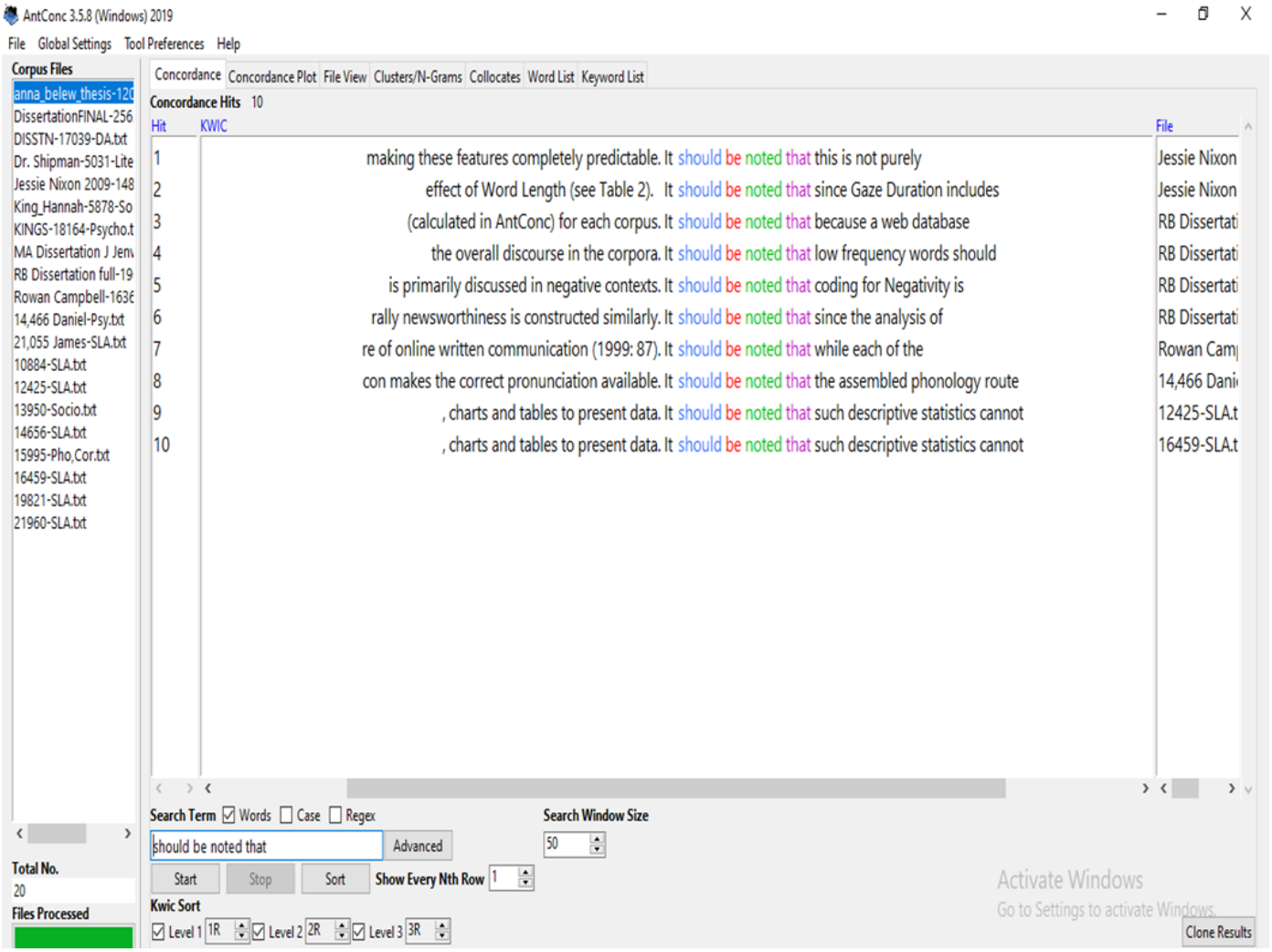
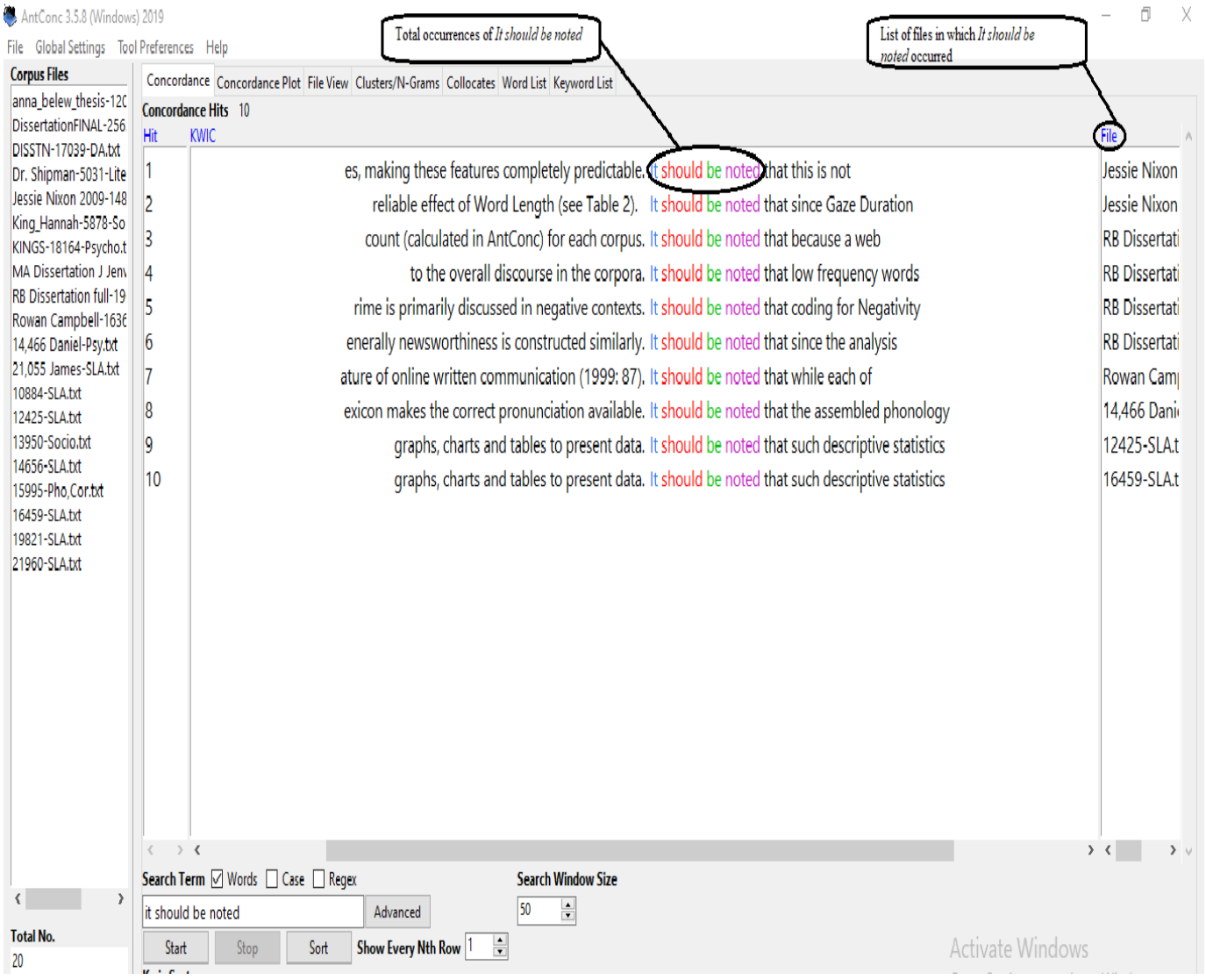


Figure 3.3 Screenshot of the Concordance lines of lexical bundle 'it should be noted'



3.7.3 File view

File view tool in AntConc, provides a detailed view of each file in corpus where a lexical bundle type occurs. In this study, the detailed file view of lexical bundles is required for the qualitative analysis of lexical bundles in different functional subcategories. Through File View, we can examine the functional use of lexical bundle in each corpus and can compare its usage in another corpus. In this study, I have done detailed qualitative analysis of some of the similar lexical bundles used in functional subcategories across the three corpora. By looking at how the similar lexical bundles were used in three different corpora can help us find any differences in the functions of lexical bundles. For this purpose, we can look at lexical bundles' meaning in context; the section in which they are used; the position at which they are used in a sentence, etc. Figures 3.4, 3.5, and 3.6 present the detailed File view of the lexical bundle, *at the same time*, used in the three corpora.

Figure 3 .4 Screenshot of the 'File View in AntConc' presenting the use of 'at the same time' in native students' corpus

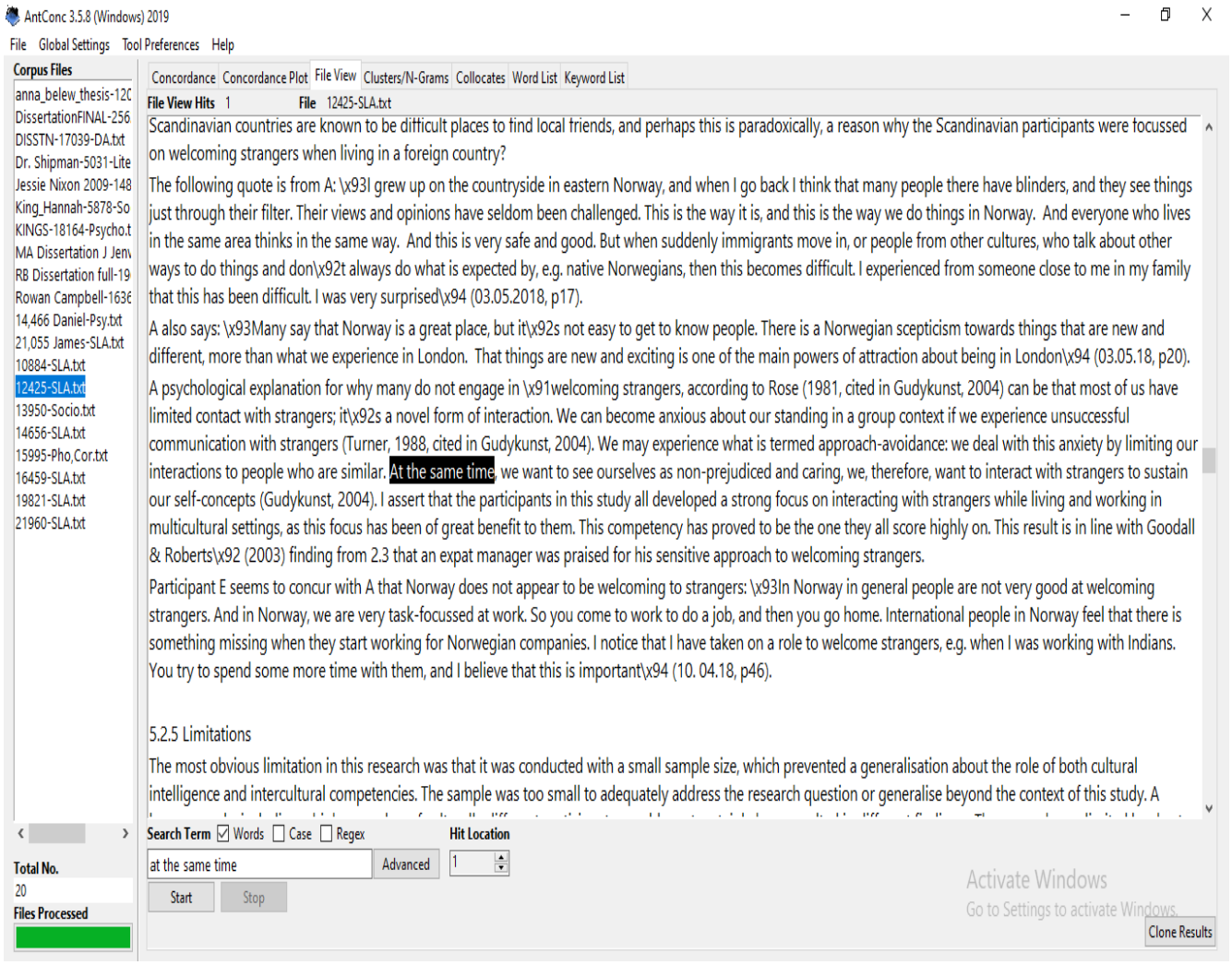


Figure 3.5. Screenshot of the 'File View in AntConc' presenting the use of 'at the same time' in the non-native students' corpus

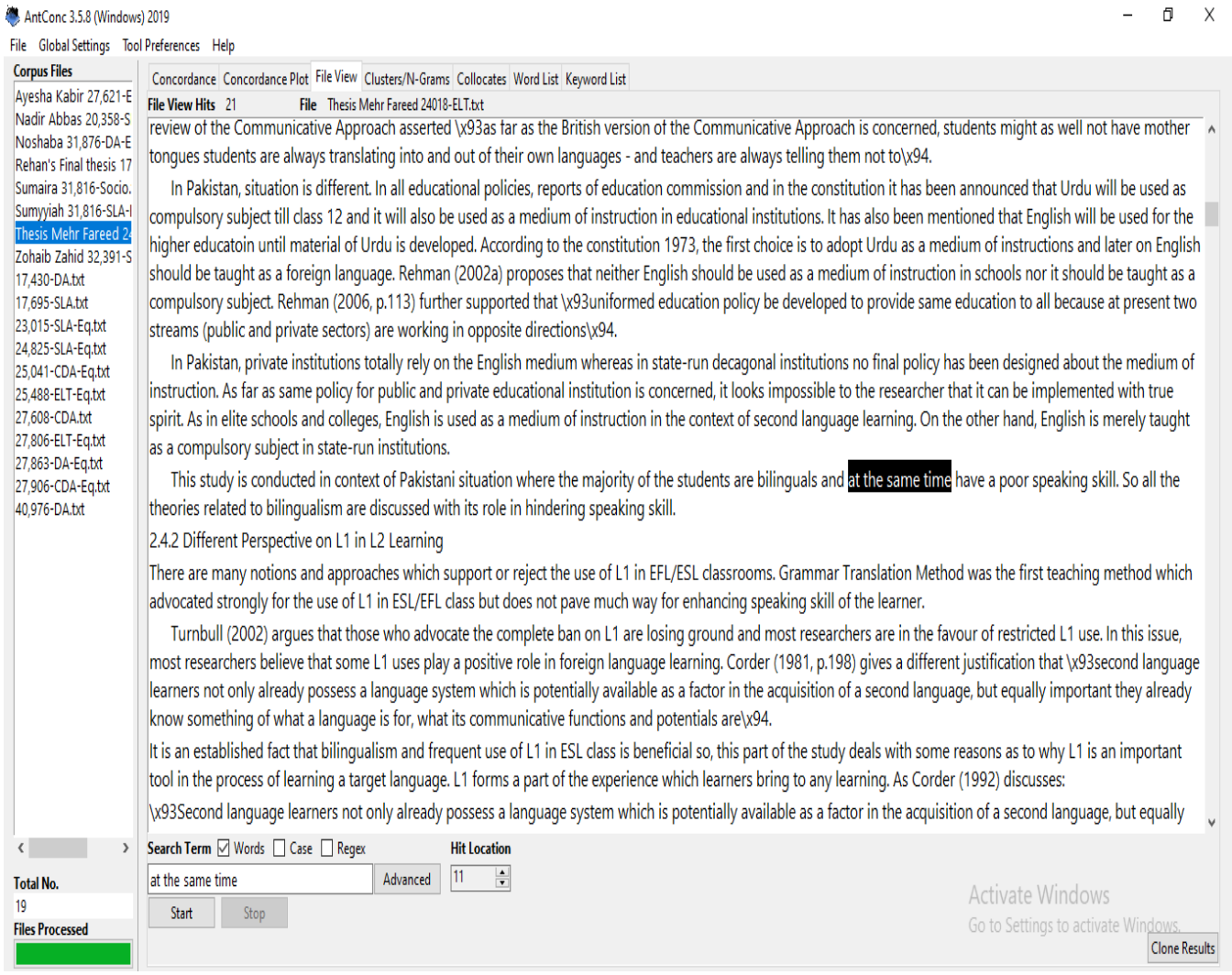
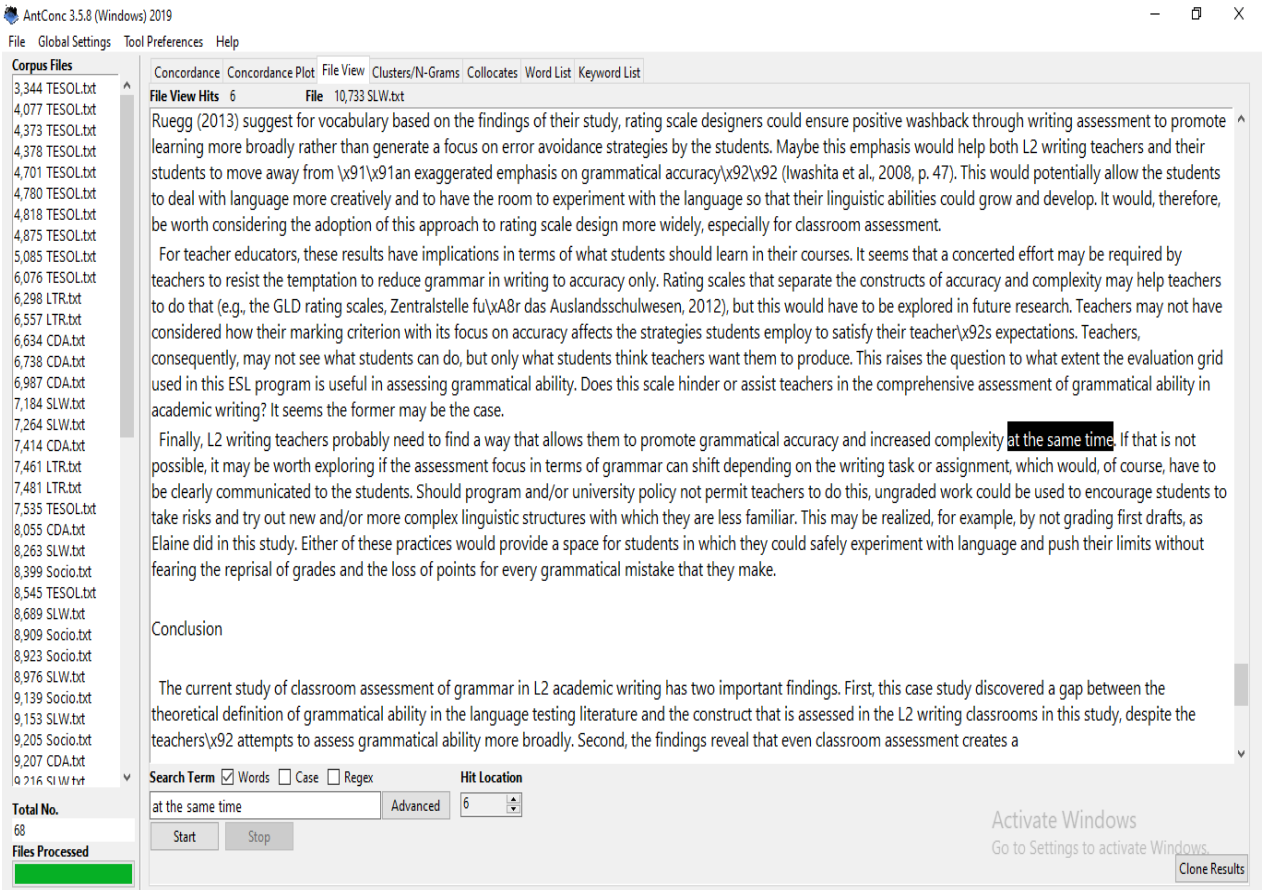


Figure 3.6 Screenshot of the 'File View in AntConc' presenting the use of 'at the same time' in the expert writers' corpus



So, from section 3.4 to 3.7, I have discussed the frequency criterion, 10 occurrences of lexical bundles, dispersion criterion, 5 texts, and size of lexical bundles, 4-words lexical bundles, selected for this study. I also discussed the rationale behind choosing the dynamic raw frequency, the dynamic dispersion criterion, for comparing different size corpora. Moreover, I

discussed the process followed for refining the final list of lexical bundles and the reasons for the refinement for extracted lexical bundles. Finally, I presented the detail of functions used in AntConc for corpus analysis to answer the research questions of this study. In the next section, 3.9, I will discuss the details and rationale of statistical test for this study along with their detailed procedures:

3.8 Statistical Analysis

Log likelihood test has been used for pairwise comparison of structural and functional subcategories of lexical bundle types and tokens across the three corpora. Log-likelihood is a statistical test performed for significant differences in frequency between two corpora. To compare the frequencies of lexical bundle types or tokens, we need the following frequencies:

1. Total number of lexical bundle types and total number of words in corpus 1
2. Total number of lexical bundle types and total number of words in corpus 2
3. Total number of lexical bundle types and total number of words in corpus 1

The rationale behind using log likelihood test was to identify the significant differences in lexical bundle types and tokens in pairwise comparison of the three corpora. For this purpose,

the test considers the total occurrences of lexical bundle types in a corpus as well as total number of words in the corpus.

The higher the log-likelihood value is, the more significant the results would be. The Log-likelihood value must be above 3.84 for the difference to be significant at $P < 0.05$ level, and above 6.63 to be significant at $P < 0.01$ level. For calculating the log-likelihood in this study, a log-likelihood calculator (Rayson, 2016) was used. As the data analysis in the current study is based on qualitative as well as a quantitative analysis, in the following section I will describe the procedures involved in the qualitative analysis of the data in the study.

3.9 Qualitative analysis

In the current study, the process for the qualitative analysis involved the structural and functional sub-categorization of the retrieved lexical bundles, and the analysis of bundle frames. In this section, I will describe the sub-categorize that were used for the analysis.

The structural sub-categories involved the sub-categorization were the 12 sub-categories proposed by Biber et al. (1999). These sub-categories are Noun phrase with of- phrase fragment, Noun phrase with other post-modifier fragments, Prepositional phrase with embedded of-phrase, prepositional phrases with post-modifier fragment, Anticipatory it + verb

/ adjective phrase, Passive verb + prepositional phrase fragment, Copula be + noun / adjective phrase, (Verb phrase) + that- clause fragment, (Verb/ adjective) + to-clause fragment.

The next step of qualitative enquiry involved the investigation of the discourse functions served by lexical bundles. For this purpose, the retrieved bundles were grouped into three main categories and ten sub-categories proposed by Hyland (2008a; 2008b). The main categories are (Research-oriented bundles, Text-oriented bundles, and Participant-oriented bundles), and the sub-categories are Location, procedure, description, quantification, transition signals, resultative signals, structuring signals, framing signals, stance features, and Engagement features.

For the qualitative analysis, bundle frames have also been used, e.g., a/the + Noun + of the/a, in + the + Noun + of/a. These two frames have been shown to be the most productive in academic writing (Biber et al., 2004; Chen & Baker, 2010).

The qualitative analysis also involved the contrastive analysis of the bundles used in the sub-categories across the three corpora. For example, the experts and the native students used different Participant-oriented bundles compared to the non-native speakers.

3.10 Conclusion

In this chapter, I have presented the methods adopted for developing the corpus for this study. I also discussed different frequent criteria set in previous studies, and the rationale for choosing different frequency criteria for different size corpora. The various features of the Concordance software, AntConc, used in the analysis of this study were also discussed in detail. The next chapter presents the results of the analysis.

Chapter 4

Results

4.1 Introduction

This chapter presents the analysis of lexical bundles used in the expert, native student, and non-native student corpora. The analysis includes bundle frequencies and usage patterns in each of the three corpora. The chapter has been divided into four main sections. The first section deals with the expert corpus, including the frequency list of the top 20 bundles, the distribution of bundle structures as well as bundle functions. The second and third sections report in the same way on the native and non-native student corpora. A comparison of the three corpora is then presented in the final section.

4.2 Expert corpus

The expert corpus is based on research articles published in reputable journals in the field of Applied Linguistics. All these journals were selected based on their impact factor and subject matter. The impact factor of these journals was 1 or above, whereas the subject matter was determined as corresponding to the subject matter of the other two corpora (see Section 3.2.4).

The size of the corpus is 505,945 words with 68 texts. After the analysis of this corpus, 95 lexical bundles were extracted with 1884 tokens. This section will analyze the use of top 20 bundles, and the distribution (structural and functional) of all the bundles in the expert corpus.

4.2.1 Top 20 bundles in the expert corpus

Table 4.1 presents the list of the top 20 lexical bundles used in the expert corpus, presented in order of frequency.

Table 4.1 Top 20 lexical bundles used by expert writers

Rank	Lexical bundle	Tokens
1.	in the context of (the)	56
2.	at the same time	54
3.	in the case of	54
4.	on the basis of (the)	47
5.	over the course of (the/a)	46
6.	on the other hand	43
7.	the end of the	43
8.	in terms of the	39
9.	it is important to (note that)	39
10.	at the beginning of (the)	38
11.	the ways in which	38

12.	as well as the	37
13.	in this case the	33
14.	on the part of (the)	33
15.	(that/in/on) the use of the (a)	33
16.	the extent to which	31
17.	in the form of	30
18.	as a result of	28
19.	in relation to the	25
20.	the results of this (study)	24

The majority (17/20) of the most frequent bundles in the list are bundles used for organising the text. For example, the most frequent bundle *In the context of*, is used to contextualize information:

(1) *Linguistic variation has consistently been found to have social meaning in its association with the status and stance of speakers* ***in the context of*** *interaction*. (ES41)

The second most frequent bundles, *at the same time* and *In the case of*, are used to connect two sentences by giving parallel information and to contextualize new information respectively:

(2) *It seems that elective use of topic knowledge helps the learners to function electively in everyday situations in the L2, while it may* ***at the same time*** *inhibit further development of their linguistic knowledge*. (ES67).

(3) ***In the case of*** *pain in clinical encounters* […]. (ES52)

The only three bundles in the top 20 which were not used to organise the text, served to describe the research or procedure, or to stress the importance of a statement: *the end of the story*

(4) *Interlocutors tended not to interrupt but waited until* ***the end of the*** *story to ask questions*. (ES11), *the use of the,*

*(*5) The use of the noun serves to background the processes themselves […]. (ES18), *it is important to*. (6) ***It is important to*** *note that all of the questions are reflexive* […]. (ES25)

So, in the expert corpus, the majority of the top 20 most frequent bundles have been used for organising the text. This shows that the expert writers use bundles more for organising the text than for describing research.

In the next section, the analysis and the results of the bundle structures will be presented.

4.2.2 Structural characteristics of lexical bundles

The previous studies have divided bundles into different structural corelates (Biber et al., 1999;; Chen and Baker, 2010; Hyland, 2008a; 2008b). In this study, we have adopted the structural classification from Hyland (2008a; 2008b) as discussed in section 2.3.3., dividing lexical bundles into four main categories: noun-based bundles, preposition-based bundles, verb-based bundles, and other structures. Table 4.2 presents the frequency of these four structural categories, in terms of types and tokens, both raw figures and percentages.

Table 4.2 Frequency of structural categories (types and tokens) in expert corpus

Structure	Types	%	Tokens	%
Noun-based bundles	**24**	**25.2%**	**435**	**23.58%**
Noun-based bundles with of-phrase fragment	16	16.84%	284	15.40%
Noun-based bundles with other post modifier fragment	8	8.42%	151	8.18%
Preposition-based bundles	**47**	**49.4%**	**1048**	**56.83%**
Preposition-based bundles with of-phrase fragment	29	30.52%	662	35.90%
Preposition-based bundles with other post modifier fragment	18	18.94%	386	20.93%
Verb-based bundles	**18**	**18.9%**	**262**	**14.19%**
Copula be + NP/Adj phrase	1	1.05%	13	0.70%

VP with active verb	0	0%	0	0%
Anticipatory it + VP/Adj phrase	6	6.31%	103	5.58%
Passive verb + PP fragment	4	4.21%	63	3.41%
VP+ that clause fragment	3	3.15%	34	1.84%
Verb/adj + to clause fragment	4	4.21%	49	2.65%
Other structures		**6.31%**		**5.34%**
	6	6.31%	99	5.34%
Total	**95**	**100%**	**1844**	**100%**

In the expert corpus, Preposition-based bundles are the most common structures, representing 49% types and 57% tokens of total bundles. The second most common structure is the Noun-based bundles, which represents about a quarter of types and tokens of total bundles. Verb -based bundles, also known as clausal bundles, are the least common bundles roughly representing 19% types and 14% tokens of total bundles.

The next section analyses each structural category and its use in turn, focusing on the most common sub-categories.

4.2.2.1 Noun -based bundles

Noun-based bundles are headed by a noun e.g., *the end of the, the context in which* etc. These bundles have two sub-categories: Noun-based bundles with of-phrase fragment, Noun-based bundles with other post-modifier fragment. Table 4.3 presents all the noun-based bundles found in the expert corpus.

Table 4.3 Noun-based bundles in expert corpus

Noun-based bundles	
Noun-based bundles with of-phrase fragment	the end of the, (that/in/on) the use of the (a), the results of this (study), the results of the, a wide range of, the nature of the, the role of the, the purpose of this, the purpose of the, the total number of, the analysis of the, the case of the, the content of the, the design of the, the meaning of the, the scope of the,
Noun-based bundles with other post-modifier fragment	the ways in which, the extent to which, the way in which, an important role in, (to) the fact that the, the relationship between the, the context in which, the degree to which

The category Noun-based bundles with of-phrase fragment represents only the noun-based bundles that consist of the fixed frame ‘the/a + Noun+ of the/a’. This Noun-based bundle frame is considered to be one of the most productive bundle frames in academic writing (Biber et al., 2004; Chen & Baker, 2010). Considering the importance of these bundle frames in academic writing, only those bundles were kept in the category ‘Noun-based bundles with of-phrase fragment’ that contained the frame ‘the/a + Noun+ of the/a’. All the other noun-based bundles with of-phrase fragment, e.g., *majority of the respondents*, *the effect of task* etc. were categorized as other structures. Moreover, ‘Noun-based bundles with of-phrase fragment’ with the frame ‘the/a + Noun+ of the/a’ represented a wide range of bundles whereas the ‘Noun-based bundles with of-phrase fragment’ without the frame ‘the/a + Noun+ of the/a’ represented a limited range of bundles (see Table 4.3 and Table 4.4). As can be seen from the Table 4.4, Noun-based bundles with of-phrase fragment are the more common NP-based bundles, representing the majority of the Noun-based bundles.

Table 4.4 Frequency and percentage of Noun-based bundles (types & tokens) in expert corpus

Structure	Types	%	Tokens	%
Noun-based bundles	**24**	**25.2%**	**435**	**23.58%**
Noun-based bundles with of-phrase fragment	16	16.84%	284	15.40%
Noun-based bundles with other post modifier fragment	8	8.42%	151	8.18%

Noun-based bundles represent 17% types and tokens of the total bundles. The majority of these bundles (13% types and tokens of total bundles) have been used for describing time, e.g., *the end of the*, (7) *Interlocutors tended not to interrupt but waited until the end of the story to ask questions*. (ES 12), denoting qualities, e.g., *the use of the*, (8) *the meso-macro movement illustrates how the use of the metaphor is superseded by other rhetorical strategies* […] (ES 22), providing justification, e.g., *the purpose of the*, (9) *The purpose of this open-ended prompt* […] (ES 09)

The rest of the 3% types and tokens of the total bundles have been used for organizing text, e.g., *the results of this, the results of the, the analysis of the.*

The Noun-based bundles with other post modifier fragments represent 8% types and tokens of the total bundles. All the Noun-based bundles with other post modifier fragments are used to

organize the text, i.e. through contextualisation, e.g. *the context in which, the ways in which,* (10) *Task-based type approaches address* ***the ways in which*** *learners may achieve* […] (ES 08), through highlighting the importance of a proposition e.g., *an important role in, the fact that the,* (11) ***The fact that the*** *writing tutor was a first-year student* […] (ES 07).

Although there are only 8 bundle types in this sub-category, there are 5 bundle types that are a variation of the frame '*the __ in/to __ which'*.

Following is a summary of the main characteristics of the Preposition-based bundles used in the expert corpus:

- In the expert corpus, Noun-based bundles represent a quarter of all bundles; Noun-based bundles with of-phrase fragment twice as common as the Noun-based bundles with other post-modifier fragments. Half of the Noun-based bundles with of-phrase fragment (13% of the total bundles) have been used to describe research.
- All the Noun-based bundles with other post modifier fragments (8% of the total bundles) have been used for organisation of the text.

4.2.2.2 Preposition-based bundles

Preposition-based bundles are headed by a preposition e.g., *in the context of*. There are two sub-categories of these bundle structures: PP with of-phrase fragment, and PP with other post-modifier fragment. These are the most common bundles representing 49% types and tokens of total bundles in the expert corpus. Table 4.5 displays all the preposition-based bundles found in the expert corpus.

Table 4.5 Preposition-based bundles in expert corpus

Preposition-based bundles	
Preposition-based bundles with of-phrase fragment	in the context of (the), in the case of, on the basis of (the), over the course of (the/a), in terms of the, at the beginning of (the), on the part of (the), in the form of, as a result of, in the process of, through the use of, at the time of, in the course of, as a way of, at the expense of, in the field of, in the light of, as a means of, as a form of, as part of a, at the level of, from the perspective of, at the university of, in the middle of, in the number

	of, for the purposes of, in the development of, in the face of, at the end of,
Preposition-based bundles with other post-modifier fragment	at the same time, on the other hand, in this case the, in relation to the, with respect to the, in the present study, in this article we, as part of the, in response to the, in line with the, with regard to the, in a way that, in the current study, in this way the, in terms of their, in addition to the, in the same way, in this section we,

As can be seen from Table 4.5 and Table 4.6, Preposition-based bundles with of-phrase fragment are twice as common as the Preposition-based bundles with other post-modifier fragment.

Table 4.6 Frequency and % of Preposition-based bundles (types & tokens) in expert corpus

Structure	Types	%	Tokens	%
Preposition-based bundles	**47**	**49.4%**	**1048**	**56.83%**
Preposition-based bundles with of-phrase fragment	29	30.52%	662	35.90%
Preposition-based bundles with other post modifier fragment	18	18.94%	386	20.93%

Preposition-based bundles with of-phrase fragment

In the expert corpus, Preposition-based bundles with of-phrase fragment are the most common bundles representing 31% types, and 36% tokens of the total bundles. Half (16% of the total bundles) of the PP-based bundles with of-phrase fragment have been used for contextualizing new information in the text e.g., *in the face of:* (12) *These learners' low writing self-efficacy judgments resulted in them viewing difficult tasks as personal threats and giving up quickly* ***in the face of*** *difficulty*. (ES40).

The other half (15% of the total bundles) of the Preposition-based bundles with of-phrase fragment have been used for describing research, such as, for referring to time, e.g., *at the time of*, (13) ***At the time of*** *writing* […] (ES27), for referring to people, e.g., *on the part of*, (14) […] ***on the part of*** *people who serve tourists*. (ES30), and for referring to the discipline, e.g., *in the field of*, (15) ***In the field of*** *math and science education, researchers have recently employed various methods*. (ES 62)

It is interesting to note that the majority (18% of types of the total bundles) of the Preposition-based bundles with of-phrase fragment are a variation of two highly productive frames (cf.2.4.2): The first one is '*in the ___ of*' (12% types, and 11% tokens of the total bundles)

where the blank slot was filled with number of words e.g., *context, case, form, process, course, field, light, number, development,* and *face*.

The second most productive structure is '*at the ___ of*' with (6% types, 6% tokens of the total bundles). The bundles in this frame are used to identify the time e.g., *at the beginning of, at the end of,* to identify the place e.g., *at the university of,* and to quantify, e.g., *at the level of.*

Preposition-based bundles with other post-modifier fragment

Preposition-based bundles with other post-modifier fragment represent 19% types and 21% tokens of the total bundles (see Table 4.6). All these bundles have been used for organizing the text, e.g., by contextualizing new information in text, e.g., *in line with the,* (16) ***In line with the*** *predictions of the Involvement Load Hypothesis* […] (ES 45), by referring to the study itself, e.g., *in the present study,* (17) ***In the present study****, learners highly valued encountering the same words repeated in different subgames*. (ES12), by adding information, e.g., *in addition to the,* (18) ***In addition to the*** *classroom observation data* […] (ES49)

The majority of these types (14% of the total bundles) involved the use of the preposition 'in', out of which 6 types are different variations of the bundle frame *'in ___ ____ the'*, e.g., *in this case the, in relation to the, in response to the, in line with the, in this way the, in addition to the*.

Following is a summary of the main characteristics of the Preposition-based bundles used in the expert corpus:

- Preposition-based bundles are by far the most common bundles (49% types, 57% tokens of the total bundles) in the expert corpus.
- The majority (34% of the total bundles) of these bundles have been used for organizing text.
- The majority of the Preposition-based bundles with of-phase fragment (15% of the total bundles) and all the Preposition-based bundles with other post-modifier fragment (19% of the total bundles) have been used for organizing the text.

4.2.2.3 Verb-based bundles

Verb-based bundles are not very common in the expert corpus representing 19% types and tokens of the total bundles. There are six sub-categories into which the verb-based bundles have been categorized (Hyland, 2008a; 2008b). These sub-categories are Copula be + Noun/adjective phrase, Verb-based bundles with active verb, Anticipatory it + verb + (adjective phrase), Passive verb + Prepositional fragment, Verb-based bundles with to-clause fragment, and Verb-based bundles with that-clause fragment.

Table 4.7 presents the list of all the verb-based bundles found in the expert corpus:

Table 4.7 Verb -based bundles in expert corpus

Verb-based bundles	
Copula be + noun /Adjective phrase	there is a need
Verb-based bundle with Active Verb	
Anticipatory it + verb + (Adjective phrase)	it is important to (note that), it is possible that, when it comes to, it should be noted, it can be argued + that, it is possible to
Passive verb + prepositional phrase fragment	can be seen in, can be seen as, (that) can be used to, is based on the
Verb-based bundles with that-clause Fragment	that is to say, that there is a, that there is no
Verb-based bundles with to-clause Fragment	to be able to, in order to be, not be able to, are more likely to

As shown in Table 4.7, anticipatory it + verb / adjective phrase, represent most of the Verb-based structure, whereas Verb-based bundles with active verb were not found in the expert corpus. Table 4.8 presents the distribution of Verb-based bundles in expert corpus.

Table 4.8 Frequency and % of Verb-based bundles (types & tokens) in expert corpus

Structure	Types	%	Tokens	%
Verb-based bundles	**18**	**18.9%**	**262**	**14.19%**
Copula be + noun/adj. phrase	1	1.05%	13	0.70%
Verb-based bundles with active verb	0	0%	0	0%
Anticipatory it + verb/adj. phrase	6	6.31%	103	5.58%
Passive verb + PP fragment	4	4.21%	63	3.41%
Verb-based bundles with that clause fragment	3	3.15%	34	1.84%
Verb-based bundles with to clause fragment	4	4.21%	49	2.65%

In the following lines, the analysis of these sub-categories will be presented.

The anticipatory it + verb/adj. phrase bundles

The anticipatory it + verb/adj. phrase bundles represent 6% of total bundles in the expert corpus. These bundles have been used for presenting writers' evaluation, e.g., *it is possible that*, (19), *Theoretically, **it is possible that** students* […] (ES17), and to engage the reader, e.g.,

it should be noted, (20) ***it should be noted****, however, that a more explicit and wide-ranging analysis of the data is being conducted* (cf. Pienemann1987a). (ES59)

Passive verb + prepositional fragment

The verb-based bundles with Passive verb + prepositional fragment represent 4% types, and tokens of the total bundles. These bundles have been used to describe research e.g., *can be seen as* (21), *Adult learners in a classroom setting* ***can be seen as*** *engaging in socialization into English* […] (ES 22).

Verb-based bundles with to-clause fragment

The Verb-based bundles with to-clause fragment also represent 4% of types and tokens of the total bundles. These bundles have also been used for describing research, such as, for describing instructions e.g., *not be able to,* (22), *Learners were informed that they would* ***not be able to*** *keep the scripted dialogues* […] (ES20), to justify, e.g., *in order to be*, (23) […] *learners ought to be given time to familiarize themselves with Alexa and other IPAs* ***in order to be*** *effective users*. (ES07)

Verb-based bundle with that-clause fragment

The Verb-based bundle with that-clause fragment represent 3% of total bundles. These bundles have also been used to describe research, such as, to make a statement, e.g., *that there is a*, (24)

Many L1 and L2 researchers claim ***that there is a*** *symbiotic relationship* […] (ES 42)

Copula be + noun/adjective phrase

The Verb-based bundles with Copula be + noun/adjective phrase represent merely 1% of total bundles. There was only one bundle in this category that was used for presenting writers' opinion, such as, presenting recommendations, e.g., *there is a need*.

Following is the summary of the main features of Verb phrase bundles and their usage in the expert corpus:

- Verb -based bundles are not very common in the expert corpus representing 19% types and tokens of the total bundles.
- The majority (13% of the total bundles) of the verb-based bundles have been used for describing research, such as providing explanations, giving instructions, and making statements. The rest of the verb-based bundles (6% of the total bundles) have been used for presenting writers' opinion, such as presenting possibilities, and giving recommendations.

4.2.2.4 Other Structures

Other structures consist of bundles that might have noun/preposition/verb component, but their structure is different from the other bundles in those categories. All the bundles categorized as other structures found in the expert corpus have been listed in Table 4.9.

Table 4.9 Other bundles in expert corpus

Other structures	
Other Structure	as well as the, (the) participants in this study, (is) one of the most, the amount of time, as well as their, the effects of task,

Table 4.10 presents the distribution of Other structures (types and tokens) in expert corpus.

Table 4.10 Frequency and % of Other structures (types & tokens) in expert corpus

Structure	Types	%	Tokens	%
Other structures		**6.31%**		**5.34%**
	6	6.31%	99	5.34%

The majority (4%) of these bundles have been used for describing research, such as, for referring to the participants, e.g., *the participants in this study, the effects of task,* (25) *During their first year at Hope College, **the participants in this study** were required* […] (ES 02).

The distribution and use of bundle structural characteristics show that the bundles used for organising the text are the most frequent bundles in the expert corpus. The majority of the preposition-based bundles (34% of the total bundles) have been used for organizing text. These are the most frequent bundles. Similarly, half of the Noun-based bundles (12% of the total bundles) have also been used for organizing text. The majority of the Verb-based bundles (13% of the total bundles) have been used for describing research, though these bundles are not very common representing 19% of the total bundles.

The main structural characteristics of lexical bundles in the expert corpus can be summarised as follows:

- In the expert corpus, Preposition-based bundles are the most common bundles, representing 49% types and 58% tokens of the total bundles, majority (34% of the total bundles) of which were used for organizing the text.

- The Noun-based bundles represent the quarter of the total bundles, half (16% of the total bundles) of which were used for organizing text, and the half (15% of the total bundles) were used for describing research.
- The Verb-based bundles are not very common, representing 19% of total bundles, the majority (13% of the total bundles) were used for describing research, and 6% of the total bundles were used for presenting writers' opinion.
- The main structural characteristic of the bundles in the expert corpus is that they are used to present information with a reference point; specify the situation, place, and time; refer to the text; and to compare and contrast the given information. 50% of the total bundles have been used for these purposes.

In the next section, the results and analysis of bundle functions in the expert corpus will be presented.

4.2.3 Functions of lexical bundles in the expert corpus

The functional taxonomy presented by Hyland (2008a; 2008b) was used to classify bundle function in the expert corpus, as discussed in section 2.3.3. The discourse functions of lexical bundles are an important part of academic discourse and academic writers use them for

different purposes related to research procedures, descriptions, coherence of the text, presentation of writers' viewpoint, engaging the reader etc.

These functions are classified into three main categories:

i. Research-oriented bundles

They are used for providing descriptions, e.g., *the design of the, can be used to* etc., to quantify, e.g., *the majority of the, a wide range of* etc., and to indicate time and place of an event, e.g., *at the end of, at the beginning of, all over the world* etc.

ii. Text-oriented bundles

They are used for organizing the text, i.e., through referring to the outcome or findings, e.g., *as a result of, the results of the* etc., through bundles that show transition from one idea to another, e.g., *on the other hand,* through bundles that put information in context, e.g., *in the context of, in the way which* etc.,

iii. Participant-oriented bundles

They are used to describe the writers' evaluation or that of other scholars, e.g., *of the view that, it is difficult to,* or to engage the reader e.g., *it is important to, it is worth noting* etc.

Table 4.11 presents the distribution of bundle functions with their corresponding frequencies in the expert corpus. The raw frequencies and percentages have been presented in the table.

Table 4.11 Frequency and % of bundle functions (types & tokens) in expert corpus

Functions	**Types**	**Tokens**
Research-oriented bundles	**39**	**640**
Location	5	115
Procedure	18	281
Quantification	6	100
Description	10	144
Text-oriented bundles	**48**	**1072**
Transition signals	6	171
Resultative signals	3	74
Structuring signals	5	86
Framing signals	34	741
Participant-oriented bundles	**8**	**132**
Stance features	5	63
Engagement features	3	69
Total	**95**	**1844**

As can be seen from Table 4.11 above, Text-oriented bundles represent 50% of total bundles in the expert corpus. This shows that the main function of lexical bundles in expert writing is to organize text. The research-oriented bundles represent 41% of total bundles. The participant-oriented bundles, representing the writers' evaluation and that of other scholars, are the least common bundles representing 8% of the total bundles used.

In short, the main function of the bundles used in expert writing appears to be organization of the text. In the next section, the detailed analysis of the bundle functions will be presented.

4.2.3.1 Research-oriented bundles

Research-oriented bundles 'help writers structure their activities and experiences of the real world' (Hyland, 2008a, p.31). There are four sub-categories of these bundles: location, procedure, description, and quantification. These bundles represent 41% of total bundles in the expert corpus. Table 4.12 presents all the research-oriented bundles found in the expert corpus:

Table 4.12 Research-oriented bundles in the expert corpus

Research-oriented bundles	
Location	the end of the, at the beginning of (the), at the university of, in the middle of, at the end of
Procedure	(that/in/on) the use of the (a), (the) participants in this study, as part of the, in the field of, an important role in, as part of a, the purpose of the, the analysis of the, the meaning of the, the effects of task, the purpose of this,
Quantification	the extent to which, a wide range of, in the number of, (is) one of the most, the amount of time, the total number of
Description	the nature of the, the role of the, as a form of, the case of the, the content of the, the design of the, the scope of the

As can be seen above, research-oriented procedure bundles represent nearly half of the research-oriented bundles. Table 4.13 presents the distribution of Research-oriented bundles in expert corpus.

Table 4.13 ***Frequency and % of Research-oriented bundles (types & tokens) in expert corpus***

Functions	**Types**	**%**	**Tokens**	**%**
Research-oriented bundles	**39**	**41.05%**	**640**	**35.05%**
Location	5	5.26	115	6.23
Procedure	18	18.94	281	15.23
Quantification	6	6.31	100	5.42
Description	10	10.52	144	7.80

In the following lines, the detail analysis of the sub-categories will be presented.

Procedure bundles

The procedure bundles represent 19% of total bundles. These bundles have been used for describing the procedures of research, e.g., *the use of the*, (26) ***The use of the*** *noun serves to background the processes* […] (ES 11), *the effects of task*, (27) *The theoretical rationale for*

examining ***the effects of task*** *environment factors such as planning time and task conditions*. (ES 33)

Description bundles

Description bundles represent 11% of total bundles. These bundles have been used to describe features of research and their characteristics, e.g., *the nature of the,* (28) *Given* ***the nature of the*** *construct we investigate* […] (ES46)*, the content of the,* (29) […] *it is not* ***the content of the*** *lesson that is the basis for learning.* (ES 49)

Quantification bundles

Quantification bundles represent 6% of total bundles. They are used to quantify, e.g., *in the number of,* (30) […] *variation* ***in the number of*** *negative particles* (ES41)*, one of the most,* (31) ***one of the most*** *salient distinguishing features* (ES34)

Location bundles

The research-oriented location bundles represent 5% of total bundles. They are used to identify the place/location/time of an event e.g., (32) *at the end of. **At the end of*** **the** *article* […] (ES04)*, in the middle of the,* (33) ***in the middle of*** *the school year* […]. (ES 52)

The main features of research-oriented bundles in the expert corpus can be summarised as follows:

- Research-oriented bundles represent 41% of the total bundles.
- The procedure bundles represent majority (19% of the total bundles) of research-oriented bundles, indicating that describing procedure of research is the most important function of these bundles.

In the next section, the results and the analysis of the text-oriented bundles will be presented.

4.2.3.2 Text-oriented bundles

Text-oriented bundles are used for the organization of text. There are four sub-categories of these bundles: Transition signals, Resultative signals, Structuring signals, and Framing signals. These bundles represent 51% of the total bundles, with text-oriented framing signals representing the majority (36%) of the total bundles. Table 4.14 presents all the text-oriented bundles found in the expert corpus:

Table 4.14 ***Text-oriented bundles in expert corpus***

Text-oriented bundles	
Transition signals	at the same time, on the other hand, as well as the, that is to say, in addition to the, as well as their
Resultative signals	as a result of, the results of this (study), the results of the
Structuring signals	in the present study, in this article we, in the current study, the purpose of this, in this section we, in the development of
Framing signals	in the context of (the), in the case of, on the basis of (the), over the course of (the/a), in terms of the, the ways in which, in this case the, on the part of (the), in the form of, in relation to the, with respect to the, in the process of, through the use of, at the time of, in response to the, in the course of, the way in which, as a way of, at the expense of, in line with the, in the light of, with regard to the, as a means of, in a way that, in this way the, (to) the fact that the, the relationship

	between the, in terms of their, at the level of, from the perspective of, in the same way, is based on the, when it comes to, in order to be, for the purposes of, in the face of, that there is a, that there is no, the context in which, the degree to which

As can be seen in Table 4.15 that framing signals represent the majority of the text-oriented bundles.

Table 4.15 ***Frequency and % of Text-oriented bundles (types & tokens) in expert corpus***

Functions	**Types**	**%**	**Tokens**	**%**
Text-oriented bundles	**48**	**50.52%**	**1072**	**58.13%**
Transition signals	6	6.31	171	9.27
Resultative signals	3	3.15	74	4.01
Structuring signals	5	5.26	86	4.66
Framing signals	34	35.78	741	40.18

Framing signals

Text-oriented framing signals are the most common bundles representing 36% types, 40% tokens of the total bundles. These bundles have been used for contextualizing new information in the text, e.g., *in the context of (the),* (34) […] *here has only been one study on IPAs* ***in the context of*** *L2 learning*. (ES07), *with respect to the*, (35) *These two tasks were hypothesized to differ* ***with respect to the*** *cognitive demands they placed on learners* […], in *the light of* (36) *When considered* ***in the light of*** *the forum data*, [….] (ES 68)

Transition signals

Transition signals represent 6% types and 9% tokens of the total bundles. These bundles have been used to organize the text by linking parts of discourse, e.g., *on the other hand*, (37) This process could be theoretically proved based on Anderson's cognitive learning theory. On the *other hand, in the teacher-cantered approach* […], *in addition to the,* (38), *In addition to the students* […] (ES 63).

Structuring bundles

Structuring bundles represent 5% types and tokens of total bundles. These bundles have been used for referring to the study itself, e.g., *the results of the*, (39) *Participants* ***in the present study*** [….], *in this article we* (40) ***In this article, we*** *explore* […] (ES 66).

Resultative signals

Resultative signals represent only 3% types and tokens of total bundles. These bundles have been used for referring to the results of the study, e.g., (41) ***The results of the*** *use of epistemic stance device* […] (ES 61). But sometimes these bundles are also used to show the outcome of a process e.g., *as a result of* (42) *it emerged that the grade assigned* ***as a result of*** *the assessment process has a negative influence* […] (ES 60)

The main features of Text-oriented bundles in the expert corpus can be summarised as follows:

- Text-oriented bundles represent the majority (51% of the total bundles) in the expert corpus. As these bundles are aimed at text-organization, the very common use of these bundles suggests that the use of bundles in the expert corpus is text-centric (by contrast to research-centric or participant-centric).
- Text-oriented framing signals represent the majority of text-oriented bundles (36% of the total bundles), indicating that the contextualization of new information is the most important function of the text-oriented bundles in the expert corpus.

The next section presents the results and analysis of the participant-oriented bundles.

4.2.3.3 Participant-oriented bundles

Participant-oriented bundles are used for presenting writers' evaluation, and also to engage the readers. There are two sub-categories of these bundles: Stance features and Engagement bundles. These are the least common bundles, representing 8% types and 7% tokens of the total bundles in the expert corpus. Table 4.16 presents all the Participant-oriented bundles found in the expert corpus.

Table 4.16 ***Participant-oriented bundles in expert corpus***

Participant-oriented bundles	
Stance features	it is important to (note that), it is possible that, to be able to, there is a need, are more likely to, it can be argued (that), it is possible to, not be able to
Engagement features	can be seen in, can be seen as (that), can be used to, it should be noted

As can be seen in Table 4.16, very few participant-oriented bundles have been used in the expert corpus. Table 4.17 presents the distribution of Participant-oriented bundles in expert corpus.

Table 4.17 ***Frequency and % of Text-oriented bundles (types & tokens) in expert corpus***

Functions	**Types**	**%**	**Tokens**	**%**
Participant-oriented bundles	**8**	**8.42%**	**132**	**7.15%**
Stance features	5	5.26	63	3.41
Engagement features	3	3.15	69	3.74

In the following lines, the detailed analysis of theses bundles will be presented.

Stance features

Participant-oriented stance features represent 5% types and tokens of the total bundles. These bundles have been used to present the writers' evaluation e.g., *it is possible to* (43) *Despite the fact that this status update is partially ambiguous and written in a decidedly emphatic style that sometimes hinders comprehension, it is possible to reconstruct its argumentative nature*. (ES46), *are more likely to* (44) *First, the current research has only studied advanced learners, who are more likely to possess* [….] (ES 44)

Engagement features

Text-oriented engagement features represent 3% types and tokens of the total bundles These bundles are used to engage the readers in the text, e.g., *can be seen in*, (45) *The statistical data of the corpus compiled for the analysis* ***can be seen in*** *Table 1*. (ES45)*:*

The main features of Participant-oriented bundles in the expert corpus can be summarised as follows:

- The Participant-oriented bundles are the least common bundles representing 8% of the total bundles in the expert corpus.
- The Stance features have been used to present writers' evaluation and these bundles represent the majority (5% of the total bundles) of the Participant-oriented bundles in the expert corpus.

The analysis of the bundle functions in the expert corpus reveals that the bundles used for organizing text are the most common representing 51% of the total bundles in the expert corpus, i.e., text-oriented bundles. In particular, the contextualization of new information was given more importance as the majority (36% of the total bundles) of text-oriented bundles have been used to contextualize new information.

Following are the main features of bundle functions in the expert corpus:

- Bundle use in the expert corpus is primarily text-oriented, representing 51% types and tokens of the total bundles. These bundles have been used for linking parts of discourse and putting new information in context. Nearly half of the text-oriented bundles consist of framing bundles that are used to contextualize new information in the text.
- Research-oriented bundles represent 41% of the total bundles, indicating their important but secondary role in the expert corpus. These bundles have been used for referring to time, location, quantitative and qualitative aspects of the research, and for describing information related to research.
- Participant-oriented bundles are the least common bundles representing 8% of the total bundles. The Stance feature represented the majority (5% of the total bundles) of these bundles. These bundles have been used for presenting writers' evaluation.

4.2.4 Conclusion: expert corpus

After the analysis of structural and functional characteristics of lexical bundles in the expert corpus, the following is the summary of the main features of bundles in the expert corpus:

- The most important feature of lexical bundle use in the expert corpus is the reliance on Text-oriented bundles (57% types, 65% tokens) which writers use to make their text more coherent. The large majority of these are NP-based and PP-based bundles (71% types, 80% tokens of the total bundles), that present information with a reference point; specify the situation, place, and time; refer to the text; and compare and contrast.

- The expert writers have used two-thirds of the verb-based bundles to aim to present the information in a more objective way. The common use of VP-based bundles (anticipatory it, and passive verb) has been instrumental in the achievement of this objective.

4.3 Native student corpus

The native student corpus is based on the Masters dissertations written by native English students in the field of Applied Linguistics. The size of the corpus is 312,981 words with 20 texts. After the analysis of this corpus, 92 lexical bundles were extracted with 1553 tokens. The current section is based on the analysis of theses bundles. As a first step of analysis, the frequency and use of the top 20 bundles will be analysed. This will follow more detailed analysis of the structural and functional characteristics of all the bundles.

4.3.1 Top 20 bundles in the native student corpus

This section is based on the analysis of the 20 most frequent bundles in the native student corpus. Table 4.18 presents the list of 20 most frequent bundles in the native student corpus.

Table4.18 ***The 20 most frequent bundles in the native student corpus***

Rank	Lexical bundle	Tokens
1.	(that) the use of the	52
2.	on the other hand (the)	49
3.	as a result of (the)	37
4.	the results of the	36
5.	(is) that there is a	31

6.	as a function of	30
7.	it is important to	29
8.	the extent to which	29
9.	in line with the	28
10.	in relation to the	28
11.	as well as the	26
12.	it is possible that	25
13.	for the purposes of	24
14.	in the context of	24
15.	the way in which	24
16.	in the case of	23
17.	in terms of the	22
18.	the total number of	21
19.	in the field of	20
20.	in the same way	20

The majority of the bundles (15/20) in the list are the bundles used for the organization of the text, e.g., *on the other hand,* (46) ***On the other hand****, qualitative research focuses on the particular* […]. (NS 03)

There are only 3/20 used for purposes other than organizing the text. These bundles have been used for describing research, e.g., *the use of the* (47) ***the use of the*** *passive, impersonal constructions, nominalisations and so on* […]. (NS 08) and for presenting writers' viewpoint e.g., *it is important to, it is possible that*. (48) *Firstly,* ***it is possible that*** *the restructuring of interlanguage prosody exists along a continuum*. (NS 10)

Following are the important features of the 20 most frequent bundles in the native corpus:

- The majority (15/20) of top 20 most frequent bundles in the native student corpus are related to organization of the text.

The next section will present the findings of the analysis of the structural characteristics of the native student corpus.

4.3.2 Structural characteristics in the native student corpus

This section will present the analysis of native student corpus in the same way as was done for the expert corpus in the section 4.2. At first, the overall distribution of bundle structures will

be presented. This will follow the detailed analysis of bundles used in 4 main structural categories.

Table 4.19 presents the frequency of structural categories, in terms of types and tokens, both raw figures and percentages:

Table. 4.19 Frequency & % of structural categories (types & tokens) in the native student corpus

Structure	Types	%	Tokens	%
Noun-based bundles	**32**	**34.77**	**525**	**33.79**
Noun-based bundles with of-phrase fragment	25	27.17	409	26.33%
Noun-based bundles with other post modifier fragment	7	7.60	116	7.46
Preposition-based bundles	**30**	**32.6**	**583**	**37.53**
Preposition-based bundles with of-phrase fragment	17	18.47	331	21.31
Preposition-based bundles with other post modifier fragment	13	14.13	252	16.22
Verb-based bundles	**25**	**27.14**	**369**	**23.74**

Copula be + NP/Adj phrase	4	4.34	85	5.47
VP with active verb	0	0	0	0
Anticipatory it + Verb + (Adj phrase)	7	7.60	110	7.08
Passive verb + PP fragment	6	6.52	79	5.08
VP+ that clause fragment	2	2.17	31	1.99
Verb/adj + to clause fragment	6	6.52	64	4.12
Other structures	**5**	**5.43**	**76**	**4.89**
	5	5.43	76	4.89
Total	**92**	**100%**	**1553**	**100%**

The Noun based bundles are the most common bundles representing 35% types and tokens of the total bundles, whereas the Preposition bundles are nearly as frequent as the Noun-based bundles representing 33% types and tokens of the total bundles. The Verb-based bundles are over one quarter, representing 27% types and tokens of the total bundles.

To summarise:

- The Noun-based and the Preposition-based bundles represent equally common bundles in the native student corpus.

- Noun-based and Preposition-based bundles together make the most of bundles (68% types, 71% tokens of the total bundles) in the native student corpus.
- Verb-based bundles are over one quarter of the bundles.

To further explore and elaborate these findings, the next section will present detailed analysis of the Structural characteristics through examples from the native student corpus.

4.3.2.1 Noun- based bundles

This section will present the analysis of the Noun-based bundles classified into two sub-categories: Noun-based bundles with of-phrase fragment, Noun-based bundles with other post-modifier fragment. Table 4.20 presents all the Noun-based bundles found in the native student corpus:

Table 4.20 ***Noun-based bundles in the native student corpus***

Noun-based bundles	
Noun-based bundles with of-phrase fragment	that) the use of the, the results of the, the total number of, (of) the use of a, a large number of, the nature of the, a small number of, a wide range of, the validity of the, the end of the, the purpose of this, the acquisition of the,

	the focus of the, the scope of this (dissertation), the vast majority of, a wide variety of, the reliability of the, the size of the, an overview of the, the design of the, the, majority of the, a summary of the, results of the study, the findings of the, the role of the,
Noun-based bundles with other post-modifier fragment	the extent to which, the way in which, the fact that the, the results from the, the difference between the, the ways in which. the context in which

Table 4.21 presents the distribution of Noun-based bundles in the native students' corpus.

Table 4.21 ***Frequency & % of Noun-based bundles in the native student corpus***

Structrue	**Types**	**%**	**Tokens**	**%**
Noun-based bundles	**32**	**34.77**	**525**	**33.79**
Noun-based bundles with of-phrase fragment	25	27.17	409	26.33%
Noun-based bundles with other post modifier fragment	7	7.60	116	7.46

Noun-based bundles with of-phrase fragment

Noun-based bundles with of-phrase fragment represent 27% types and tokens of the total bundles. These bundles are three times more common than the Noun-based bundles with other post-modifier fragments. The majority (21% types and tokens of the total bundles) of Noun-based bundles with of-phrase fragment have been used for describing research.

For example, these bundles were used to quantify, e.g., *the total number of, a wide range of, the size of the* etc.: (50) *Frequency of turn-taking was also analyzed by dividing* ***the total number of*** *turns by the length of time taken* […] (NS 20), to denote qualities of something, e.g., *the use of the, the nature of the, the validity of the* (51) *They managed to input largely well-formatted data with minimal instruction in* ***the use of the*** *tool* […] (NS 15)

There were only 6% bundles of the total bundles, in this category used for organizing the text, for example, bundles used for referring to results section, e.g., *the results of the* (52) ***The results of the*** *preliminary surveys are best considered as a whole* […]. (NS 06)

So, Noun-based bundles with of-phrase fragment have been mainly used for describing research.

Regarding the use of bundle frames in the Noun-based bundles with of-phrase fragment, the research has shown that '*the ___ of the'* is the most productive Noun-based bundle frame. The same was found in the native student corpus. These Noun-based bundle frames represented 16% types and tokens of the total bundles in the native corpus.

Noun based bundles with other post-modifier fragment

These bundles represent 8% types and tokens of the total bundles, the majority (5% types and tokens of the total bundles) of which have been used for organizing the text e.g., *the way in which, the difference between the, the fact that the, the ways in which, and the context in which*

(53) *The study aims to present* ***the extent to which*** *SPR has been used.* (NS 17)

Following is the summary of Noun-based bundles in native students' corpus:

- Noun-based bundles are the most common bundles representing 35% types and tokens of the total bundles in the native student corpus. The majority (24% types and tokens of the total bundles) have been used for describing research.
- The majority of Noun-Phrase with of-phrase fragment (21% types and tokens of the total bundles) have been used for describing research

- The majority of the Noun-based bundles with other phrase fragment (5% types and tokens of the total bundles) have been used for organizing text.
- Noun-based bundles have been mainly used for describing research in the native student corpus.

The next section will present the results and analysis of the Preposition-based bundles.

4.3.2.2 Preposition-based bundles

Preposition-based bundles consist of two sub-categories: Preposition-based bundles with of-phrase fragment, Preposition-based bundles with other post-modifier fragment. Preposition-based bundles represent 33% types and tokens of total bundles. Table 4.22 displays all the Preposition-based bundles found in the native student corpus.

Table 4.22 ***Preposition-based bundles in the native student corpus***

Preposition-based bundles with of-phrase fragment	as a result of (the), as a function of, for the purposes of, in the context of, in the case of, in terms of the, in the field of, as part of a, on the basis of, as part of the, at the time of, at the university of, for each of the, in the use

	of, with the exception of, in the form of, in terms of their
Preposition-based bundles with other post-modifier fragment	on the other hand (the), in line with the, in relation to the, in the same way, of the present study, with regard to the, at the same time, in addition to the, in the present study, in other words the, as a starting point, in contrast to the, on the one hand

As can be seen from Table 4.23, the Preposition-based bundles with of-phrase fragment are more frequent (19% types 21% tokens of the total bundles) than the Preposition-based bundles with other post-modifier fragment (14% types 16% tokens of the total bundles).

Table 4.23 Frequency & % of Preposition-based bundles in the native student corpus

Structure	Types	%	Tokens	%
Preposition-based bundles	**30**	**32.6**	**583**	**37.53**
Preposition-based bundles with of-phrase fragment	17	18.47	331	21.31
Preposition-based bundles with other post modifier fragment	13	14.13	252	16.22

Preposition-based bundles with of-phrase fragment

Preposition-based bundles with of-phrase fragment represent 18% types and tokens of the total bundles. The majority of these bundles (11% types and tokens of the total bundles) have been used for organizing text. e.g., *as a result of,* (54) *However, **as a result of** the multi-channel recording, the whole process of observation* [….], *in the context of* (55) ***In the context of** ongoing research*, […]

The rest of these bundles (8% types of the total bundles) have been used for describing research, e.g., *in the field of* (56) ***in the field of** second-language acquisition* [….], *at the university of* etc (57) *The student population **at the University of** Essex* (UOE) […]

As for the Preposition-based bundles with of-phrase fragment frames, 8% of these bundles have been used in the bundle frames: '*in the ___ of*' and only 2% bundles were used with the frame *'at the ___ of'*.

Preposition-based bundles with other post-modifier fragment

Preposition-based bundles with other post-modifier fragment represent 14.13% types and tokens of the total bundles. All of these bundles have been used for organizing the text. e.g., *on the other hand,* (58) ***On the other hand****, saccade planning in Chinese may require a degree of lexical processing* […] (NS 06), *at the same time,* (59) ***At the same time****, we want to see ourselves as non-prejudiced and caring* […], *in line with the* (60) *This is not* ***in line with the*** *predictions of the PTH*. (NS 11)

Following is the summary of Preposition-based bundles in native students' corpus:

- Preposition-based bundles represent 33% types and 38% tokens of the total bundles, the majority which (25% types and tokens of the total bundles) have been used for organizing the text.
- The majority (11% of the total bundles) of Preposition-based bundles with of-phrase fragment, and all the (14% of the total bundles) of Preposition-based bundles with other post-modifier fragment have been used for organizing the text.
- So, Preposition-based bundles have been mainly used for organizing text in the native student corpus.

In the next section, the results of verb-based bundles will be presented.

4.3.2.3 Verb-based bundles

There are 6 sub-categories of the Verb-based bundles. These sub-categories are: Copula be + Noun/adjective phrase, Verb-based bundles with active verb, Anticipatory it + verb + (adjective phrase), Passive verb + Prepositional fragment, Verb-based bundles with to-clause fragment, and Verb-based bundles with that-clause fragment. All the Verb-based bundles found in the native student corpus have been categorized and displayed in Table 4.24.

Table 4.24 ***Verb-based bundles in the native student corpus***

Verb-based bundles	
Copula be + noun /Adjective phrase	is one of the, there is also a, this is not the, , is not the case
Verb-based bundle with Active Verb	
Anticipatory it + verb + (Adjective phrase)	it is important to, it is possible that, it is possible to, it is worth noting, it is difficult to, it is interesting to, it should be noted
Passive verb + prepositional phrase fragment	as can be seen (from/in), can be found in (the), can be used to, participants were asked to, (to) be included in the, can be used as

Verb-based bundles with that-clause Fragment	(is) that there is a, is that it is
Verb-based bundles with to-clause Fragment	to be able to, in order to answer (the), (be/are) more likely to be, in order to provide, is likely to be, are likely to be

As can be seen from Table 4.25, 'anticipatory it+ noun/adj phrase' bundles are the most common (8% of types and 7% of tokens) bundles in the native student corpus, whereas no Verb-based bundle with active verb has been used in the native student corpus (see Table 4.25).

Table 4.25 ***Frequency and % of Verb-based bundles in the native student corpus***

Structure	Types	%	Tokens	%
Verb-based bundles	**25**	**27.14**	**369**	**23.74**
Copula be + NP/Adj phrase	4	7.60	85	5.47
VP with active verb	0	0	0	0
Anticipatory it + Verb + (Adj phrase)	7	7.60	110	7.08
Passive verb + PP fragment	6	6.52	79	5.08
VP+ that clause fragment	2	1.08	31	1.99
Verb/adj + to clause fragment	6	4.34	64	4.12

The detailed analysis of all the sub-categories of the Verb-based bundles will be presented below:

Anticipatory it + verb + (Adjective phrase)

The Anticipatory it + verb + (Adjective phrase) bundles represent 8% types and tokens of the total bundles. These bundles have been used for engaging the reader e.g., *it is important to, it is worth noting, it is interesting to, it should be noted,* (61) ***It should be noted*** *that while each of the examples above was incorrect in its judgement* […]. (NS 10), and to present authors' evaluation, e.g., *it is possible that, it is possible to, it is difficult to* (62) *Conjecturally,* ***it is possible that*** *L2 users could differ* […]. (NS 13)

Passive verb + Prepositional phrase fragment

The Verb-based bundles with passive verb represent 7% types and tokens. the majority (4% of the total bundles) of these were used for describing research e.g., *can be used to, participants were asked to, can be included in, can be used as*. (63) *On each occasion,* ***participants were asked to*** *complete* […]. (NS 15). The other 3% of the total bundles were used for referring to some section or table of the text e.g., *as can be seen, can be found in* etc.

Verb-based bundles with to-clause fragment

The Verb-based bundles with to-clause fragment represent 4% types and tokens of the total bundles. Almost all of these bundles (3% of the total bundles) used for presenting writers' evaluation, e.g., (be/are) *more likely to be, is likely to be, are likely to be*. (64) *Therefore, online measures are likely to be more accurate* […]. (NS 08)

Copula be + noun/adjective phrase

The Copula be + noun/adjective phrase bundles represent 4% types and tokens of the total bundles in the native student corpus. All of these bundles have been used for describing research, e.g., *there is also a, this is not the, is not the case* (65) *There is also a noticeable difference* […]. (NS 16)

Verb-based bundles with that-clause fragment

The Verb-based bundles with that-clause fragment represent only 1% of the total bundles. These bundles have been used for describing research, e.g., *that there is a, is that it is*.

Following is the summary of Verb-based bundles in native students' corpus:

- Verb-based bundles represent 27% types and tokens of the total bundles in the native student' corpus.

- Half of the Verb-based bundles (13% of the total bundles) have been used for describing research, 9% (of the total bundles) of these bundles have been used for presenting writers' evaluation, and 5% (of the total bundles) the bundles were used for engaging the reader.

- All the Copula be+ verb/ adj phrase bundles (4% of the total bundles), Verb-based bundles with that-clause fragment (1% of the total bundles), the Verb-based bundles with to-clause fragment (3% of the total bundles)., and the majority of Verb-based bundles with passive fragment (4% of the total bundles) have been used for describing research

- All the Anticipatory it+ Verb bundles (8% types and tokens of the total bundles) have been used for presenting writers' evaluation and to engage the readers.

- The verb-based bundles have been mainly used for describing research and presenting writers' evaluation.

4.3.2.4 Other Structures

Other structures represent only 5% types and tokens of the total bundles in the native student corpus. Table 4.26 presents all the other structure bundles found in the native student corpus:

Table 4.26 ***Other structures in the native student corpus***

Other structures	
Other Structure	as well as the, one of the most, whether or not the, each of the three, and the use of

Table 4.27 presents the frequency and proportion of Other structures in native students' corpus.

Table 4.27 ***Frequency and % of Other structures in the native student corpus***

Structure	**Types**	**%**	**Tokens**	**%**
Other structures	**5**	**5.43**	**76**	**4.89**
	5	5.43	76	4.89

Other structures represent 5% of bundles in the native student corpus, half of which have been used for organizing text, e.g., *as well as the* (66) *Information Seeking Anxiety Scale (ISAS)* ***as well as the*** *mean for the subscales* […]. (NS 14), whereas half the bundles were used for describing research, e.g., *one of the most* (67) ***One of the most*** *important factors* […]. (NS 04) etc.,

The important features of the structural features of bundles in the native student corpus can be summarized as follows:

- Noun-based bundles are the most common bundles representing 35% types and tokens of the total bundles in the native student corpus. The majority (24% of the total bundles) of these bundles have been used for describing research.
- The Preposition-based bundles are nearly as common as the Noun-based bundles representing 33% types and tokens of the total bundles. The majority of these bundles (25% types and tokens of the total bundles) have been used for organising text.
- The Verb-based bundles represent 27% types and tokens of the total bundles, half of these bundles (13% types and tokens of the total bundles) have been used for describing research, whereas the 9% (of the total bundles) have been used for presenting writers' evaluation and engaging the readers.

So, the Noun-based bundles, and the Verb-based bundles have been mainly used for describing research, whereas the Preposition-based bundles have been mainly used for organizing text in the native corpus.

4.3.3 Bundle functions in the native corpus

In this section, I will present the results, and the analysis of bundle functions found in the native student corpus. For the analysis of bundle functions, I will follow the same pattern as was followed in the analysis of bundle functions of the expert corpus. At first, I will present the analysis of the overall distribution of bundle functions in the three main categories: Research-oriented bundles, Text-oriented bundles, and participant-oriented bundles. This will follow the detailed analysis of all bundle functions in the sub-categories.

Table 4.28 presents the frequency distribution of bundle functions, in terms of types and tokens, in the native student corpus. The raw frequencies, percentages, and relative frequencies have been presented in the table.

Table 4.28 ***Frequency & % of functional categories in the native student corpus***

Functions	**Types**	**%**	**Tokens**		**%**
			ABS	**REL**	
Research-oriented bundles	**43**	**46.73%**	**701**	**191.68**	**45.13%**
Location	3	3.26	45	14.37	2.89
Procedure	16	17.39	291	51.44	18.73
Quantification	13	14.13	207	58.78	13.32

Description	11	11.95	158	67.09	10.17
Text-oriented bundles	**37**	**40.20%**	**673**	**223.31**	**43.33%**
Transition signals	9	9.78	164	52.39	10.56
Resultative signals	5	5.43	108	34.50	6.95
Structuring signals	5	5.43	74	10.86	2.18
Framing signals	18	19.56	327	125.56	25.30
Participant-oriented bundles	**12**	**13.04%**	**179**	**81.14**	**11.52%**
Stance features	6	6.52	84	65.17	5.40
Engagement features	6	6.52	95	15.97	6.11
Total	**92**	**100%**	**1553**	**496.13**	**100%**

Research-oriented bundles are the most common bundles representing 47% types and tokens of the total bundles, indicating that the native students assign importance to describing research. The text-oriented bundles represent the 40% of the total bundle types and tokens. The participant-oriented bundles are the least common bundles accounting 14% types and tokens of the total bundles.

In the next section, the detailed analysis of the research-oriented bundles will be presented.

4.3.3.1 Research-oriented bundles

Research-oriented bundles represent 47% types and tokens of the total bundles. These bundles are used for describing research. There are four sub-categories of these bundles: location bundles, procedure bundles, quantification bundles, and description bundles. Table 4.29 presents all the research-oriented bundles found in the native student corpus.

Table 4.29 ***Research-oriented bundles in the native student corpus***

Research-oriented bundles	
Location	at the time of, at the university of, the end of the
Procedure	as a function of, in the field of, as part of a, as part of the, in the use of, the purpose of this, participants were asked to, the acquisition of the, the focus of the, the scope of this (dissertation)
Quantification	the extent to which, the total number of, a large number of, a small number of, a wide range of, is one of the, one of the most, for

	each of the, each of the three, the vast majority of, a wide variety of
Description	(that) the use of the (of), the use of a, the nature of the, the validity of the, in the form of, the reliability of the, the size of the, an overview of the, is that it is, the design of the, the majority of the, a summary of the, the role of the

The detailed analysis of the sub-categories will now be presented.

Table 4.30 presents the distribution of Research-oriented bundles in native students' corpus.

Table 4.30 ***Frequency & % of Research-oriented bundles in the native student corpus***

Structure	Types	%	Tokens	%
Research-oriented bundles	**43**	**46.73%**	**701**	**45.13%**
Location	3	3.26	45	2.89
Procedure	16	17.39	291	18.73
Quantification	13	14.13	207	13.32
Description	11	11.95	158	10.17

Procedure bundles

Procedure bundles represent 17% types and tokens of the total bundles in the native students' corpus. These bundles have been used to describe the procedures of research. For example, *the use of the* (69) *Therefore,* ***the use of the*** *validated Information Seeking Anxiety Scale questionnaire* […] (NS 06), *participants were asked to* (70) ***participants were asked to*** *complete an initial interview form* […]. (NS 12)

Quantification bundles

Quantification bundles represent 14% types and tokens of the total bundles. These bundles have been used for quantifying, e.g., *a small number of,* (71) *The survey was piloted by* ***a small number of*** *close teaching colleagues* [….] (NS 09), *one of the most,* (72) ***One of the most*** *common uses of SPR is in the investigation* […]. (NS 01)

Description bundles

Description bundles represent 12% types and tokens of the total bundles. These bundles have been used for describing qualitative features of research, or giving descriptions of different features of research, e.g., *the nature of the,* (73) *This difference could well be attributed to* ***the***

nature of the *sample group in this research* […]. (NS 01), *the design of the*, (74) ***The design of the*** *study meant that the teachers self-reported on their WCF practices*. (NS 11)

Location bundles

Location bundles are the least used research-oriented bundles representing only 3% of the total bundles. These bundles have been used to show place and time, e.g., *at the time of*, *at the university of* (74) […] *all postgraduate students in the Language and Linguistics Department at the University of Essex*. (NS 07)

Following is the summary of Research-oriented bundles used in native students' corpus:

- Research-oriented bundles are the most common bundles representing 47% types and tokens of the total bundles indicating the importance of research-oriented bundles in the native student corpus.

4.3.3.2 Text-oriented bundles

Text-oriented bundles are used for organizing the text. There are four sub-categories of these bundles: transition signals, resultative signals, structuring signals, and framing signals**.** In the native corpus, text-oriented bundles represent 40% types and 45% of the total bundles. Table 4.31 presents all the text-oriented bundles found in the native student corpus:

Table 4.31 ***Text-oriented bundles in the native student corpus***

Text-oriented bundles	
Transition signals	on the other hand (the), as well as the, at the same time, in addition to the, there is also a, and the use of, in other words the, in contrast to the, on the one hand
Resultative signals	as a result of (the), the results of the, the results from the, results of the study, the findings of the
Structuring signals	of the present study, in the present study
Framing signals	(is) that there is a, in line with the, in relation to the, for the purposes of, in the context of, the way in which, in the case of, in terms of the, in the same way, in order to answer (the), on the basis of, with regard to the, the fact that the, with the exception of, whether or not the, in terms of their, in order to provide, as a starting point, the difference

	between the, the ways in which, the context in which

Table 4.32 presents the distribution of Text-oriented bundles in native students' corpus

Table 4.32 ***Frequency & % of Text-oriented bundles in the native student corpus***

Structure	Types	%	Tokens	%
Text-oriented bundles	**37**	**40.20%**	**673**	**43.33%**
Transition signals	9	9.78	164	10.56
Resultative signals	5	5.43	108	6.95
Structuring signals	5	5.43	74	2.18
Framing signals	18	19.56	327	25.30

Framing signals

Text-oriented framing bundles are the most frequent text-oriented bundles representing 20% types and tokens of the total bundles in native students' corpus. These bundles have been used for contextualizing new information in discourse, e.g., *in line with the* (75) *Privacy and*

confidentiality issues were addressed ***in line with the*** *University's guidelines*. (NS 18), *the way in which,* (76) *I will be examining* ***the ways in which*** *conflicts around acceptability and accent are reproduced*. (NS 18)

Transition signals

Text-oriented transition signals represent 10% types and tokens of the total bundles in the native students' corpus. These bundles have been used for linking sentences through showing contrast, addition, and paraphrasing the information, e.g., *in contrast to the* (77) **I*n contrast to the*** *aforementioned studies,* […]. (NS 14), *in other words the*, (78) ***In other words, the*** *regions defined as spill over and critical may be the same*. (NS 09)

Resultative signals

Resultative signals represent 5% types and tokens of the total bundles. These bundles have been used for two purposes, i.e., to refer to the results of the study, and to show the outcome of some process in the study. For example, the results of the, (79) ***the results of the*** *analysis undertaken in this study* […]. (NS 10), *as a result of* (80) *This theory assumes that the greater number of long fixations among younger readers occurs* ***as a result of*** *more frequent cognitive intervention* […]. (NS 17)

Structuring signals

Structuring signals are the least common bundles representing only 2% of the total bundles in the native corpus. These bundles have been used for referring to the study, e.g., *in the present study* etc.

Following is the summary of Text-oriented bundles used in native students' corpus:

- The text-oriented bundles represent 40% types and tokens of the total bundles in the native corpus.
- Half of the text-oriented bundles (20% types and tokens of the total bundles) represent framing signals that have been used for contextualizing new information in the text, indicating that Text-oriented bundles have been mainly used for contextualizing new information in the native corpus.

4.3.3.3 Participant-oriented bundles

Participant-oriented bundles are the least common bundles representing 13% types and tokens of the total bundles). These bundles are used for presenting writers' evaluation, and for engaging the readers. Table 4.33 presents all the Participant-oriented bundles found in the native student corpus:

Table 4.33 ***Frequency & % of Text-oriented bundles in the native student corpus***

Participant-oriented bundles	
Stance features	it is important to, it is possible that, to be able to, can be found in (the), can be used to, (be/are) more likely to be, is likely to be, it is possible to, this is not the, are likely to be, it is difficult to, (to) be included in the, is not the case
Engagement features	as can be seen (from/in), it is worth noting, can be used as, it should be noted, it is interesting to

Table 4.34 presents the distribution of Participant-oriented bundles in native students' corpus

Table 4.34 ***Frequency & % of Research-oriented bundles in the native student corpus***

Structure	**Types**	**%**	**Tokens**	**%**
Participant-oriented bundles	**12**	**13.04%**	**179**	**11.52%**
Stance features	6	6.52	84	5.40
Engagement features	6	6.52	95	6.11

Stance features

Stance features represent 7% of types and tokens of the total bundles. These bundles have been used for presenting the authors' viewpoint, e.g., *it is possible that* (81) ***It is possible that*** *the type of questions asked may again be relevant here* […]. (NS 13), *is likely to* be (82) *This* ***is likely to be*** *true for all research in this field* […]. (NS 08)

Engagement features

Engagement features represent 7% of the total bundles. They have been used to for engaging with the reader, e.g., *it is important to,* (83) *It is important to* remember […]. (NS 11), *as can be seem from,* (84) ***As can be seen*** *from Tables 4.5.1*[…]. (NS 12)

Following is the summary of Participant-oriented bundles used in native students' corpus:

- Participant-oriented bundles are used for presenting writers' opinion and for engaging the readers, but they are not very common in the native student corpus. These bundles represent just over 10% types and tokens of the total bundles.

4.3.3.4 Summary of the bundle functions in the native student corpus

Following are the important features of the bundle functions in the native student corpus:

- The research-oriented bundles, used for describing research, are the most common bundles representing 47% of bundle types and tokens in the native student corpus.
- The text-oriented bundles, used for organizing text, represent 40% types and tokens of the total bundles. The text-oriented framing signals represent half of text-oriented bundles (23% types and tokens of the total bundles). This indicates that the text-oriented bundles have been mainly used for contextualizing new information in the native student corpus.
- The Participant-oriented bundles are the least common bundles representing 14% of the total bundles. The stance features and the engagement features represented equally common bundles.

4.4 Non-native student corpus

This section presents the analysis of the bundles found in the non-native student corpus. Like the previous sections on expert corpus and native student corpus, this section will begin with the analysis of the 20 most frequent bundles in the non-native student corpus. This will follow the analysis of the structural and functional characteristics of bundles in the non-native student corpus.

4.4.1 The 20 most frequent bundles in the non-native student corpus

This section is based on the analysis of the 20 most frequent bundles in the non-native student corpus. Table 4.35 presents the 20 most frequent bundles in the non-native student corpus. These bundles are listed in order of frequency.

Table 4.35 ***Top 20 lexical bundles in the non-native student corpus***

Rank	Lexical bundle	Tokens
1.	agreed with the statement (that)	459
2.	of the respondents agreed	180
3.	with the statement and	167
4.	on the other hand (the)	168
5.	(is/are/were) of the view that	150
6.	majority of the respondents	146

7.	with the help of	127
8.	(in) the analysis of the (data)	123
9.	of the present study (the)	114
10.	on the basis of (the)	112
11.	that of the respondents	51
12.	(but) at the same time (the)	97
13.	in the use of	96
14.	that most of the	96
15.	the results of the (study)	89
16.	in the form of	85
17.	to find out the	84
18.	in the process of	83
19.	(and) the use of the	53
20.	the findings of the	75

The majority of the bundles (13/20) in the list have been used for describing research e.g., *agreed with the statement, of the respondents agreed, majority of the respondents, the use of the* etc. (85) *The students responded differently but 48.0% students* ***agreed with the statement*** *that their attitude is the main hurdle in learning English language*. (NNS 02), *the use of the,*

(86) *As the words like wicket, lift, battery and capsule are the words which have no equivalents in Urdu while the words like election, ban, nursing, and break do have their equivalent words in Urdu, but these words are not in* ***the use of the*** *public in common conversation frequently*. (NNS 05)

There are 6/20 bundles that have been used for the purpose of organising the text e.g., *at the same time*, (87) ***At the same time*** *they can positively utilize these technologies in the betterment of language learning of their pupils*. (NNS 10).

There is only bundle in the list that has been used to present the writers' evaluation e.g., *of the view that*. (88) *As Duff is* ***of the view that*** […]. (NNS 18)

It is worth noting that the tokens of the first 10 (4% types of the total bundles) represent 25% tokens of the total bundles in the non-native corpus. This indicates that the non-native students have used a very small number of bundles highly frequently.

It is also worth mentioning that the bundle '*agreed with the statement'* is so frequent in the non-native student corpus because in the majority of the non-native student texts (11/19) questionnaires have been used for data collection. As the non-native students repeatedly

referred to results of the questionnaires in the results section, the use of the bundle, *agreed with the statement*, has occurred highly frequently. The other highly frequent bundles, e.g., *of the respondents agreed, with the statement and,* are also highly frequent for the same reason.

Following are the main features of top 20 bundles used in non-native students' corpus:

- Non-native students appear to rely on very few bundles which they use very frequently in their writing i.e., the top 4% of bundle types represent 26% of the total bundle tokens in the non-native student corpus.
- Noun-native students appear to use bundles more frequently for describing research than organizing the text. As has been shown in Table 4.35, the majority of the bundles (13/20) have been used for purposes related to describing research.

To further elaborate the use of bundles in the non-native corpus, the next section will present the analysis of the structural characteristics of the bundles.

4.4.2 Structural characteristics of bundles in the non-native student corpus

This section is based on the results of the structural features of bundles in non-native students' corpus. The results will be presented following the same pattern as was following for the other two corpora. At first, the general distribution of bundle structures will be presented. Then, a

detailed analysis of all the sub-categories of bundle structures will follow. Table 4.36 presents the distribution of bundle types and tokens in the non-native students' corpus.

Table 4.36 ***Frequency & % of structural categories in the non-native student corpus***

Structure	Types	%	Tokens	%
Noun-based bundles	**50**	**20.65%**	**1016**	**15.78%**
Noun-based bundles with of-phrase fragment	39	16.11	859	12.78%
Noun-based bundles with other post modifier fragment	11	4.54	202	3.00
Preposition-based bundles	**65**	**26.85%**	**2417**	**35.95%**
Preposition-based bundles with of-phrase fragment	37	15.28	1407	20.93
Preposition-based bundles with other post modifier fragment	28	11.57	1010	15.02
Verb-based bundles	**91**	**37.58%**	**2477**	**36.83%**
Copula be + NP/Adj phrase	21	8.67	355	5.28
VP with active verb	11	4.54	755	11.23
Anticipatory it + VP/Adj phrase	10	4.13	216	3.21

Passive verb + PP fragment	13	5.37	204	3.03
VP+ that clause fragment	17	7.02	605	9.00
Verb/adj + to clause fragment	19	7.85	342	5.08
Other structures	**36**	**14.87%**	**765**	**11.83%**
	36	14.87	765	11.83
Total	**242**	**100%**	**6720**	**100%**

The Verb-based bundles, also known as the clausal bundles, are the most common bundles representing 38% types and tokens of the total bundles in the non-native corpus.

Preposition-based bundles represent 27% types, and 36% tokens of the total bundles represent, whereas the Noun-based bundles 21% types and 16% tokens of the total bundles.

The main feature of the Structural distribution of lexical bundles in non-native students' corpus are as follows:

- The verb-based bundles are the most common bundles representing over one third types and tokens of the total bundles.
- The Preposition-based bundles represent one quarter of the total bundles.

- The Noun-based bundles are less common, representing 21% types and 15% tokens of the total bundles.

The detailed analysis of the bundle structures will be presented in the next section.

4.4.2.1 Noun-based bundles

All the Noun-based bundles found in the non-native corpus have been listed in Table 4.37.

Table 4.37 ***Noun-based bundles in the non-native student corpus***

Noun-based bundles	
Noun-based bundles with of-phrase fragment	(in) the analysis of the (data), the results of the (study), (and) the use of the, the findings of the, the end of the, the attitude of the, a large number of, the majority of the, a part of the, a lot of time, the role of the, an essential part of, the need of the, the total number of, a wide range of, the population of the, the purpose of this, the importance of the, the light of the, the performance of the, the significance of the, an important part of, the purpose of the, the responses of the, an integral part of, the aim of the, the beginning of the, the result of the,

	the context of the, the results of this, the meanings of the, the needs of the, the status of a, a small number of, the characteristics of the, the review of the, the sample of the, the status of the, the topic of the
Noun-based bundles with other post-modifier fragment	the view that the, a vital role in, (plays)an important role in, (the) the ways in which, the degree to which, the extent to which, the fact that the, the context in which, the answer to the (research questions), the relationship between the, the study the present,

As can be seen from Table 4.37 that the Noun-based bundles with of-phrase fragment are thrice as common as the Noun-based bundles with other post-modifier fragment. Table 4.38 presents the distribution of Noun-based bundles in non-native students' corpus.

Table 4.38 ***Frequency & % of Noun-based bundles in the non-native student corpus***

Structure	Types	%	Tokens	%
Noun-based bundles	**50**	**20.65%**	**1016**	**15.78%**
Noun-based bundles with of-phrase fragment	39	16.11	859	12.78%
Noun-based bundles with other post modifier fragment	11	4.54	202	3.00

Noun-based bundles with of-phrase fragment

Noun-based bundles with of-phrase fragment represent 16% types and tokens of the total bundles in the non-native corpus. The majority (14% of the total bundles) of these bundles have been used for describing research, such as, to denote the quality of something, *the use of the, the attitude of the, the population of the, the importance of the* etc., (89) The attitude of the learners influences their TL use in any social environment. (NNS 06), to show time e.g., *the end of the, the beginning of the*, (90) In the beginning of the college session […] (NNS 11) to quantify e.g., *a large number of, the majority of the,* (91) *A large number of the teachers* […] (NNS 19).

The rest (2% of the total bundles) of the Noun-based bundles with of-phrase fragment have been used for organizing text, such as, to show results of the study e.g., *the findings of the, the result of the, the results of this,* (92) ***The results of the*** *post-test* […] (NNS 04).

The NP bundle frame *'the ___ of the'* e.g., *the use of the, the attitude of the, the meaning of the* etc. has been the most frequent bundle frame in the non-native corpus. Majority of the bundles with of-phrase fragment (13% types and tokens of the total bundles) have been used in this frame.

Noun-based bundles with other post modifier fragment

Noun- based bundles with other post-modifier fragments represent 5 % types and tokens of the total bundles. The majority (3% of the total bundles) of these bundles have been used for describing research, such as, to show measurement, e.g., *the degree to which, the extent to which* (93) *The primary objective is to explore* ***the extent to which*** *bilingualism hinders* […] (NNS 01), to refer to the research, e.g., *the answer to the* etc. The rest (2% of the total bundles) have been used for organizing text, such as, e.g., *the way in which, the context in which,* etc.

Following are the main features of Noun-based bundles in non-native students' corpus:

- Noun-based bundles are the least common and represent 21% (of the total bundles) of bundle types and tokens of the total bundles. The majority of these bundles (17% types and tokens of the total bundles) have been used for describing research,
- The majority of Noun-based bundles with of-phrase fragment (14% types and tokens of the total bundles), and almost all of Noun-based bundles with other post-modifier fragment (3% types and tokens of the total bundles) have been used for organizing the text.

- The Noun-based bundles have been mainly used for describing research in the non-native corpus.

In the next section, the results and analysis of the Preposition-based bundles will be presented.

4.4.2.2 Preposition-based bundles

Preposition-based bundles start with a preposition. These bundles have two sub-categories: Preposition-based bundles with of-phrase fragment, and Preposition-based bundles with post-modifier fragment. These bundles represent 27% types and 36% tokens of the total bundles in the non-native student corpus. Table 4.39 presents all the Preposition-based bundles found in the non-native corpus.

Table 4.39 ***Preposition-based bundles in the non-native student corpus***

Preposition-based bundles with of-phrase fragment	with the help of, on the basis of (the), in the use of, in the form of, in the process of, in the field of, by the use of, (significance) of the study the, about the use of, at the end of, in the light of, (but) with the passage of (time), for the development of, of the use of, for the purpose of, regarding the use of, in the context of, on the use of, to the use of, on the

	part of (the), towards the use of, for the sake of, as one of the, as a result of, at the time of, of the importance of, for the use of, in the area of, in spite of the, in the development of, for the collection of, in this type of, for the selection of, by majority of the, in front of the, in the fields of, in the presence of
Preposition-based bundles with other post-modifier fragment	with the statement and, on the other hand (the), of the present study (the), (but) at the same time (the), in the same way, in this way the, as compared to the, in this chapter the, in the present study (are), for the present study, in such a way, in this study the, in a better way, in other words the, on the one hand, in accordance with the, in the study the, for the present research (is), in this regard the, in a way that, (and)on the other side, as a tool to, in the present research, in this section the, between the use of, from the perspective of, in the study as, of the findings of

Table 4.40 presents the distribution of Preposition-based bundles in non-native students' corpus.

Table 4.40 ***Frequency & % of Preposition-based bundles in the non-native student corpus***

Structure	Types	%	Tokens	%
Preposition-based bundles	**65**	**26.85%**	**2417**	**35.95%**
Preposition-based bundles with of-phrase fragment	37	15.28	1407	20.93
Preposition-based bundles with other post modifier fragment	28	11.57	1010	15.02

Preposition-based bundles with of-phrase fragment

Preposition-based bundles with of-phrase fragment represent 15% types and 21% tokens of the total bundles. The majority of these bundles (11% of the total bundles) have been used for describing research. For example, they have been used for referring to procedures, e.g., *in the use of, in the form of, with the help of,* (94) *Now* ***with the help of*** *e-dictionaries, learners thought flow is no longer disrupted as much as before*. [through]. (NNS 09), for indicating time, e.g., *at the time of,* (95) ***At the time of*** *creation of Pakistan* […]. (NNS 01), and to

quantify, e.g., *by majority of the*, (96*) The scholars have investigated the interactions and position between languages spoken* ***by majority of the*** *population* […] (NNS 05)

The rest (5% of the total bundles) of the Preposition-based bundles with of-phrase fragment have been used for organization of the text. For example, these bundles have been used to contextualize new information in the text, e.g., *in the light of, on the basis of*, (97), to justify, e.g., *for the sake of*, (98) […] ***for the sake of*** *survival and growth in that community* (NNS 03), to present the outcome, e.g., *as a result of*, (100) ***As a result of*** the study […] (NNS 09).

The most common Preposition-based bundles with of-phrase fragment frame, *'in the ___ of '* represent 3% types and tokens of the total bundles.

Preposition phrase with post-modifier fragment

Preposition phrase with post-modifier fragment represent 12% types and 15% tokens of the total bundles. Almost all (11% of the total bundles) these bundles have been used for organizing the text, such as, to show contrast, e.g., *on the other side*, (99), …. to refer to the section of study, e.g., *in the present study*, (100) *The data* ***in the present study*** *to contextualize new information*. (NNS 06), e.g., *in this way the*, (101) ***In this way, the*** *people who have positive beliefs to compare*. (NNS 04), to compare, e.g., *as compared to the*, (102) […] *the*

*female language learners are less **anxious as compared to the** male English language learners*. (NNS 07), to paraphrase, e.g., (103) ***In other words, the** educational development* […] (NNS 08).

Following are the main features of Preposition-based bundles in non-native students' corpus:

- Preposition-based bundles represent 27% types and tokens of total bundles in the non-native corpus. Half of these bundles (15% of the total bundles) have been used for organizing the text, and most of the other half (12% types and tokens of the total bundles) have been used for describing research.
- Almost all the Preposition-based bundles with of-phrase fragment (11% types and tokens of the total bundles) have been used for describing research
- All the Preposition-based bundles with other post-modifier fragment (11% types and tokens of the total bundles) have been used for organizing text.

In the next section, the results and analysis of the Verb-based bundles will be presented.

4.4.2.3 Verb-based bundles

The Verb-based bundles contain verbs. There are six sub-categories of these bundles. These sub-categories are ,Copula be + Noun/adjective phrase, Verb-based bundles with active verb, Anticipatory it + verb + (adjective phrase), Passive verb + Prepositional fragment, Verb-based bundles with to-clause fragment, and Verb-based bundles with that-clause fragment. These are the most common bundles representing 38% types and tokens of the total bundles in the non-native corpus. Table 4.41 presents all the verb-based bundles found in the non-native student corpus.

Table 4.41 Verb-based bundles in the non-native student corpus

Verb-based bundles	
Copula be + noun /Adjective phrase	is one of the (main), of the respondents were, the present study is, there is a need (to), is the case with, of this study was (to), has been used as, (objective/aim) of the study was, is the result of, be helpful for the, same is the case, is the use of, of the study is, the present study was, he is of the, is evident from the, is the product of , of this research is, of this

	research was, are some of the, in which they are
Verb-based bundle with Active Verb	agreed with the statement (that), of the respondents agreed, the study will be, and they do not (have), become a part of, play a vital role, chapter deals with the, find it difficult to, this chapter deals with, discussed in detail in, used in this study
Anticipatory it + verb + (Adjective phrase)	it has been observed (that), it can be seen, it was observed that (the), it is important to, it refers to the, it is believed that it is necessary to, it is also a, it is needed to, when it comes to
Passive verb + prepositional phrase fragment	is used in the, can be used to, can be seen in (the), is based on the, is used as a, were found to be, can be used in, is related to the, can be used for, are found to be, were selected for the, are based on the, can be divided into

Verb-based bundles with that-clause Fragment	(is/are/were) of the view that, that of the respondents, that most of the, (is) of the opinion that, that the use of, that there is a, of the fact that, such a way that, that is why the, that there is no, that they do not (have/any), that the majority of that they are not, that is why it, that it is the, to the fact that, that there are some
Verb-based bundles with to-clause Fragment	to find out the, to know about the, to take part in, to participate in the, in order to make, in order to understand, in order to get study was conducted to, this research was to, to be able to, when they try to, in order to find (out), of the study to, the researcher tried to, to talk about the, participants were asked to, the study was to, to collect the data, to interact with the

Table 4.42 presents the distribution of Verb-based bundles in non-native students' corpus.

Table 4.42 Frequency & % of Verb-based bundles in the non-native student corpus

Structure	Types	%	Tokens	%
Verb-based bundles	**91**	**37.58%**	**2477**	**36.83%**
Copula be + noun/adj. phrase	21	8.67	355	5.28
Verb-based bundles with active verb	11	4.54	755	11.23
Anticipatory it + verb/adj. phrase	10	4.13	216	3.21
Passive verb + PP fragment	13	5.37	204	3.03
Verb-based bundles with that clause fragment	17	7.02	605	9.00
Verb-based bundles with to clause fragment	19	7.85	342	5.08

Copula be+ noun/adjective phrase

These bundles represent 9% types and tokens of the total bundles. The majority (5% of the total bundles) of these bundles have been used to describe research, e.g., to refer to the study, e.g., *objective of the study was*, (104) *First objective* ***of this study was*** […] (NNS 13), to refer to the procedure, e.g., *has been used as*, (105), to quantify, e.g., *is one of the*, (106) *As survey method* ***is one of the*** *scientific traditions* […] (NNS 14),

The rest (3% of the total bundles) have been used to organize text, such as, to present the outcome, e.g., *is the result of,* and to compare, e.g., *is the case with*, (107) *As* ***is the case with*** *any research* […] (NNS 17),

Only 1% (of the total bundles) have been used for presenting writers' evaluation, e.g., *there is a need,* (108) *However,* ***there is a need*** *of proper and suitable syllabus* […] (NNS 18).

Verb-based bundles with to-clause fragment

The Verb-based bundles with to-clause fragment represent 8% types and tokens of the total bundles. All these bundles have been used for describing research. such as, to justify, e.g., *in order to make,* (109) *As teachers are not involved in the process of policy making, so* ***in order to make*** *up the deficiencies in the curriculum* […] (NNS 19)., and to refer to the study itself., e.g., *of the study was,* (110) *The major aim of* ***the study was to*** *check different sections* […] (NNS 10).

Verb-based bundles with that-clause fragment represent

The Verb-based bundles with that-clause fragment represent 7% of bundle types and 9% tokens of the total bundles. The majority of these bundles (5% types and tokens of the total bundles) have been used for describing research, such as, for making statements, e.g., *that there is no,*

(111) *Through this research it was determined* ***that there is no*** *specific set of materials* [….] (NNS 05), and to quantify, e.g., *that most of the*, (112) *The overall mean score 3.35 shows* ***that most of the*** *teachers* […] (NNS 03).

Only 1% types and tokens of the total bundles were used for presenting writers' opinion, e.g., *of the opinion that* etc., and a few of the bundles (1%) have been used for organizing the text, e.g., *that is why the, that is why it* etc.

Verb phrase with active verb

Verb Phrases with active verbs, represent 5% types and 11% tokens of the total bundles. All these bundles have been used for describing research, such as, referring to the responses in questionnaire, e.g., *agreed with the statement*, (115) […] *and majority of the respondents agreed with the statement that interactive technology-based activities arouse your interest in studies*. (NNS 02), and to refer to the study itself, e.g., *this chapter deals with*, (116) *This chapter deals with the methodological procedure of the current study*. (NNS 12)

Passive verb+ prepositional phrase fragment

These bundles represent 5% types and tokens of the total bundles. All of these bundles have been used for describing research, such as, referring to the procedures of research, e.g., *is used*

in the, can be used to, (117) *Audio recorders like talking tins, pegs or cards* ***can be used to*** *reinforce the learning* […] (NNS 04), referring to the section of study, e.g., *can be seen in*, (118) *It* ***can be seen in*** *the figure 4.11 that 55 participants strongly agreed* […] (NNS 01).

Anticipatory it + verb phrase/adjective

These bundles represent 4% types and tokens of the total bundles. Almost all of these bundles 3% types and tokens of the total bundles) have been used for presenting writers' observation, e.g., *it has been observed*, (119) ***It has been observed*** *that the students do not use English language excessively* […] (NNS), and to present the writers' opinion, e.g., *it is needed to*, (120), *To have better results* ***it is needed to*** *give proper training to the teachers*. (NNS 06)

Following is the summary of Verb-based bundles in non-native students' corpus:

- Verb-based bundles are the most common bundles with 38% types and tokens of the total bundles in the non-native corpus. The majority of these bundles (28% types and tokens of the total bundles) have been used for describing research.

- All the Verb-based bundles with to-clause fragment (8% types and tokens of the total bundles), verb-based bundles with active verb (5% types and tokens of the total

bundles), Passive verb with Prepositional phrase fragment (5% types and tokens of the total bundles) have been used for describing research.

- The majority of Verb-based bundles with that-clause fragment (5% types and tokens of the total bundles), and Copula be + noun/adj. phrase (5% types and tokens of the total bundles) have been used for describing research.

- The Verb-based bundles have been mainly used for describing research in the non-native students' corpus.

In the next section, the results and analysis of the Other structures will be presented

4.4.2.4 Other structures

Other structures consist of bundles that have noun/preposition/verb component, but these structures do not represent the most productive frames that have been included in the main structural categories. For example, the category Noun-based bundles with of-phrase fragment only represents the bundles with frame 'the/a+ Noun+ of the/a' i.e., *the results of the*. On the other hand, the bundle *majority of the respondents* has been categorized as the Other structure because it does not contain the bundle frame 'the/a+ Noun+ of the/a'. Other structures in the

non-native corpus account for 15% of types and tokens of the total bundles. Table 4.43 presents all the Other structures found in the non-native corpus.

Table 4.43 Other structures in the non-native student corpus

Other Structures	majority of the respondents, as well as the, findings of the study, one of the most, the use of technology (in), (the) objectives of the study, keeping in view the, statement of the problem, most of the time, very important role in, of the study and, all over the world, and the role of, as well as in, of the present research, of the study in, population of the study, both male and female, and at the same, findings of the research, of the most important, of the study this, part of the study, and its use in, of male and female, well aware of the, findings of the present, of the questionnaire the, purpose of the study, the collection of data, the purpose of research, between two or more, findings of this research, in real life situations, point of view of, result of the study

Table 4.44 presents the distribution of Other Structure in non-native students' corpus

Table 4.44 Frequency & % of other structures in the non-native student corpus

Structure	Types	%	Tokens	%
Other structures	**36**	**14.87%**	**765**	**11.83%**

The majority (10% of the total bundles) of these bundles have been used for describing research. For example, for referring to procedures of research, e.g., *the use of technology*, for referring to the study itself, e.g., *objectives of the study*, (121) *Main* ***objective of the study*** *was to review* […]. (NSS 06), for quantifying, e.g., *majority of the respondents*, and for referring to place, e.g., *all over the world*, (122) ***All over the world*** *essay writing is an integral part of the students' academic career* […] (NSS 15).

The rest (5% of the total bundles) of the bundles have been used for organizing the text, e.g., *as well as the, findings of the study* etc.

Following is the summary of the Structural characteristics of lexical bundles in non-native students' corpus:

- The Verb-based bundles are the most common bundles, representing 38% types and tokens of the total bundles in the non-native corpus. The majority of the Verb-based bundles (21% of the total bundles) have been used for describing research.
- The Preposition-based bundles represent 27 % types and tokens of the total bundles, half of which (15% types and tokens of the total bundles) have been used for organizing text, and half (12% types and tokens of the total bundles) have been used for describing research.
- The Noun-based bundles represent 21% types and tokens of the total bundles, the majority (16% of the total bundle types and tokens) of which have been used for describing research.
- So, more than half the types and tokens of the total bundles have been used for describing research in the non-native corpus.

4.4.3 Bundle functions in non-native corpus

In this section, I will present the results, and the analysis of bundle functions found in the non-native student corpus. For the analysis of bundle functions, I will follow the same pattern as was followed in the analysis of bundle functions of the expert and native student corpora. At

first, I will present the analysis of the overall distribution of bundle functions in the three main categories: Research-oriented bundles, Text-oriented bundles, and participant-oriented bundles. The analysis of the overall distribution of the bundle functions will follow the detailed analysis of all bundle functions in the sub-categories. For the illustration of bundle functions, examples from the non-native students' corpus will be presented.

Table 4.45 presents the frequency distribution, in terms of types and tokens, in the native student corpus. The raw frequencies and percentages have been presented in the table.

Table 4.45 Frequency & % of functional categories in the non-native student corpus

Functions	**Types**	**%**	**Tokens**	**%**
Research-oriented bundles	**133**	**54.95%**	**3956**	**58.86%**
Location	6	2.47	154	2.29
Procedure	88	36.36	2817	41.91
Quantification	19	7.85	547	8.13
Description	20	8.26	438	6.51
Text-oriented bundles	**96**	**39.66%**	**2322**	**34.55%**
Transition signals	11	4.54	444	6.60

Resultative signals	15	6.19	440	6.54
Structuring signals	33	13.63	620	9.22
Framing signals	37	15.28	818	12.17
Participant-oriented bundles	**13**	**5.37%**	**442**	**6.57%**
Stance features	10	4.13	382	5.68
Engagement features	3	1.23	60	0.89
Total	**242**	**100%**	**6720**	**100%**

Research-oriented bundles are the most common bundles representing over half (55% types and tokens of the total bundles) in the non-native student corpus. This suggests that the main feature of non-native student corpus is to describe research.

The text-oriented bundles, used for organizing text, represent 40% types and tokens of the total bundles, whereas the participant-oriented bundles are the least common bundles representing only 5% of the total bundles.

So, describing research, especially describing the procedures of research is an important feature of bundle functions in the non-native corpus.

In the next section, the detailed analysis of the bundle functions will be presented.

4.4.3.1 Research-oriented bundles

Research-oriented bundles are used for describing research. There are four sub-categories of these bundles: Location bundles, procedure bundles, quantification bundles, and description bundles. In the non-native corpus, these bundles represent more than half of the total bundles. Table 4.46 presents all the research-oriented bundles found in the non-native corpus:

Table 4.46 Research-oriented bundles in the non-native student corpus

Research-oriented bundles	
Location	at the end of, the end of the, all around the world, at the time of, the beginning of the
Procedure	agreed with the statement (that), of the respondents agreed, with the statement and, (in) the analysis of the (data), in the field of, by the use of, of the respondents were, of the use of, the use of technology (in), a vital role in, (the) objectives of the study, on the use of, to the use of, is used in the, statement of the problem, can be used to, a part of the, very important role in, (plays)an important role in (the), of the study and, and the role of, the study will be, to participate in the, an essential part of, of this study was (to), the need of the, has been used as, of the importance of, of the study in, (objective/aim) of the study was, become a part of, for the use of, is used as a, play a vital role, the population

	of the, the purpose of this, were found to be, population of the study, the importance of the, the light of the, the performance of the, the significance of the, an important part of, both male and female, in the development of, is the use of, of the study is, study was conducted to, the purpose of the, an integral part of, findings of the research, for the collection of, of the most important, of the study this, part of the study, the aim of the, the answer to the (research questions), can be used for, of male and female, are found to be, is the product of, of the questionnaire the, of the study to, of this research is, of this research was, purpose of the study, the collection of data, the meanings of the, the needs of the, the purpose of research, the researcher tried to, were selected for the, between the use of, can be divided into, in the fields of, of the findings of, participants were asked to, the sample of the, the topic of the, to collect the data
Quantification	majority of the respondents, that most of the, is one of the (main), one of the most, (but) wih the passage of (time), a large number of, the majority of the, most of the time, a lot of time, as one of the, the total number of, a wide range of, the degree to which, the extent to which, that the majority of, a small number of, are some of the, between two or more, by majority of the, that there are some

Description	in the form of, (and) the use of the, (significance) of the study the, the attitude of the, the role of the, is the case with, in the area of, that they are not, he is of the, the status of a, the characteristics of the, the review of the, the status of the, used in this study

As can be seen above, the procedure bundles represent the majority of the research-oriented bundles in the non-native corpus. Table 4.47 presents the distribution of Research-oriented bundles in non-native students' corpus.

Table 4.47 Frequency & % of Research-oriented bundles in the non-native student corpus

Functions	**Types**	**%**	**Tokens**	**%**
Research-oriented bundles	**133**	**54.95%**	**3956**	**58.86%**
Location	6	2.47	154	2.29
Procedure	88	36.36	2817	41.91
Quantification	19	7.85	547	8.13
Description	20	8.26	438	6.51

Procedure bundles

The research-oriented procedure bundles represent the majority (36% types and 42% tokens of the total bundles) in the non-native corpus. Procedure bundles have been used for describing the procedures, and the qualitative features of the research, e.g., *agreed with the statement*, (125) […]14 teachers *agreed with the statement,* 3 teachers were uncertain in their opinion […] (NNS 12), *the use of the*, (126) […] towards *the use of the* Internet for English language learning (NNS 01), *plays an important role,*(127) *This shows that interaction* ***plays an important role*** *in enhancing speaking skills of second language learners*. (NNS 07)

Quantification bundles

Quantification bundles represent 8% of the total bundles. These bundles have been used for quantifying, e.g., *majority of the respondents*, (128) ***Majority of the respondents*** *favoured the idea of enhancing the process of second language learning with the help of interactive technology*. (NNS 02)

Description bundles

Research-oriented description bundles also represent 8% types and tokens of the total bundle. These bundles have been used for presenting description of different features of research, e.g., *in the form of,* (129) *This collocation appears* ***in the form of*** *a Noun Phrase* […] (NNS19)

Location bundles

Location bundles represent merely 2% types and tokens of the total bundles. These bundles have been used for describing place and time of an event, e.g., *at the end of, all around the world* etc.

Following are the main features of Research-oriented bundle in non-native students' corpus:

- Research-oriented bundles are the most common bundles representing 55% types and tokens of the total bundles.
- The Procedure bundles represent the majority (36% types and tokens of the total bundles) of the research-oriented bundles in the non-native corpus.

So, describing research, especially describing the procedures of research is the main feature of bundle functions in the non-native student corpus.

In the next section, the analysis of text-oriented bundles will be presented.

4.4.3.2 Text-oriented bundles

Text-oriented bundles are used for organizing text. These bundles have four sub-categories: transition signals, resultative signals, structuring signals, and framing signals. In the non-native corpus, these bundles represent 40% of total bundles. Table 4.48 presents all the text-oriented bundles found in the non-native student corpus.

Table 4.48 Text-oriented bundles in the non-native student corpus

Text-oriented bundles	
Transition signals	on the other hand (the), (but) at the same time (the), as well as the, as well as in, and they do not (have), in other words the, in spite of the, on the one hand, and at the same, and its use in, (and)on the other side
Resultative signals	the results of the (study), the findings of the, findings of the study, as a result of, is the result of, the responses of the, the result of the, the results of this, findings of the present, findings of this research, result of the study

Structuring signals	of the present study (the), in this chapter the, in the present study (are), for the present study, in this study the, the present study is, of the present research, in the study the, the present study was, for the present research (is), chapter deals with the, this research was to, in the present research, in this section the, this chapter deals with, discussed in detail in, in the study as, the relationship between the, the study the present, the study was to
Framing signals	with the help of, on the basis of (the), that of the respondents, in the use of, to find out the, in the process of, about the use of, that the use of, in the light of, for the development of, in the same way, in this way the, the view that the, for the purpose of, to know about the, as compared to the, regarding the use of, in the context of, that there is a, in such a way, on the part of (the), towards the use of, for the sake of, keeping in view the, of the fact that, to take part in, it refers to the, is based on the, in a better way, such a way that, that is why

	the, that there is no, the ways in which, that they do not (have/any), in order to make, the fact that the, be helpful for the, in order to understand, is related to the, same is the case, the context in which, in accordance with the, in order to get, in this regard the, in a way that, in this type of, that is why it, that it is the, the context of the. to be able to, to the fact that, well aware of the, when they try to, as a tool to, for the selection of, in order to find (out), to talk about the, are based on the, from the perspective of, in front of the, in real life situations, in the presence of, in which they are, to interact with the, when it comes to

As can be seen in Table 4.48 and Table 4.49, the framing signals and the structuring signals account for majority of the text-oriented bundles.

Table 4.49 Frequency & % of Text-oriented bundles in the non-native student corpus

Functions	Types	%	Tokens	%
Text-oriented bundles	**96**	**39.66%**	**2322**	**34.55%**
Transition signals	11	4.54	444	6.60
Resultative signals	15	6.19	440	6.54
Structuring signals	33	13.63	620	9.22
Framing signals	37	15.28	818	12.17

Framing signals

Framing signals represent 15% types and tokens of the total bundles. These bundles have been used to contextualize new information in the text, e.g., *on the basis of*, (130) *Khan only expresses Sidhwa as pioneering women narrative writer and her work* ***on the basis of*** *different aspects* […]. (NNS 01), *in order to understand*, (131) ***in order to understand*** *any social or political* [….] (NNS 16)

Structuring signals

Structuring signals represent 14% types and tokens of the total bundles. These bundles have been used for referring to the section, the chapter, or the text itself in the non-native corpus, e.g., (132) ***In this chapter, the*** *data have been interpreted in detail*. (NNS 08), *of the present study,* (133) *Aim* ***of the present study*** *is to evaluate English textbook for class 9th*. (NNS 16)

Resultative signals

Resultative signals represent 6% types and tokens of the total bundles. These bundles have been used for referring to the results, e.g., *the results of the*, (134) ***The results of the*** *responses by the participants* [….]. (NNS 02) At times, these bundles were used to refer to the outcome, e.g., *as a result of,* (135) *Pakistan came into being* ***as a result of*** *the division of the subcontinent in 1947.* (NNS 01)

Transition bundles

Transition bundles represent 5% types and tokens of the total bundles. These bundles have been used to link sentences in the discourse, e.g., *on the other hand, at the same time.* etc., (136*) Government education institutions are provided with these facilities though the process is in preliminary stage.* ***At the same time****, results indicate that private education institutions have also realized the important role of ICT in learning and teaching*. (NNS 01)

Following are the main features of Text-oriented bundle in non-native students' corpus:

- Text-oriented bundles represent 40% types and tokens of the total bundles in the non-native corpus.
- The framing signals, and structuring signals represent the majority of the text-oriented bundles (29% types and tokens of the total bundles), suggesting that the text-oriented bundles were mainly used for contextualizing new information, and for referring to the sections of the text.

In the next section, the results of participation-oriented bundles will be presented.

4.4.3.3 Participant-oriented bundles

Participant-oriented bundles are used for presenting authors' evaluation, and to engage the readers. These are the least common bundles representing 5% of the total bundles in the non-native corpus. Table 4.50 presents all the Participant-oriented bundles found in the non-native student corpus.

Table 4.50 Participant-oriented bundles in the non-native student corpus

Participant-oriented bundles	
Stance features	it is necessary to, there is a need (to), can be used in, find it difficult to, it is also a, it is needed to, it is believed that, point of view of, (is) of the opinion that, it has been observed (that), it was observed that (the), (is/are/were) of the view that
Engagement features	it is important to, it can be seen, can be seen in (the), is evident from the

Table 4.51 presents the distribution of Participant-oriented bundles in non-native students' corpus.

Table 4.51 Frequency & % of Participant-oriented bundles in the non-native student corpus

Functions	**Types**	**%**	**Tokens**	**%**
Participant-oriented bundles	**13**	**5.37%**	**442**	**6.57%**
Stance features	10	4.13	382	5.68
Engagement features	3	1.23	60	0.89

Stance features

Participant-oriented stance features represent 4% types and tokens of the total bundles. These bundles have been used to present authors' evaluation and that of the other scholars mentioned in the text, e.g., *it is necessary to,* (137) ***It is necessary to*** *mention all the positive and negative aspects of attitude* […] (NNS 06), *of the view that,* (138) *Tomlinson (2011) is* ***of the view that*** […]. (NNS 19)

Engagement features

Engagement features represent only 1% types and tokens of the total bundles. These bundles have been used for engaging the readers by guiding them through the text or referring to the evidence in the text, e.g., *as can be seen,* (139) ***as can be seen*** *in Table 4.5* […]. (NNS 16)

Following are the main features of participant-oriented bundle in non-native students' corpus:

- The participant-oriented bundles are not very common in the non-native student corpus, representing only 5% of the total bundles, suggesting that presenting writers' evaluation and engaging the reader are given least importance in the non-native student corpus.

4.5 Comparison of bundle use in the expert and the native student corpora

In this section, the two corpora, expert corpus and native student corpus, will be compared. For the comparison, the top 20 bundles in each corpus will be compared at first. The relative frequencies will also be given for comparison as the two corpora are different in size. In the comparison of top 20 bundles, I will show the bundles that are similar in both the corpora. I will also compare the frequencies and the use of those bundles. This will lead to the comparison of structural and functional characteristics of the bundles in the expert and the native student corpora.

4.5.1 Comparison of the top 20 bundles in expert and the native student corpora

The comparison of the top 20 bundles used in both corpora will be presented in Table 4.52. For the comparison of the two different size corpora, it was essential to present the absolute frequencies (ABS) and the relative frequencies (REL) in the table. The similar bundles in both the corpora have been bolded in the list. Table 4.52 presents the list of top 20 bundles in the expert and the native student corpora.

Table 4.52 The 20 highly frequent bundles in Expert writers' corpus & native students' corpus

Rank	Expert writers	ABS	REL	Native students	ABS	REL
1.	**in the context of** (the)	56	10.96	(that) **the use of the**	52	16.61
2.	at the same time	54	10.57	**on the other hand** (the)	49	15.65
3.	**in the case of**	54	10.57	**as a result of** (the)	37	11.82
4.	on the basis of (the)	47	9.20	the results of the	36	11.50
5.	over the course of (the/a)	46	9.00	(is) that there is a	31	9.90
6.	**on the other hand**	43	8.41	as a function of	30	9.58
7.	the end of the	43	8.41	**it is important to**	29	9.26
8.	**in terms of the**	39	7.63	**the extent to which**	29	9.26
9.	**it is important to** (note that)	39	7.63	in line with the	28	8.94
10.	at the beginning of (the)	38	7.43	**in relation to the**	28	8.94
11.	the ways in which	38	7.43	**as well as the**	26	8.30
12.	**as well as the**	37	7.24	it is possible that	25	7.98
13.	in this case the	33	6.46	for the purposes of	24	7.66
14.	on the part of (the)	33	6.46	**in the context of**	24	7.66
15.	(that/in/on) **the use of the** (a)	33	6.46	the way in which	24	7.66
16.	**the extent to which**	31	6.06	**in the case of**	23	7.34
17.	in the form of	30	5.87	**in terms of the**	22	7.02
18.	**as a result of**	28	5.48	the total number of	21	6.70
19.	**in relation to the**	25	4.89	in the field of	20	6.39
20.	the results of this (study)	24	4.69	in the same way	20	6.39

Table 4.52 shows that half of the top 20 bundles in the expert and the native student corpora have been shared.

The top 18% bundles of the total bundles in expert corpus and top 16% bundles of the total bundles, in the native student corpus have been used for organization of the text. For example, the bundles, e.g., *in the context of, on the other hand, as a result of, in terms of the, in relation to the* were used in both the corpora.

The most frequent bundle in the expert corpus, *in the context of,* has been used for organization of the text, whereas the most frequent bundle in the native student corpus, *the use of the*, has been used for describing research.

Considering the relative frequencies, top 3% bundles in the expert corpus occurred 10 or more times, whereas top 4% bundles in the native student corpus occurred 10 or more times, showing similar frequency patterns in both the corpora.

Following are the main features of top 20 bundles in expert and native student corpora:

- The expert and the native student corpora are close to each other in their use of top 20 bundles as they share half of their top 20 bundles.
- The focus of top 20 bundle use in both the corpora is on the organization of the text.

The next section will compare the structural characteristics of bundles in the expert and the native student corpora.

4.5.2 Comparison of the structural characteristics in the expert and the native student corpora

In this section I will compare the structural characteristics of bundles in the expert and the native student corpora. At first, the distribution of the structural categories of bundles in both the corpora will be compared. The results of loglikelihood test have been indicated in the table where there are significant differences in the use of bundles between the expert and the native student corpora. Table 4.53 displays the distribution of bundle types and tokens in the expert and the native student corpora.

Table 4.53 Frequency & percentages of bundle structures (types & tokens) in expert & native student corpora

	Types				**Tokens**			
Structure	**Experts**	**%**	**Natives**	**%**	**Experts**	**%**	**Natives**	**%**
Noun-based bundles	**24**	**25.2%**	**32^{++}**	**34.77**	**435**	**23.58%**	**525^{++++}**	**33.79**
Noun-based bundles with of-phrase fragment	16	16.84%	25^{++}	27.17	284	15.40%	409^{++++}	26.33%

Noun-based bundles with other post modifier fragment	8	8.42%	7	7.60	151	8.18%	116	7.46
Preposition-based bundles	**47**	**49.4%**	**30**	**32.6**	**1048**	**56.83%**	**583^{-}**	**37.53**
Preposition-based bundles with of-phrase fragment	29	30.52%	17	18.47	662	35.90%	331^{--}	21.31
Preposition-based bundles with other post modifier fragment	18	18.94%	13	14.13	386	20.93%	252	16.22
Verb-based bundles	**18**	**18.9%**	**25^{++}**	**27.14**	**262**	**14.19%**	**369^{++++}**	**23.74**
Copula be + NP/Adj phrase	1	1.05%	4	4.34	13	0.70%	85^{++++}	5.47
VP with active verb	-	0%	-	0	-	0%	-	0
Anticipatory it + VP/Adj phrase	6	6.31%	7	7.60	103	5.58%	110^{++++}	7.08
Passive verb + PP fragment	4	4.21%	6	6.52	63	3.41%	79^{++++}	5.08
VP+ that clause fragment	3	3.15%	2	2.17	34	1.84%	31	1.99
Verb/adj + to clause fragment	4	4.21%	6	6.52	49	2.65%	64^{++++}	4.12
Other structures	**6**	**6.31%**	**5**	**5.43**		**5.34%**	**76**	**4.89**
	6	6.31%	5	5.43	99	5.34%	76	4.89
Total	**95**	**100%**	**92^{++}**	**100%**	**1844**	**100%**	**1553^{++++}**	**100%**

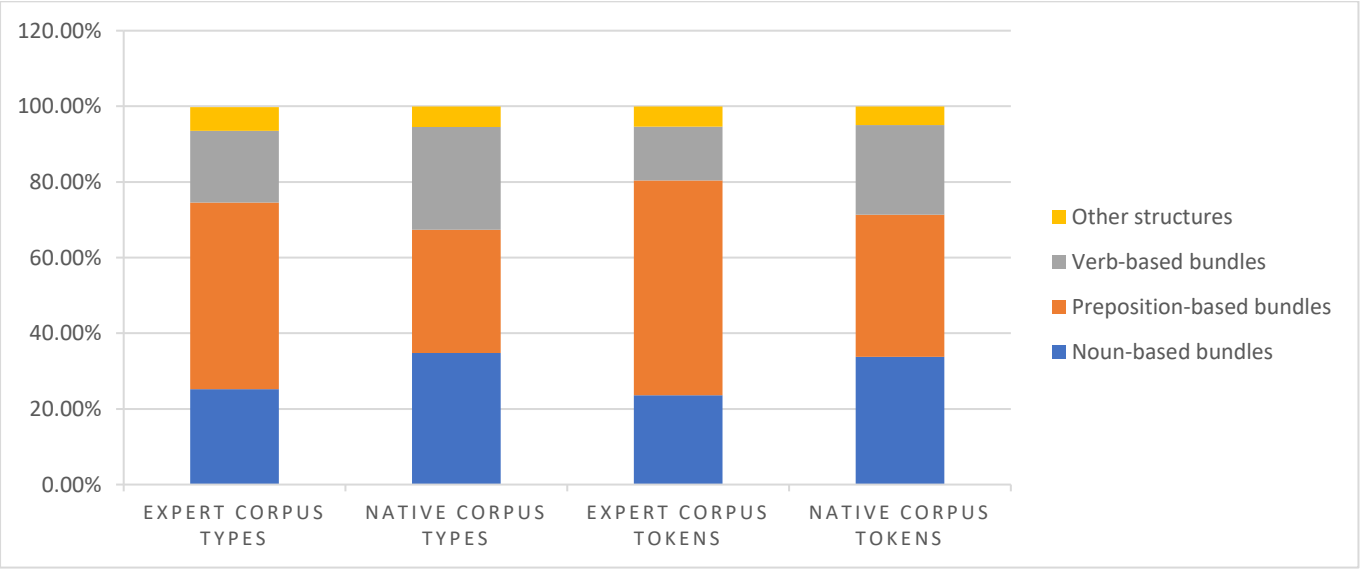
LEGEND (--) Statistically significant underuse in native corpus (at $p < 0.01$, critical value 6.63) (-) Statistically significant underuse in native corpus (at $p < 0.05$, critical value 3.84) (++++) Statistically significant overuse in native corpus (at $p < 0.0001$, critical value 15.13) (+++) Statistically significant overuse in native corpus(at $p < 0.001$, critical value 10.83) (++) Statistically significant overuse in native corpus at $p < 0.01$, critical value 6.63 (+) Statistically significant overuse in native corpus (at $p < 0.05$, critical value 3.84)

In the expert corpus, Preposition-based bundles are the most common bundles representing 49% types and tokens of the total bundles, the Noun-based bundles represent one quarter of

bundles, and the Verb-based bundles are not very common representing 19% of the total bundles. In the expert corpus, Preposition-based bundles have been used significantly more frequently than in the native student corpus.

In the native student corpus, Noun-based bundles representing 35% types and tokens of the total bundles, and Preposition-based bundles representing 33% types and tokens of the total bundles are equally the most common bundles, whereas the Verb-based bundles represent 27% types and tokens of the total bundles. In the native student corpus, Noun-based bundles, and Verb-based bundles have been used more frequently than the expert writers (see Figure 4.1).

Figure 4.1 Types & tokens of bundle structures in the expert & the native student corpora



4.5.2.1 Noun-based bundles

In this section, I will compare the Noun-based bundle sub-categories in the expert and the native student corpora. Noun-based bundles represent 25% of total bundles in the expert corpus, whereas these bundles represent 35% types and tokens of total bundle in the native student corpus (see Table 4.54).

Table 4.54 **Frequency & % of Noun-based bundles in the expert & the native corpora**

	Types				Tokens			
Structure	**Experts**	**%**	**Natives**	**%**	**Experts**	**%**	**Natives**	**%**
Noun-based bundles	**24**	**25.2%**	**32^{++}**	**34.77**	**435**	**23.58%**	**525^{++++}**	**33.79**
Noun-based bundles with of-phrase fragment	16	16.84%	25^{++}	27.17	284	15.40%	409^{++++}	26.33%
Noun-based bundles with other post modifier fragment	8	8.42%	7	7.60	151	8.18%	116	7.46

Noun-based bundles with of-phrase fragment

In the expert and the native student corpora, the majority of the Noun-based bundles with of-phrase fragment have been used for describing research. However, the native students have

used significantly more bundle types and tokens of these bundles than the expert writers. The similar trend was observed in the use of the Noun-Phrase bundle frame 'the__ of the' common in both the corpora. The native students used significantly more bundle tokens in this frame than the expert writers. Table 4.55 presents the use of bundles in this frame in both the corpora:

Table 4.55 Comparison of types & tokens in the bundle frame 'The__ of the' used in the expert and native corpora

Structure	**Types**		**Tokens**		**LOGL**
	Experts	**Natives**	**Experts**	**Natives**	
Noun-based bundles					
'The ___ of the'	12	15	212	242	42.33 (++++)

LEGEND (++++) Statistically significant overuse in native corpus (at $p < 0.0001$, critical value 15.13)

Noun-based bundles with other post-modifier fragment

The expert and the native student corpora were quite similar in their use of Noun-based bundles with other post-modifier fragment as both of them used these bundles for organizing the text. In the expert and the native student corpora, the use of Noun-based bundles was found similar,

however, the native students used significantly more bundle types and tokens than the expert writers.

4.5.2.2 Preposition-based bundles

The expert writers have used significantly more tokens of Preposition-based bundles than the native students (see Table 4.56).

Table 4.56 Frequency & % of Preposition-based bundles in the expert & the native corpora

Structure	Types				Tokens			
	Experts	**%**	**Natives**	**%**	**Experts**	**%**	**Natives**	**%**
Preposition-based bundles	**47**	**49.4%**	**30**	**32.6**	**1048**	**56.83%**	**583**	**37.53**
Preposition-based bundles with of-phrase fragment	29	30.52%	17	18.47	662	35.90%	331	21.31
Preposition-based bundles with other post modifier fragment	18	18.94%	13	14.13	386	20.93%	252	16.22

Preposition-based bundles with of-phrase fragment

In both corpora, the Preposition-based bundles are the most common bundles and the majority of these bundles have been used for organizing the text, however, expert writers have used

significantly more tokens of these bundles than the native students. The two most common PP-based bundles with of-frame '*in the ___ of*' and *'at the ___ of*' have been used significantly more frequently in the expert corpus. Table 4.57 and Table 4.58 present the use of bundle types and tokens in these frames.

Table 4.57 Comparison of types & tokens in the bundle frame 'in the __ of' used in the expert and native corpora

Structure	Types		Tokens		LOGL
	Experts	Natives	Experts	Natives	
Preposition-based bundles					
in the ___ of	11	5	254	91	21.59 (----)

LEGEND (----) Statistically significant underuse in native student corpus (at $p < 0.0001$, critical value 15.13)

Table 4.58 Comparison of types & tokens in the bundle frame 'at the__ of' used in the expert and native corpora

Structure	Types		Tokens		LOGL
	Experts	Natives	Experts	Natives	
Noun-based bundles					
at the ___ of	6	2	110	30	18.17 (----)

LEGEND (----) Statistically significant underuse in native corpus (at $p < 0.0001$, critical value 15.13)

As can be seen in Table 4.58, the expert writers have used significantly more bundle tokens in these two frames.

Preposition-based bundles with other post-modifier fragment

The expert writers have used more preposition-based bundles with other-fragment, however, there is no significant difference in the use of these bundles between the two corpora. In both the corpora, the majority of the Preposition-based bundles have been used for organizing text, however, expert writers have used these bundles significantly more frequently than the native students.

4.5.2.3 Verb-based bundles

Verb-based bundles are not very common in the expert corpus representing 19% types, 14% tokens of the total bundles, whereas in the native corpus, these bundles represent 27% types and tokens of the total bundles. The native students have used significantly more Verb-based bundle types and tokens than the expert writers (see Table 4.59).

Table 4.59 Frequency & % of Verb-based bundles (types & tokens) in the expert & the native student corpora

Structure	Types				Tokens			
	Experts	**%**	**Natives**	**%**	**Experts**	**%**	**Natives**	**%**
Verb-based bundles	**18**	**18.9%**	**25^{++}**	**27.14**	**262**	**14.19%**	**369^{++++}**	**23.74**
Copula be + NP/Adj phrase	1	1.05%	4	4.34	13	0.70%	85^{++++}	5.47
VP with active verb	-	0%	-	0	-	0%	-	0
Anticipatory it + VP/Adj phrase	6	6.31%	7	7.60	103	5.58%	110^{++++}	7.08
Passive verb + PP fragment	4	4.21%	6	6.52	63	3.41%	79^{++++}	5.08
VP+ that clause fragment	3	3.15%	2	2.17	34	1.84%	31	1.99
Verb/adj + to clause fragment	4	4.21%	6	6.52	49	2.65%	64^{++++}	4.12

Following are the main features of the Verb-based bundles used in expert and native student corpora:

- In both the expert and the native student corpora, the majority of Verb-based bundles have been used for describing research.
- In both corpora, all the anticipatory it + verb/noun phrase bundles have been used for presenting writers' opinion and for engaging the readers.

- In both corpora, all the Verb-based bundles with that-clause fragment, Copula be + noun/adjective phrase, and the majority of Passive verb with Prepositional fragment have been used for describing research.
- In the expert corpus, all the Verb-based bundles with to-clause fragment have been used for describing research, however, in the native corpus, all of these bundles have been used for presenting writers' evaluation.
- In both corpora, the Verb-based bundles have been mainly used for describing research, however, the native student have used these bundles significantly more frequently than the expert writers.

4.5.2.4 Other Structures

The other structures are not very common in expert and the native student corpora, representing 5% of the total bundles in each of the two corpora. Similarly, half of these bundles have been used for describing research in both the corpora. There was no significant difference in the use of these bundles between the two corpora (see Table 4.60).

Table 4.60 Frequency & % of Other Structures (types & tokens) in expert & native student corpus

Structure	Types				Tokens			
	Experts	**%**	**Natives**	**%**	**Experts**	**%**	**Natives**	**%**
Other structures	**6**	**6.31%**	**5**	**5.43**		**5.34%**	**76**	**4.89**
	6	6.31%	5	5.43	99	5.34%	76	4.89

The other structures account for over 5% of the total bundle types and tokens in the native texts, almost similar percentage represented by the other structures in the expert corpus.

4.5.2.5 Summary of the comparison of bundle structures:

- In the expert corpus, Preposition-based bundles are the most common bundles representing 49% types and tokens of the total bundles, the Noun-bundles represent one quarter of bundles, and the Verb-based bundles are not very common representing 19% of the total bundle types and tokens.

- In the native corpus, Noun-based bundles represent 35% types and tokens of the total bundles, and Preposition-based bundles 33% types and tokens of the total bundles are

equally the most common bundles, whereas the Verb-based bundles represent 27% types and tokens of the total bundles.

- In both corpora, the majority of Noun-based bundles are used for describing research, however, the native students have used significantly more Noun-based bundles than the expert writers.
- In both corpora, the majority of the Preposition-based bundles were used for organizing text, however, the expert writers have used significantly more Preposition-based bundles than the native students.
- In both corpora, the majority of the Verb-based bundles were used for describing research, however, the native students have used significantly more Verb-based bundles than the expert writers.

So, there are significant differences in the use of bundle structures in the expert and the native student corpora. The bundles used for organizing text are significantly more frequent in the expert corpus than in the native student corpus. On the other hand, the bundles used for describing research are significantly more common in the native student corpus than in the expert writers' corpus.

4.5.3 Comparison of the functional characteristics in the expert and the native student corpora

In this section I will compare the functional characteristics of bundles in the expert and the native student corpora. At first, the distribution of the functional categories of bundles in both corpora will be compared. The results of loglikelihood test have been indicated in the tables where there is significant difference in the use of bundles between the expert and the native student corpora.

Table 4.61 displays the distribution of bundle types and tokens in the expert and the native student corpora:

Table 4.61 Frequency & % of bundle functions (types & tokens) in the expert & the native student corpora

	Types				**Tokens**			
Functions	**Experts**	**%**	**Natives**	**%**	**Experts**	**%**	**Natives**	**%**
Research-oriented bundles	**33**	**41.05%**	**40^{++}**	**46.73%**	**533**	**35.05%**	**635^{++++}**	**45.13%**
Location	5	5.26	3	3.26	115	6.23	45^{--}	2.89
Procedure	15	18.94	12	17.39	224	15.23	184^{++}	18.73
Quantification	6	6.31	13^{++}	14.13	100	5.42	207^{++++}	13.32

Description	7	10.52	12^{+}	11.95	94	7.80	199^{++++}	10.17
Text-oriented bundles	**54**	**50.52%**	**40**	**40.20%**	**1179**	**58.13%**	**739**	**43.33%**
Transition signals	6	6.31	9	9.78	171	9.27	164^{++++}	10.56
Resultative signals	3	3.15	5	5.43	74	4.01	108^{++++}	6.95
Structuring signals	5	5.26	5	5.43	86	4.66	74^{+}	2.18
Framing signals	40	35.78	21	19.56	848	40.18	393^{----}	25.30
Participant-oriented bundles	**8**	**8.42%**	**12^{+}**	**13.04%**	**132**	**7.15%**	**179^{++++}**	**11.52%**
Stance features	6	5.26	7	6.52	102	3.41	113^{++++}	5.40
Engagement features	2	3.15	5	6.52	30	3.74	66^{++++}	6.11
Total	**95**	**41.05%**	**92^{++}**	**100%**	**1844**	**100%**	**1553^{++++}**	**100%**

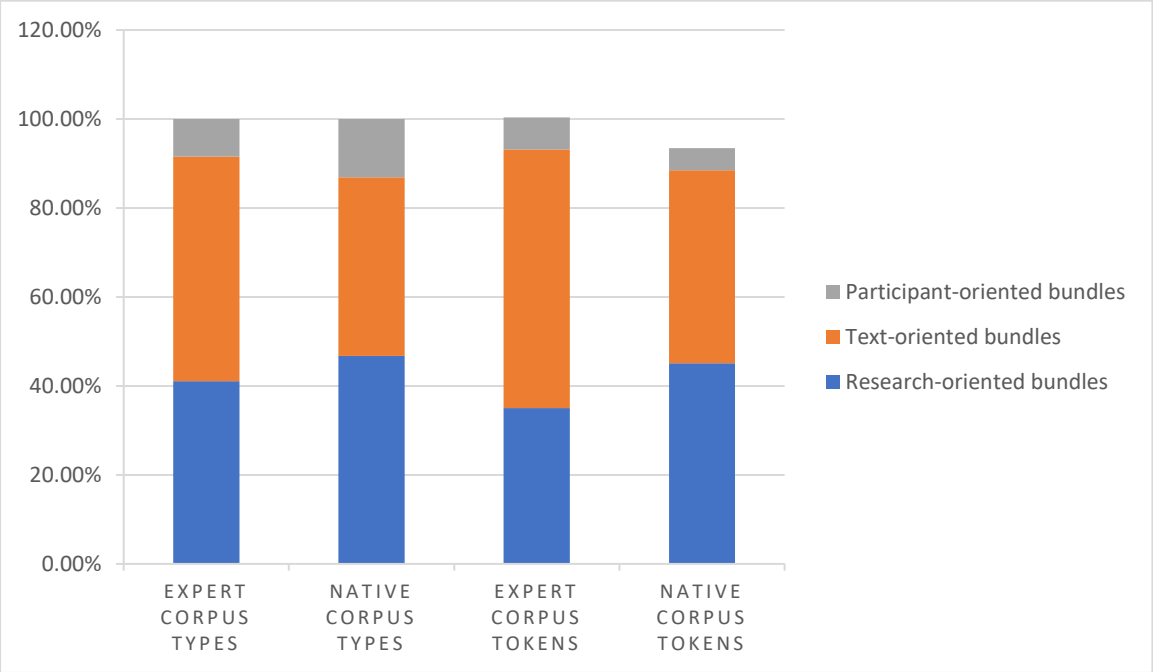
LEGEND (--) Statistically significant less frequent use in native corpus (at $p < 0.01$, critical value 6.63) (-) Statistically significant less frequent use in native corpus (at $p < 0.05$, critical value 3.84) (++++) Statistically significant more frequent use in native corpus (at $p < 0.0001$, critical value 15.13) (+++) Statistically significant more frequent use in native corpus(at $p < 0.001$, critical value 10.83) (++) Statistically significant more frequent use in native corpus at $p < 0.01$, critical value 6.63 (+) Statistically significant more frequent use in native corpus (at $p < 0.05$, critical value 3.84)

In the expert corpus, Text-oriented bundles are the most common bundles representing 51% types and tokens of the total bundles. This shows that the main function of lexical bundles in expert writing is to organize text. The research-oriented bundles represent 41% types and tokens of the total bundle. The participant-oriented bundles are the least common bundles representing 8% types and tokens of the total bundles.

In the native student corpus, Research-oriented bundles are the most common bundles representing 47% types and tokens of the total bundles, indicating that the native students use bundles predominantly for describing research. The text-oriented bundles represent 40% of total bundle types and tokens. The participant-oriented bundles are the least common bundles representing 14% types and tokens of the total bundles (see Figure 4.2).

The detailed analysis of the functional sub-categories of the two corpora will be presented in the next section.

Figure.4.2 Types & tokens of bundle functions in the expert & the native student corpora



4.5.3.1 Research-oriented bundles

Research-oriented bundles are used for describing research. In the expert corpus, research-oriented bundles represent 40% types and tokens of the total bundles, whereas in the native student corpus these are the most common bundles representing 47% types and tokens of the total bundles.

Table 4.62 Frequency & % of Research-oriented bundles (types & tokens) in the expert & the native student corpora

	Types				Tokens			
Functions	**Experts**	**%**	**Natives**	**%**	**Experts**	**%**	**Natives**	**%**
Research-oriented bundles	**33**	**41.05%**	**40^{++}**	**46.73%**	**533**	**35.05%**	**635^{++++}**	**45.13%**
Location	5	5.26	3	3.26	115	6.23	45^{--}	2.89
Procedure	15	18.94	12	17.39	224	15.23	184^{++}	18.73
Quantification	6	6.31	13^{++}	14.13	100	5.42	207^{++++}	13.32
Description	7	10.52	12^{+}	11.95	94	7.80	199^{++++}	10.17

Procedure bundles

In the expert and the native student corpora, the Procedure bundles are the most common Research-oriented bundles, however the native students have used significantly more tokens of these bundles.

Description bundles

The native students have used significantly more types and tokens of Description bundles than the expert writers.

Quantification bundles

The native students have used significantly more types and tokens of Quantification bundles than the expert writers.

Location bundles

The expert writers have used significantly more tokens of Location bundles than the expert writers.

The results of the Text-oriented bundles will be presented in the next section.

4.5.3.2 Text-oriented bundles

Text-oriented bundles are used for organizing the text. In the expert corpus, these bundles are the most common bundles representing 51% of total bundles. In the native corpus, text-oriented bundles represent 40% types and tokens of total bundles (see Table 4.63).

Table 4.63 Frequency & % of Text-oriented bundles (types & tokens) in the expert & the native student corpora

	Types				Tokens			
Functions	**Experts**	**%**	**Natives**	**%**	**Experts**	**%**	**Natives**	**%**

Text-oriented bundles	**54**	**50.52%**	**40**	**40.20%**	**1179**	**58.13%**	**739**	**43.33%**
Transition signals	6	6.31	9	9.78	171	9.27	164^{++++}	10.56
Resultative signals	3	3.15	5	5.43	74	4.01	108^{++++}	6.95
Structuring signals	5	5.26	5	5.43	86	4.66	74^{+}	2.18
Framing signals	40	35.78	21	19.56	848	40.18	393^{----}	25.30

Framing signals

In both corpora, Framing signals are the most common text-oriented bundles (36% of the total bundles in expert corpus, and 23% of the total bundles in native students' corpus).

However, the expert writers have used significantly more framing signals than the native students, indicating that the expert writers use more bundles in order to contextualize new information.

Transition signals

In the expert and the native student corpora, the similar bundle types of Transition signals have been used, however, the native students used significantly more bundle tokens than the expert writers.

Structuring signals

The native students used significantly more bundle tokens than the expert writers.

Resultative signals

In the expert corpus, the Resultative signals have been used significantly more frequently in the native student corpus.

4.5.3.3 Participant-oriented bundles

Participant-oriented bundles are used for presenting writers' evaluation and for engaging the readers. These are the least common bundles, representing 8% types and tokens of the total bundles in the expert corpus and 13% (types and tokens of the total bundles in the native student corpus (see Table 4.64).

Table 4.64 Frequency & % of Participant-oriented bundles (types & tokens) in the expert & the native student corpora

Functions	Types				Tokens			
	Experts	**%**	**Natives**	**%**	**Experts**	**%**	**Natives**	**%**
Participant-oriented bundles	**8**	**8.42%**	**12^{+}**	**13.04%**	**132**	**7.15%**	**179^{++++}**	**11.52%**
Stance features	6	5.26	7	6.52	102	3.41	113^{++++}	5.40
Engagement features	2	3.15	5	6.52	30	3.74	66^{++++}	6.11

The use of the Participant-oriented bundles has been found to be similar in the experts and the native student corpus, however the native students have used significantly more tokens of these bundles than the expert writers.

4.5.4 Conclusion

Following are the main features of the bundles use in expert and native student corpora:

- The native students have relied more on lexical bundles in their texts, as is evident by the significantly more frequent use of bundle types and tokens in the native student corpus.

- The main characteristics of bundle structures and functions in the expert corpus is that they focus on the organization of the text and contextualizing the information in their writing, by using predominantly preposition-based and text-oriented bundles.

- By contrast, the focus of bundle structures and functions in the native student corpus is divided roughly equally between describing research and organizing the text. They used approximately equal numbers of Noun-based bundles (mostly used for describing research) and preposition-based bundles (mostly used for organizing the text). Similarly, they used roughly as many research-oriented bundles (used for describing research) and text-oriented bundles (used for organizing the text). But compared to expert writers, bundle use in native student corpus was far more focused on describing research, as is evident in their significantly more frequent use of Noun-based bundles, Verb-based bundles, research-oriented bundles, and participant-oriented bundles.

4.6 Comparison of the bundle use in expert and the non-native corpora

In this section, the two corpora: expert and non-native student corpora will be compared. For the comparison, the top 20 bundles in each corpus will be compared at first. The relative frequencies will also be given for comparison as the two corpora are different in size. In the comparison of top 20 bundles, I will show the bundles that are similar in both the corpora. I will also compare the frequencies and the use of those bundles. This will lead to the comparison of structural and functional characteristics of the bundles in the expert and the non-native student corpora.

4.6.1 Comparison of the top 20 bundles in expert and the non-native student corpora

For the comparison of the top 20 bundles used in both corpora will be presented in a table. For the comparison of the two different size corpora, it was essential to present the absolute frequencies (ABS) and the relative frequencies (REL) in the table. The similar bundles in both the corpora have been bolded in the list. Table 4.65 presents the list of top 20 bundles in the expert and the non- native student corpora:

Table 4.65 Top 20 bundles in the expert & the non-native corpora

Rank	Experts	ABS	REL	Non-natives	ABS	REL
1.	in the context of (the)	56	10.96	agreed with the statement (that)	459	91.26
2.	**at the same time**	54	10.57	of the respondents agreed	180	35.78
3.	in the case of	54	10.57	with the statement and	167	33.20
4.	**on the basis of** (the)	47	9.20	**on the other hand** (the)	168	33.40
5.	over the course of (the/a)	46	9.00	(is/are/were) of the view that	150	29.82
6.	**on the other hand**	43	8.41	majority of the respondents	146	29.02
7.	the end of the	43	8.41	with the help of	127	25.25
8.	in terms of the	39	7.63	(in) the analysis of the (data)	123	24.45
9.	it is important to (note that)	39	7.63	of the present study (the)	114	22.66
10.	at the beginning of (the)	38	7.43	**on the basis of** (the)	112	22.26
11.	the ways in which	38	7.43	(but) **at the same time** (the)	97	19.28
12.	**as well as the**	37	7.24	in the use of	96	19.08
13.	in this case the	33	6.46	that most of the	96	19.08
14.	on the part of (the)	33	6.46	the results of the (study)	89	17.69
15.	(that/in/on) the use of the (a)	33	6.46	**in the form of**	85	16.90
16.	the extent to which	31	6.06	to find out the	84	16.70
17.	**in the form of**	30	5.87	in the process of	83	16.50
18.	as a result of	28	5.48	the findings of the	75	14.91
19.	in relation to the	25	4.89	in the field of	70	13.91
20.	the results of this (study)	24	4.69	**as well as the**	67	13.32

As can be seen from Table 4.65, 5 of the top 20 bundles have been shared by the expert and the non-native corpora. Importantly, all the five shared bundles have been significantly more frequent in the non-native student corpus.

Top 18% of the total bundles in expert corpus have been used for the organization of the text. For example, the bundles, e.g., *in the context of, on the other hand, as a result of, in terms of the, in relation to the* are used for organization of the text.

On the other hand, more than half of the top 20 bundles in the non-native corpus have been used for describing the research. For example, the bundles, e.g., *agreed with the statement, majority of the respondents, in the use of* etc. This indicates that the non-native student focus on describing research.

Considering the raw frequencies of top 20 bundles, the top 10% types of the total bundles in the expert corpus represent the one quarter of the bundle tokens, whereas only 4% types of the total bundles in the non-native corpus represent more than quarter of the bundle tokens. This indicates that only a few bundle types have been used highly frequently in the non-native corpus.

Considering the relative frequencies, only top 3/20 bundles in the expert corpus occurred 10 or more times, whereas all the top 20 bundles in the non-native student corpus occurred 10 or more times, again showing the highly frequent use of bundles in the non-native student corpus.

The important features of the analysis of the top 20 bundles in the expert and the non-native corpora are as follows:

- The expert and the native student corpora tend to be different in their use of top 20 bundles as they share less than half of their top 20 bundles.
- The focus of both the corpora is different as almost all the top 20 bundles in the expert corpus have been used for the organization of the text, whereas more than half of the top 20 bundles in the non-native corpus have been used for describing the research.
- The non-native students have used bundles significantly more frequently, as is evident from the significant frequent use of shared bundles, as well as the significantly frequent use of the top 4% bundles in the non-native corpus.

4.6.2 Comparison of the structural characteristics in the expert and the non-native student corpora

In this section I will compare the structural characteristics of bundles in the expert and the non-native student corpora. At first, the distribution of the structural categories of bundles in both the corpora will be compared. The distribution of structural characteristics will follow the detailed analysis of the comparison of the use of bundles in both the corpora. In Table 4.66, bundle types and tokens have been compared. The results of loglikelihood test have been indicated in the table where there are significant differences in the use of bundles between the expert and the non-native student corpora.

Table 4.66 Frequency & % of bundle structures (types & tokens) in the expert & the non-native student corpora

Structure	Types				Tokens			
	Experts	%	N-Natives	%	Experts	%	N-Natives	%
Noun-based bundles	**24**	**25.2%**	**50^{++++}**	**20.65%**	**435**	**23.58%**	**1016^{++++}**	**15.78%**
Noun-based bundles with of-phrase fragment	16	16.84%	39 $^{++}$	16.11	284	15.40%	859^{++++}	12.78%
Noun-based bundles with other post modifier fragment	8	8.42%	11	4.54	151	8.18%	202^{++++}	3.00

Preposition-based bundles	**47**	**49.4%**	**65^{++++}**	**26.85%**	**1048**	**56.83%**	**2417^{++++}**	**35.95%**
Preposition-based bundles with of-phrase fragment	29	30.52%	37^{++}	15.28	662	35.90%	1407^{++++}	20.93
Preposition-based bundles with other post modifier fragment	18	18.94%	28^{++}	11.57	386	20.93%	1010^{++++}	15.02
Verb-based bundles	**18**	**18.9%**	**91^{++++}**	**37.58%**	**262**	**14.19%**	**2477^{++++}**	**36.83%**
Copula be + NP/Adj phrase	1	1.05%	21^{++++}	8.67	13	0.70%	355^{++++}	5.28
VP with active verb	-	0%	11^{+++}	4.54	-	0%	755^{++++}	11.23
Anticipatory it + VP/Adj phrase	6	6.31%	10	4.13	103	5.58%	216^{++++}	3.21
Passive verb + PP fragment	4	4.21%	13^{+}	5.37	63	3.41%	204^{++++}	3.03
VP+ that clause fragment	3	3.15%	17^{++}	7.02	34	1.84%	605^{++++}	9.00
Verb/adj + to clause fragment	4	4.21%	19^{+++}	7.85	49	2.65%	342^{++++}	5.08
Other structures	**6**	**6.31%**	**36^{++++}**	**14.87%**		**5.34%**	**765^{++++}**	**11.83%**
	6	6.31%	36^{++++}	14.87	99	5.34%	765^{++++}	11.83
Total	**95**	**100%**	**242^{++++}**	**100%**	**1844**	**100%**	**6720^{++++}**	**100%**

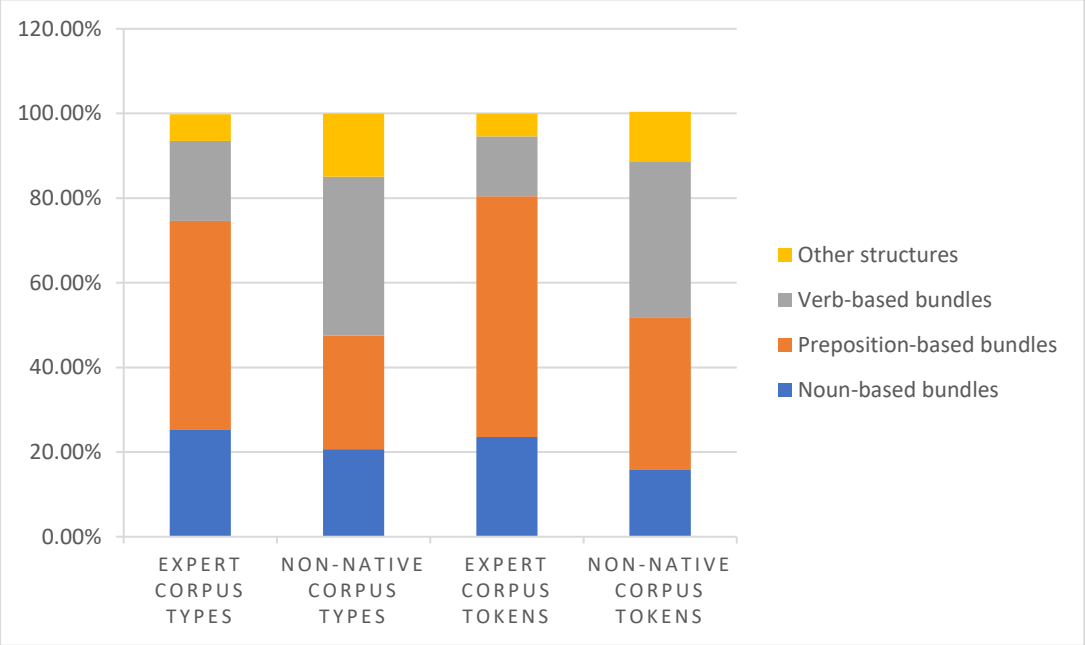
LEGEND (--) Statistically significant underuse in native corpus (at $p < 0.01$, critical value 6.63) (-) Statistically significant underuse in native corpus (at $p < 0.05$, critical value 3.84) (++++) Statistically significant overuse in native corpus (at $p < 0.0001$, critical value 15.13) (+++) Statistically significant overuse in native corpus(at $p < 0.001$, critical value 10.83) (++) Statistically significant overuse in native corpus at $p < 0.01$, critical value 6.63 (+) Statistically significant overuse in native corpus (at $p < 0.05$, critical value 3.84

Following are the main features of the distribution of structural characteristics of bundles in expert and non-native student corpora:

- The expert writers have used twice as many Preposition-based bundles (49% types, 57% tokens of the total bundles) as have been used by the non-native students (27% types, 36% tokens of the total bundles).
- The non-native students have used twice as many Verb-based bundles (38% types, 37% tokens of the total bundles) as have been used by the expert writers (19% types, 16% tokens of the total bundles)
- In both the corpora, the Noun-based bundles represent almost similar proportion of bundles, i.e., 25% of the total bundles in the expert corpus, 21% of the total bundles in the non-native student corpus.
- The non-native students have used significantly more bundle types and tokens than the expert writers.

Figure 4.3 shows the distribution of structural characteristics of bundles in expert and non-native student corpora:

Figure.4.3 Types & tokens of bundle structures in the expert & the non-native student corpora



4.6.2.1 Noun-based bundles

In this section, I will compare the Noun-based bundle sub-categories in the expert and the non-native student corpora. Table 4.67 presents the distribution of Noun-based bundles in expert and non-native student corpora.

Table 4.67 Frequency & % of Noun-based bundles (types & tokens) in the expert & the non-native student corpora

Structure	Types				Tokens			
	Experts	**%**	**N-Natives**	**%**	**Experts**	**%**	**N-Natives**	**%**
Noun-based bundles	**24**	**25.2%**	**50^{++++}**	**20.65%**	**435**	**23.58%**	**1016^{++++}**	**15.78%**
Noun-based bundles with of-phrase fragment	16	16.84%	39 $^{++}$	16.11	284	15.40%	859^{++++}	12.78%
Noun-based bundles with other post modifier fragment	8	8.42%	11	4.54	151	8.18%	202^{++++}	3.00

Noun-based bundles with of-phrase fragment

In both the corpora, the majority of the Noun-based bundles with of-phrase fragment have been used for describing research, However, the non-native students have used significantly more types and tokens of these bundles than the expert writers.

Similarly, the Noun-phrase bundle frame ‘the__ of the’ was the most common Noun-based bundle frame in both the corpora, but the non-native students used significantly more bundle

tokens in this frame as compared to the expert writers. Table 4.68 presents the comparison of the use of this bundle frame in the two corpora:

Table 4.68 Comparison of bundle frame 'the ___ of the' used in the expert & the non-native student corpora

Structure	**Experts**		**Non-atives**		**LOGL**
	ABS	**REL**	**ABS**	**REL**	
Noun-based bundles					
The ___ of the	212	40.90	694	137.98	273.03 (++++)

LEGEND (+++) Statistically significant overuse in non-native corpus (at $p < 0.0001$, critical value 15.13)

Noun-based bundles with other post-modifier fragment

In the expert corpus all the Noun-based bundles with other post-modifier fragment have been used for organizing the text, whereas in the non-native corpus half of these bundles, 3% types and tokens of the total bundles have been used for organizing the text. The use of Noun-based bundles in expert and non-native student corpora can be summarized as follows:

- In both the corpora, the majority of the Noun-based bundles with of-phrase fragment have been used for describing research, and the majority of the Noun-based bundles with other post-modifier fragment have been used for organizing the text. However, the non-native students have used significantly more types and tokens of these bundles than the expert writers.

- The non-native students have used significantly more bundle types and tokens of Noun-based bundles than the expert writers.

4.6.2.2 Preposition-based bundles with of-phrase fragment

Preposition-based bundles represent 49% types and 57% tokens of the total bundles in the expert corpus, whereas in non-native student corpus they represent 27% types and 36% tokens of the total bundles (see Table 4.69). So, the expert writers have used twice as many Preposition-based bundles than were used by the non-native students.

Table 4.69 Frequency & % of Preposition-based bundles (types & tokens) in the expert & the non-native student corpora

Structure	Types				Tokens			
	Experts	%	N-Natives	%	Experts	%	N-Natives	%
Preposition-based bundles	**47**	**49.4%**	**65**$^{++++}$	**26.85%**	**1048**	**56.83%**	**2417**$^{++++}$	**35.95%**

Preposition-based bundles with of-phrase fragment	29	30.52%	37^{++}	15.28	662	35.90%	1407^{++++}	20.93
Preposition-based bundles with other post modifier fragment	18	18.94%	28^{++}	11.57	386	20.93%	1010^{++++}	15.02

Preposition-based bundles with of-phrase fragment

In the expert corpus, Preposition-based bundles with of-phrase fragment are the most common bundles. Over half of these bundles have been used for organizing the text, whereas in the non-native student corpus the majority of these bundles have been used for describing research.

The analysis of the two Preposition-based bundle frames has been given below:

The most common PP-based frame, *'in the ___ of* 'was used significantly more frequently in the non-native corpus (see Table 4.70). On the other hand, the other most common PP-based structure *'at the ___ of '* that was used to identify place or time, has been used significantly more frequently in the expert corpus (see Table 4.71).

Table 4.70 Comparison of bundles frame 'in the ___ of' used in the expert & the non-native students corpora

Structure	Experts		Non-natives		LOGL
	ABS	REL	ABS	REL	
Preposition-based bundles					
'in the___ of'	254	50.20	460	91.46	61.52 (++++)

LEGEND (+++) Statistically significant overuse in non-native corpus (at p < 0.0001, critical value 15.13)

Table 4.71 Comparison of bundles frame 'at the ___ of' used in the expert & the non-native students corpora

Structure	Experts		Non-natives		LOGL
	ABS	REL	ABS	REL	
Preposition-based bundles					
at he ___ of	110	21.74	65	12.92	11.44 (---)

(---) Statistically significant underuse in non-native corpus (at p < 0.0001, critical value 10.83)

Preposition-based bundles with other post-modifier fragment

In both the expert and the non-native student corpora, the Preposition-based bundles with other post-modifier fragment have been used for organizing the text. The use of Preposition-based bundles in expert and non-native student corpora can be summarized as follows:

- The expert writers have used twice as many Preposition-based bundles (49% types and 57% tokens of the total bundles) as were used by the non-native students (27% types and 36% tokens of the total bundles).
- The expert writers have used more than twice as many Preposition-based bundles (34% of the total bundles) for organizing the text as were used by the non-native students (15% of their total bundles).
- In both the corpora, all the Preposition-based bundles with other-phrase fragment have been used for organizing the text.
- The non-native students have used significantly more tokens of Preposition-based bundles than the expert writers.

4.6.2.3 Verb-based bundles

Verb-based bundles are not very common in the expert corpus representing 19% types and tokens of the total bundles. The majority of these bundles (12% of the total bundles) have been

used for describing research, while (7% of the total bundles) were used for describing writers' opinion and engaging the readers.

In the non-native corpus, Verb-based bundles are the most common bundles representing 38% types and tokens of the total bundles. The majority (21% of the total bundles) of these bundles have been used for describing research, 12% types and tokens of the total bundles have been used for organizing text, whereas 7% types and tokens of the total bundles have been used for presenting writers' evaluation and for engaging the reader. So, in both the corpora, the majority of the Verb-based bundles have been used for describing research. Table 4.72 presents the distribution of Verb-based bundles in expert and non-native student corpora.

Table 4.72 Frequency & % of Verb-based bundles (types & tokens) in the expert & the non-native student corpora

Structure	Types				Tokens			
	Experts	**%**	**N-Natives**	**%**	**Experts**	**%**	**N-Natives**	**%**
Verb-based bundles	**18**	**18.9%**	**91^{++++}**	**37.58%**	**262**	**14.19%**	**2477^{++++}**	**36.83%**
Copula be + NP/Adj phrase	1	1.05%	21^{++++}	8.67	13	0.70%	355^{++++}	5.28
VP with active verb	-	0%	11^{+++}	4.54	-	0%	755^{++++}	11.23
Anticipatory it + VP/Adj phrase	6	6.31%	10	4.13	103	5.58%	216^{++++}	3.21
Passive verb + PP fragment	4	4.21%	13^{+}	5.37	63	3.41%	204^{++++}	3.03

VP+ that clause fragment	3	3.15%	17^{++}	7.02	34	1.84%	605^{++++}	9.00
Verb/adj + to clause fragment	4	4.21%	19^{+++}	7.85	49	2.65%	342^{++++}	5.08

The main features of Verb-based bundles used in expert and no-native student corpora are as follows:

- The non-native students have used twice as many Verb-based bundles (38% types and tokens of the total bundles) as were used by expert writers (19% types and tokens of the total bundles).
- In both the corpora, the majority of Verb-based bundles (12% of the total bundles in expert corpus, 21% of the total bundles in the non-native corpus) have been used for describing research.
- In expert corpus, no Verb-based bundle with active verb was found, however, these bundles represent 5% types and 11% tokens of the total bundles in the non-native corpus.
- The non-native students have used significantly more Verb-based bundle types and tokens than the expert writers.

4.6.2.4 Other structures

The other structures are not very common in the expert corpus, representing 5% of the total bundles (see Table 4.73). The half of these bundles have been used for describing research.

Table 4.73 Frequency & % of Other Structures (types & tokens) in the expert & the non-native student corpora

Structure	Types				Tokens			
	Experts	**%**	**N-Natives**	**%**	**Experts**	**%**	**N-Natives**	**%**
Other structures	**6**	**6.31%**	**36^{++++}**	**5.43**		**5.34%**	**765**	**11.83%**
	6	6.31%	36^{++++}	5.43	99	5.34%	765	11.83

In the non-native corpus, the other structures represent 15% types and tokens of the total bundles. The majority of these bundles (10% of the total bundles) have been used for describing research, e.g., *objective of the study, well aware of the* etc. The non-native students have used significantly more tokens of other structures than the expert writers.

Following is the summary of the structural characteristics of bundle use in expert and non-native student corpora:

- The distribution of structural characteristics of bundles is different in the expert and the non-native student corpora and shows that the expert writers and the non-native students not only use bundles differently but also for different purposes.
- The expert writers have used more bundles for organizing the text, whereas the non-native students have used more bundles for describing research. The expert writers have used twice as many Noun-based bundles (12% of the total bundles), and the Preposition -based bundles (38% of the total bundles) for organizing text as were used by the non-native students (4% Noun-based bundles, 15% Preposition-based bundles).
- The non-native students have used more than twice as many Verb-based bundles (21% of the total bundles) for describing research as were used by expert writers (12% of the total bundles).
- The non-native students have used significantly more types and tokens than the expert writers.

4.6.3 Comparison of the functional characteristics in the expert and the Non-native student corpora

In this section I will compare the functional characteristics of bundles in the expert and the non-native student corpora. At first, the distribution of the functional categories of bundles in both the corpora will be compared. The distribution of functional characteristics will follow the detailed analysis of the comparison of the use of bundles in both the corpora. In Table 4.74, bundle types and tokens have been compared. The results of loglikelihood test have been indicated in the table where there are significant differences in the use of bundles between the expert and the non-native student corpora.

Table 4.74 displays the distribution of bundle types and tokens in the expert and the native student corpora:

Table 4.74 Frequency & % of bundle functions (types & tokens) in the expert & the non-native student corpora

Functions	Types				Tokens			
	Experts	**%**	**N-Natives**	**%**	**Experts**	**%**	**N-Natives**	**%**
Research-oriented bundles	**33**	**41.05%**	**138^{++++}**	**54.95%**	**533**	**35.05%**	**3857^{++++}**	**58.86%**
Location	5	5.26	6	2.47	115	6.23	154^{+}	2.29

Procedure	15	18.94	94^{++++}	36.36	224	15.23	2729^{++++}	41.91
Quantification	6	6.31	19^{++}	7.85	100	5.42	547^{++++}	8.13
Description	7	10.52	19^{+}	8.26	94	7.80	427^{++++}	6.51
Text-oriented bundles	**54**	**50.52%**	**91^{++}**	**39.66%**	**1179**	**58.13%**	**2421^{++++}**	**34.55%**
Transition signals	6	6.31	12	4.54	171	9.27	455^{++++}	6.60
Resultative signals	3	3.15	15^{++}	6.19	74	4.01	440^{++++}	6.54
Structuring signals	5	5.26	21^{++}	13.63	86	4.66	428^{++++}	9.22
Framing signals	40	35.78	43	15.28	848	40.18	1098^{++++}	12.17
Participant-oriented bundles	**8**	**8.42%**	**13**	**5.37%**	**132**	**7.15%**	**442^{++++}**	**6.57%**
Stance features	6	5.26	10	4.13	102	3.41	382^{++++}	5.68
Engagement features	2	3.15	3	1.23	30	3.74	60^{++}	0.89
Total	**95**	**100%**	**242^{++++}**	**100%**	**1844**	**100%**	**6720^{++++}**	**100%**

LEGEND (--) Statistically significant underuse in non-native corpus (at $p < 0.01$, critical value 6.63) (-) Statistically significant underuse in non-native corpus (at $p < 0.05$, critical value 3.84) (++++) Statistically significant overuse in non-native corpus (at $p < 0.0001$, critical value 15.13) (+++) Statistically significant overuse in non-native corpus(at $p < 0.001$, critical value 10.83) (++) Statistically significant overuse in non-native corpus at $p < 0.01$, critical value 6.63 (+) Statistically significant overuse in non-native corpus (at $p < 0.05$, critical value 3.84

As can be seen from Table 4.74, the two corpora in total contrast to each other. Both the corpora show completely different characteristic of bundle functions.

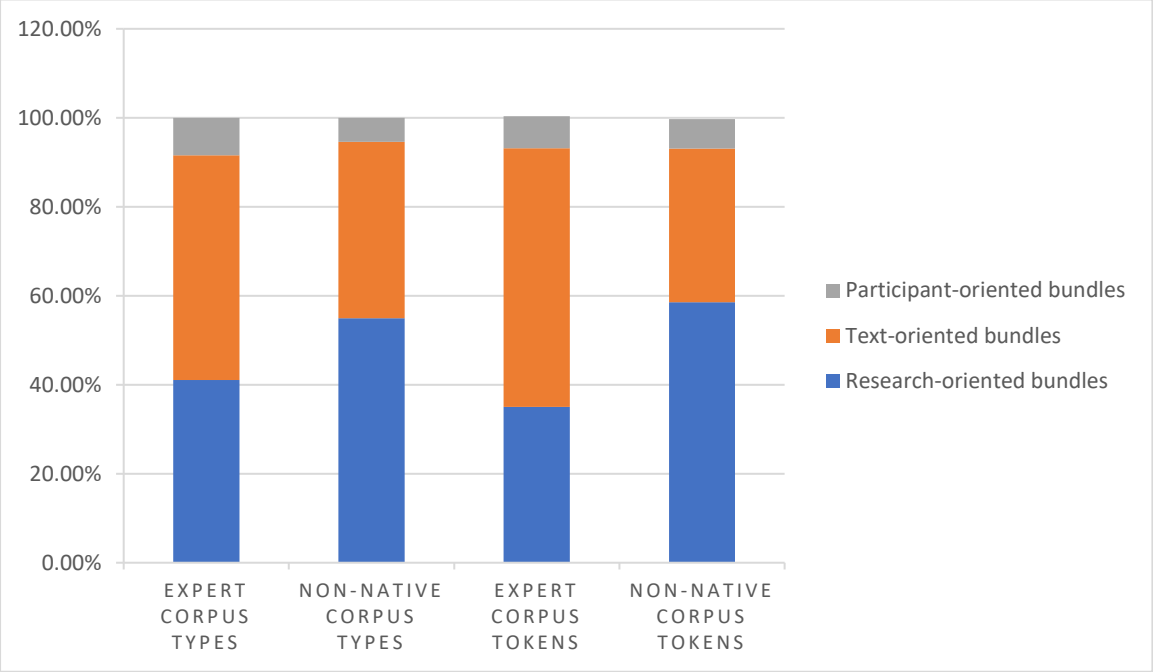
In the expert corpus, the text-oriented bundles, used for organizing text, are the most common bundles representing 51% types and tokens of the total bundles. The Research-oriented bundles represent 40% of the total bundles whereas the Participant-oriented bundles are the least common bundles representing 8% types and tokens of the total bundles.

In the non-native corpus, the research-oriented bundles, used for describing research, are the most common bundles representing 57% types and tokens of the total bundles.

So, both corpora are in contrast to each other in their bundle functions, with expert writers using more than half of their total bundles for organizing text, whereas the non-native students used more than half of their total bundles for describing research.

Figure 4.4 displays the distribution of functional characteristics of bundles in expert and non-native student corpora.

Figure.4.4 Types & tokens of bundle functions in the expert & the non-native student corpora



4.6.3.1 Research-oriented bundles

In the expert corpus, research-oriented bundles represent 40% types and tokens of the total bundles, whereas in the non-native corpus, they represented 57% types and tokens of the total bundles (see Table 4.75).

Table 4.75 Frequency & % of Research-oriented bundles (types & tokens) in the expert & the non-native student corpora

Functions	Types				Tokens			
	Experts	**%**	**N-Natives**	**%**	**Experts**	**%**	**N-Natives**	**%**
Research-oriented bundles	**33**	**41.05%**	**138^{++++}**	**54.95%**	**533**	**35.05%**	**3857^{++++}**	**58.86%**
Location	5	5.26	6	2.47	115	6.23	154^{+}	2.29
Procedure	15	18.94	94^{++++}	36.36	224	15.23	2729^{++++}	41.91
Quantification	6	6.31	19^{++}	7.85	100	5.42	547^{++++}	8.13
Description	7	10.52	19^{+}	8.26	94	7.80	427^{++++}	6.51

Following are the main feature of Research-oriented bundles used in expert and non-native student corpora:

- The non-native students have used far more Research-oriented bundles representing 57% types and tokens of the total bundles, whereas in the expert corpus these bundles represent 40% of the total bundles.

- The non-native students have used twice as many Procedures bundles (38% types and tokens of the total bundles) as were used by expert writers, representing 19% types and tokens of the total bundles.
- The non-native students have used significantly more Research-oriented bundles than the expert writers.

4.6.3.2 Text-oriented bundles

Text-oriented bundles are used for organizing the text. In the expert corpus, these bundles are the most common bundles representing 51% of total bundles, whereas in the non-native corpus, text-oriented bundles represent 40% types and tokens of the total bundles (see Table 4.76).

Table 4.76 Frequency & % Text-oriented bundles (types & tokens) in the expert & the non-native student corpora

Functions	Types				Tokens			
	Experts	%	N-Natives	%	Experts	%	N-Natives	%
Text-oriented bundles	**54**	**50.52%**	**91**$^{++}$	**39.66%**	**1179**	**58.13%**	**2421**$^{++++}$	**34.55%**

Transition signals	6	6.31	12	4.54	171	9.27	455^{++++}	6.60
Resultative signals	3	3.15	15^{++}	6.19	74	4.01	440^{++++}	6.54
Structuring signals	5	5.26	21^{++}	13.63	86	4.66	428^{++++}	9.22
Framing signals	40	35.78	43	15.28	848	40.18	1098^{++++}	12.17

Following are the main feature of Text-oriented bundles used in expert and non-native student corpora:

- In both the corpora, framing signals represent the majority of the text-oriented bundles, however, the expert writers have used more than twice as many framing signals as were used by the non-native students, indicating that the expert writers give far more importance to contextualizing new information than the non-native students.

- In expert corpus, Text-oriented bundles, used for organizing text, are the most common bundles representing 51% types and tokens of the total bundles, whereas these bundles represent 40% types and tokens of the total bundles in the native student corpus.

- Text-oriented framing signals, used for contextualizing new information, represent twice as many bundles (36% types and tokens of the total bundles) as were used by the non-native students (18% of the total bundles).

- The non-native students have used significantly more Text-oriented bundle types and tokens than the expert writers.

4.6.3.3 Participant-oriented bundles

Participant-oriented bundles are used for presenting writers' evaluation and engaging the readers. These are the least common bundles, representing 8% types and 7% tokens of total bundles in the expert corpus. Similarly, in the non-native student corpus, Participant-oriented bundles are the least common bundles representing 5% types and tokens of the total bundles (see Table 4.77).

Table 4.77 Frequency & % Participant-oriented bundles (types & tokens) in the expert & the non-native student corpora

Functions	Types				Tokens			
	Experts	**%**	**N-Natives**	**%**	**Experts**	**%**	**N-Natives**	**%**
Participant-oriented bundles	**8**	**8.42%**	**13**	**5.37%**	**132**	**7.15%**	**442^{++++}**	**6.57%**
Stance features	6	5.26	10	4.13	102	3.41	382^{++++}	5.68
Engagement features	2	3.15	3	1.23	30	3.74	60^{++}	0.89

- In both the corpora, Participant-oriented bundles are the least common bundles representing 5% of the total bundles in each corpus.
- In both the corpora, majority of the Participant-oriented bundles are represented by Stace features.
- The non-native students have used significantly more Participant-oriented bundle tokens than expert writers.

4.6.3.4 Conclusion

Following are the main features of the bundle use in the expert and the non-native student corpora:

- The expert writers used more bundles for organizing the text and contextualizing information in the text. This is evident in the expert corpus as text-oriented bundles represent 51% of the total bundles, and most of which (36% of the total bundles) consist of text-oriented framing signals.
- The non-native writers used more bundles for describing research and giving procedural details. This is evident in the non-native corpus as research-oriented bundles account

for 57% of their total bundles, and most of which (39% of the total bundles) consist of procedural bundles.

4.7 Comparison of the bundle use in the native student and the non- native student corpora

In this section, the two corpora: the native and non-native student corpora will be compared. For the comparison, the top 20 bundles in each corpus will be compared at first. The relative frequencies will also be given for comparison as the two corpora are different in size. In the comparison of top 20 bundles, I will show the bundles that are similar in both the corpora. I will also compare the frequencies and the use of those bundles. This will lead to the comparison of structural and functional characteristics of the bundles in the native and non-native student corpora.

4.7.1 Comparison of the top 20 bundles in the native student and the non-native student corpora

For the comparison of the top 20 bundles used in both corpora will be presented in Table 4.78. For the comparison of the two different size corpora, it was essential to present the absolute frequencies (ABS) and the relative frequencies (REL) in the table. The similar bundles in both the corpora have been bolded in the list.

Table 4.78 Comparison of the top 20 bundles (types & tokens) in native& non-native student corpora

Rank	Native students	ABS	REL	Non-native students	ABS	REL
1.	(that) the use of the	52	16.61	agreed with the statement (that)	459	91.26
2.	**on the other hand** (the)	49	15.65	of the respondents agreed	180	35.78
3.	as a result of (the)	37	11.82	with the statement and	167	33.20
4.	**the results of the**	36	11.50	**on the other hand** (the)	168	33.40
5.	(is) that there is a	31	9.90	(is/are/were) of the view that	150	29.82
6.	as a function of	30	9.58	majority of the respondents	146	29.02
7.	it is important to	29	9.26	with the help of	127	25.25
8.	the extent to which	29	9.26	(in) the analysis of the (data)	123	24.45
9.	in line with the	28	8.94	of the present study (the)	114	22.66
10.	in relation to the	28	8.94	on the basis of (the)	112	22.26
11.	**as well as the**	26	8.30	(but) at the same time (the)	97	19.28
12.	it is possible that	25	7.98	in the use of	96	19.08
13.	for the purposes of	24	7.66	that most of the	96	19.08
14.	in the context of	24	7.66	**the results of the** (study)	89	17.69
15.	the way in which	24	7.66	in the form of	85	16.90
16.	in the case of	23	7.34	to find out the	84	16.70
17.	in terms of the	22	7.02	in the process of	83	16.50
18.	the total number of	21	6.70	the findings of the	75	14.91
19.	**in the field of**	20	6.39	**in the field of**	70	13.91
20.	in the same way	20	6.39	**as well as the**	67	13.32

As can be seen from Table 4.78, less than half of the top 20 bundles have been shared by the native and the non-native student corpora. Importantly, all four shared bundles have been used significantly more frequently in the non-native student corpus.

In the native student corpus, the majority of the top 20 bundles have been used for organizing the text, whereas in the non-native corpus the majority of the top 20 bundles have been used for describing research.

Considering the raw frequencies of top 20 bundles, the top 10% of bundle types in the native student corpus represent the 20% of the bundle tokens, whereas only top 4% bundle types in the non-native corpus represent more than quarter of the bundle tokens. This indicates that only a few bundle types have been frequently used in the non-native corpus.

Considering the relative frequencies of the top 20 bundles in both the corpora, only top 4/20 bundles in the native corpus occurred 10 or more times, whereas all the top 20 bundles in the non-native student corpus occurred 10 or more times, again showing the frequent use of bundles in the native student corpus.

The important features of the analysis of the top 20 bundles in the native and the non-native student corpora are as follows:

- The native and the non-native student corpora tend to be different in their use of the top 20 bundles as they share less than half of their top 20 bundles.
- The focus of bundle use appears to be different in both the corpora. The organization of the text is the focus of the native student corpus whereas the description of research seems to be the focus in the non-native student corpus.

4.7.2 Comparison of the bundle structure in the native student and the non-native student corpora

Table 4.79 presents the distribution of structural characteristics of lexical bundles used in native and non-native student corpora.

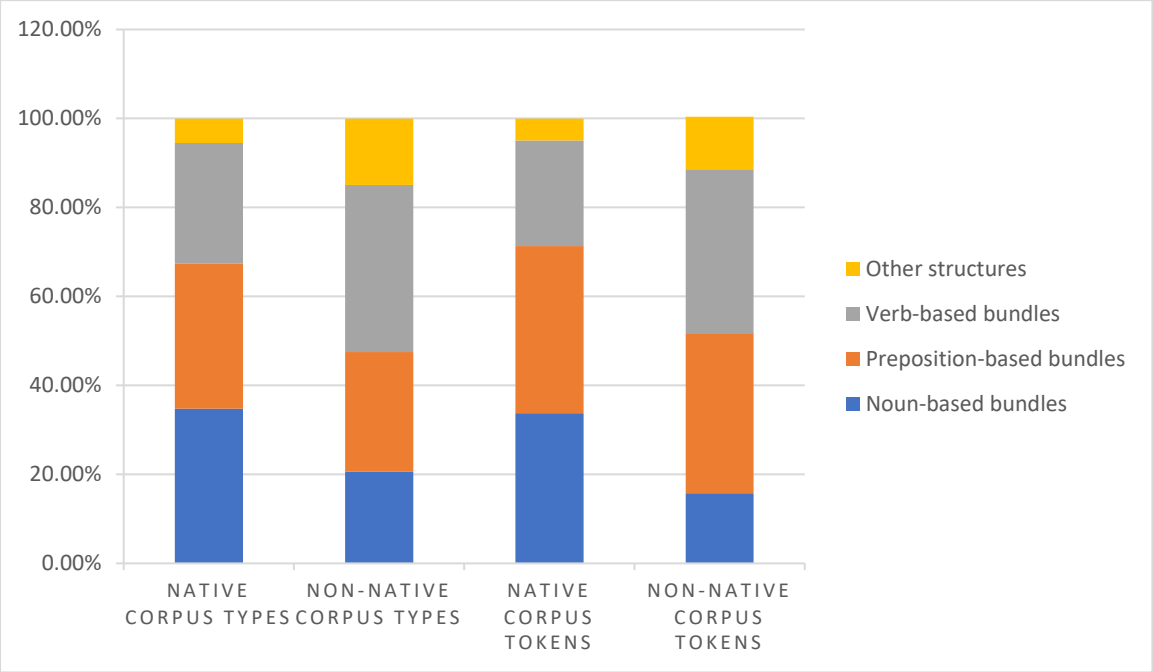
Table 4.79 Frequency & % of the bundle structure (types & tokens) in the native student & the non-native student corpora

Structure	Types				Tokens			
	Natives	%	N-Natives	%	Natives	%	N-Natives	%
Noun-based bundles	**32**	**34.77**	**50**	**20.65%**	**525**	**33.79**	**1016**$^{++++}$	**15.78%**
Noun-based bundles with of-phrase fragment	25	27.17	39	16.11	409	26.33%	859^{++++}	12.78%

Noun-based bundles with other post modifier fragment	7	7.60	11	4.54	116	7.46	202^{++++}	3.00
Preposition-based bundles	**30**	**32.6**	**65**	**26.85%**	**583**	**37.53**	**2417^{++++}**	**35.95%**
Preposition-based bundles with of-phrase fragment	17	18.47	37	15.28	331^{--}	21.31	1407^{++++}	20.93
Preposition-based bundles with other post modifier fragment	13	14.13	28	11.57	252	16.22	1010^{++++}	15.02
Verb-based bundles	**25**	**27.14**	**91^{++++}**	**37.58%**	**369**	**23.74**	**2477^{++++}**	**36.83%**
Copula be + NP/Adj phrase	4	4.34	21	8.67	85	5.47	355^{++++}	5.28
VP with active verb	-	0	11	4.54	-	0	755^{++++}	11.23
Anticipatory it + VP/Adj phrase	7	7.60	10	4.13	110	7.08	216^{++++}	3.21
Passive verb + PP fragment	6	6.52	13^{+}	5.37	79	5.08	204^{++++}	3.03
VP+ that clause fragment	2	2.17	17^{++}	7.02	31	1.99	605^{++++}	9.00
Verb/adj + to clause fragment	6	6.52	19^{+++}	7.85	64	4.12	342^{++++}	5.08
Other structures	**5**	**5.43**	**36^{++++}**	**14.87%**	**76**	**4.89**	**765^{++++}**	**11.83%**
	5	5.43	36^{++++}	14.87	76	4.89	765^{++++}	11.83
Total	**92^{++}**	**100%**	**242^{++++}**	**100%**	**1553^{++++}**	**100%**	**6720^{++++}**	**100%**

- In the native student corpus, the Noun-based and Preposition-based bundles are equally the most common bundles each representing 35% of the total bundles.
- In the non-native corpus, the Verb-based bundles are the most common bundles representing 38% types and tokens of the total bundles.
- The non-native students have used significantly more bundles (types and tokens) than the expert writers (see Figure 4.5).

Figure.4.5 Types & tokens of bundle structures in the native student & the non-native student corpora



4.7.2.1 Noun-based bundles

Noun-based bundles are the most common bundles (35% types and tokens of the total bundles) in the native student corpus. More than half of these bundles have been used for describing research, whereas these bundles represent 21% bundle types and 16% tokens of the total bundles in the non-native student corpus (see Table 4.80).

Table 4.80 Frequency & % of the Noun-based bundles (types & tokens) in the native student & the non-native student corpora

Structure	Types				Tokens			
	Natives	**%**	**N-Natives**	**%**	**Natives**	**%**	**N-Natives**	**%**
Noun-based bundles	**32**	**34.77**	**50**	**20.65%**	**525**	**33.79**	**1016^{++++}**	**15.78%**
Noun-based bundles with of-phrase fragment	25	27.17	39	16.11	409	26.33%	859^{++++}	12.78%
Noun-based bundles with other post modifier fragment	7	7.60	11	4.54	116	7.46	202^{++++}	3.00

Noun-based bundles with of-phrase fragment

In both the corpora, the majority of the Noun-based bundles with of-phrase fragment have been used for describing research, however, the non-native students have used significantly more bundle tokens than the native students.

Similarly, the Noun-phrase bundle frame 'the__ of the' was the most common Noun-based bundle frame in both the corpora, but the non-native students used significantly more bundles in this frame as compared to the expert writers. Table 4.81 presents the comparison of the use of this bundle frame in the two corpora:

Table 4.81 Comparison of the bundle frame 'the ___ of the' used in the native & the non-native student corpora

Structure	**Native Students**		**Non-atives**		**LOGL**
	ABS	**REL**	**ABS**	**REL**	
Noun-based bundles					
The ___ of the	242	40.90	694	137.98	273.03 (++++)

LEGEND (+++) Statistically significant overuse in non-native corpus (at $p < 0.0001$, critical value 15.13)

Noun based with other post-modifier fragment

- In both the corpora, the majority of the Noun-based bundles with other post-modifier fragment have been used for organizing text, however, the non-native students have used more tokens than the native students.
- The non-native students have used significantly more types and tokens of Noun-based bundles than the native students.

4.7.2.2 Preposition-based bundles with of-phrase fragment

In the native student corpus, Preposition-based bundles represent 33% types and 38% tokens of the total bundles, whereas in the non-native student corpus, these bundles represent 27% types and 36% tokens of the total bundles (see Table 4.82).

Table 4.82 Frequency & % of the Preposition-based bundles (types & tokens) in the native student & the non-native student corpora

Structure	Types				Tokens			
	Natives	%	N-Natives	%	Natives	%	N-Natives	%

Preposition-based bundles	**30**	**32.6**	**65**	**26.85%**	**583**	**37.53**	**2417^{++++}**	**35.95%**
Preposition-based bundles with of-phrase fragment	17	18.47	37	15.28	331^{--}	21.31	1407^{++++}	20.93
Preposition-based bundles with other post modifier fragment	13	14.13	28	11.57	252	16.22	1010^{++++}	15.02

In the native student corpus, the majority of Preposition-based bundles with of-phrase fragment (11% of the total bundles) have been used for organizing the text, whereas in the non-native student corpus the majority of these bundles have been used for describing research.

The analysis of the two Preposition-based bundle frames has been given below:

The most common PP-based frame, *'in the ___ of* 'was used significantly more frequently in the non-native corpus (see Table 4.83). On the other hand, the other most common PP-based structure *'at the ___ of '* that was used to identify place or time, has been used significantly more frequently in the expert corpus (see Table 4.84).

Table 4.83 Comparison of the bundle frame 'in the ___ of' in the native & the non-native student corpora

Structure	Native Students		Non-natives		LOGL
	ABS	REL	ABS	REL	
Preposition-based bundles					
'in the___ of'	254	50.20	460	91.46	61.52 (++++)

LEGEND (+++) Statistically significant overuse in non-native corpus (at p < 0.0001, critical value 15.13)

Table 4.84 Comparison of the bundle frame 'at the ___ of' in the native & the non-native student corpora

Structure	Native Students		Non-natives		LOGL
	ABS	REL	ABS	REL	
Preposition-based structure					
at he ___ of	110	21.74	65	12.92	11.44 (---)

(---) Statistically significant underuse in non-native corpus (at p < 0.0001, critical value 10.83)

Preposition-based bundles with other post-modifier fragment

In both the corpora, the majority of Preposition-based bundles with other post-modifier fragment have been used for organizing text.

Following are the important features of the comparison of preposition-based bundles in the two corpora:

- In both the corpora, the majority of Preposition-based bundles (21% of the total bundles in the native corpus, 15% of the total bundles in the non-native corpus) have been used for organizing text.
- In both the corpora, all the Preposition-based bundles with other post-modifier fragment have been used for organizing text.
- The non-native students have used significantly more types and tokens of Preposition-based bundles than were used by the native students.

4.7.2.3 Verb-based bundles

In the native corpus, the Verb-based bundles represent 27% types and tokens of the total bundles. The majority (17% of the total bundles) of these bundles have been used for describing research, whereas 10% types and tokens of the total bundles have been used for presenting writers' evaluation and for engaging the reader. In the non-native corpus, these are the most common bundles representing 38% and tokens of the total bundles. The majority (28% of the total bundles) of these bundles have been used for describing research, whereas 7% types and tokens of the total bundles have been used for presenting writers' evaluation and for engaging the reader. Table 4.85 presents the distribution of Verb-based bundles in native and non-native student corpora.

Table 4.85 Frequency & % of the Verb bundles (types & tokens) in the native student & the non-native student corpora

Structure	Types				Tokens			
	Natives	%	N-Natives	%	Natives	%	N-Natives	%
Verb-based bundles	**25**	**27.14**	**91^{++++}**	**37.58%**	**369**	**23.74**	**2477^{++++}**	**36.83%**
Copula be + NP/Adj phrase	4	4.34	21^{++++}	8.67	85	5.47	355^{++++}	5.28
VP with active verb	-	0	11	4.54	-	0	755^{++++}	11.23

Anticipatory it + VP/Adj phrase	7	7.60	10	4.13	110	7.08	216	3.21
Passive verb + PP fragment	6	6.52	13	5.37	79	5.08	204^{++}	3.03
VP+ that clause fragment	2	2.17	17^{++}	7.02	31	1.99	605^{++++}	9.00
Verb/adj + to clause fragment	6	6.52	19^{+++}	7.85	64	4.12	342^{++++}	5.08

- In both the corpora, the majority of the Verb-based bundles have been used for describing research.
- In the native corpus, the Verb-based bundles represent 27% of the total bundles, whereas these bundles are the most common bundles in the non-native corpus representing 38% of the total bundles.
- The non-native students have used more than twice as many Verb-based bundles (28% of the total bundles) for describing research as were used by the native students (12% of the total bundles).
- The non-native students have used significantly more types and tokens of Verb-based bundles than the native students.

4.7.2.4 Other structures

In the native student corpus, other structures represent 5% of total bundles, the majority of which (3% of the total bundles) have been used for describing research.

Table 4.86 Frequency & % of Other Structures (types & tokens) in the native student & the non-native student corpora

Structure	Types				Tokens			
	Natives	**%**	**N-Natives**	**%**	**Natives**	**%**	**N-Natives**	**%**
Other Structures	**5**	**5.43**	**36^{++++}**	**5.43**	**76**	**4.89**	**765^{++++}**	**11.83%**
	5	5.43	36^{++++}	5.43	76	4.89	765^{++++}	11.83

In the non-native corpus, the other structures represent 15% types and tokens of the total bundles. The majority of these bundles (10% of the total bundles) have been used for describing research, e.g., *objective of the study, well aware of the* etc. The non-native students have used significantly more other structures than native students.

- In the native corpus, Noun-based bundles and the Preposition-based bundles are equally the most common bundles, each representing 35% types and tokens of the total bundles.

- In the non-native corpus, the Verb-based bundles are the most common bundles representing 38% types and tokens of the total bundles.
- In both the corpora, the majority of the Noun-based bundles (24% of the total bundles in the native corpus, and 16% (of the total bundles in the non-native corpus) have been used for describing research.
- In both the corpora, the majority of the Preposition-based bundles 21% (of the total bundles in the native corpus), and 15% (of the total bundles in the non-native corpus) have been used for organizing text.
- The non-native students have used more than twice as many verb-based bundles (28% of the total bundles) for describing research as were used by the native students (12% of the total bundles).
- The non-native students have used significantly more tokens than the native students.

4.7.3 Comparison of the bundle functions in the native and the non-native student corpora

In this section I will compare the functional characteristics of bundles in the native and the non-native student corpora. At first, the distribution of the functional categories of bundles in both the corpora will be compared. The distribution of functional characteristics will follow the

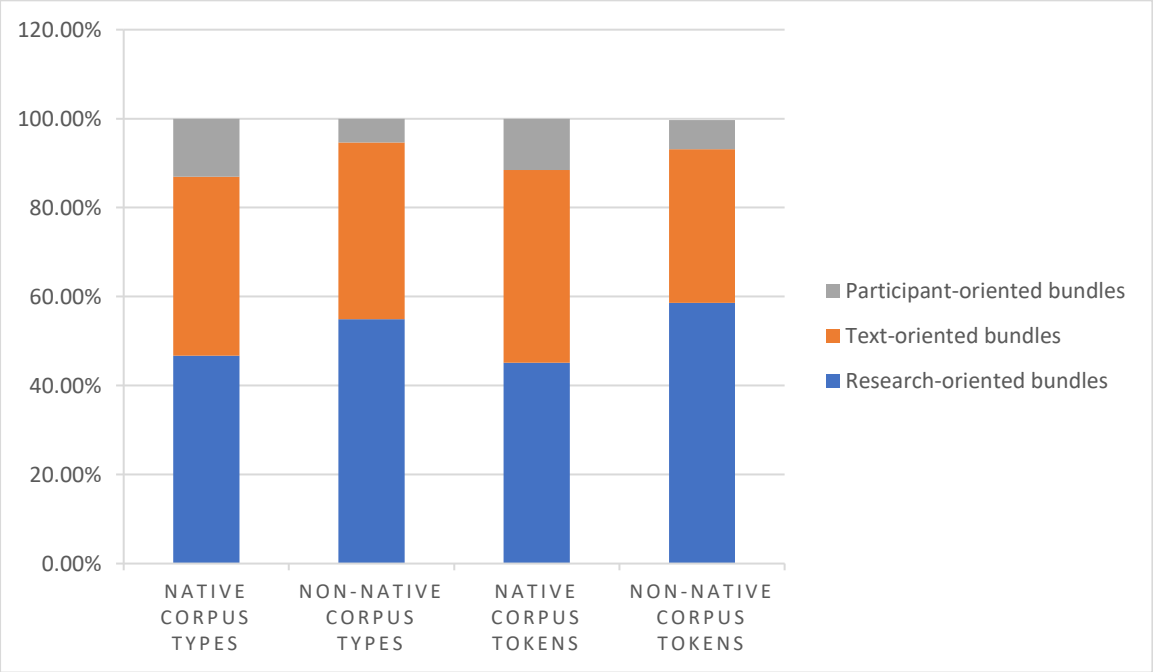
detailed analysis of the comparison of the use of bundles in both the corpora. In Table 4.87, bundle types and tokens have been compared. The results of loglikelihood test have been indicated in the table where there are significant differences in the use of bundles between the native and the non-native student corpora. Table 4.87 presents the table that displays the distribution of bundle types and tokens in the expert and the native student corpora.

Table 4.87 Frequency & % of bundle functions (types & tokens) in the native student & the non-native student corpora

	Types				**Tokens**			
Functions	**Natives**	**%**	**N-Natives**	**%**	**Natives**	**%**	**N-Natives**	**%**
Research-oriented bundles	**43**	**46.73%**	**138**$^{++++}$	**54.95%**	**635**	**45.13%**	**3857**$^{++++}$	**58.86%**
Location	3	3.26	6	2.47	45	2.89	154^{++++}	2.29
Procedure	16	17.39	94^{++++}	36.36	184	18.73	2729^{++++}	41.91
Quantification	13	14.13	19	7.85	207	13.32	547^{++++}	8.13
Description	11	11.95	19	8.26	199	10.17	427^{++++}	6.51
Text-oriented bundles	**37**	**40.20%**	**91**	**39.66%**	**739**	**43.33%**	**2421**$^{++++}$	**34.55%**
Transition signals	9	9.78	12	4.54	164	10.56	455^{++++}	6.60
Resultative signals	5	5.43	15	6.19	108	6.95	440^{++++}	6.54
Structuring signals	5	5.43	21^{+}	13.63	74	2.18	428^{++++}	9.22
Framing signals	18	19.56	43	15.28	393	25.30	1098^{++++}	12.17
Participant-oriented bundles	**12**	**13.04%**	**13**	**5.37%**	**179**	**11.52%**	**442**$^{++++}$	**6.57%**
Stance features	6	6.52	10	4.13	113	5.40	382^{++++}	5.68
Engagement features	6	6.52	3	1.23	66	6.11	60^{--}	0.89
Total	**92**	**100%**	**242**$^{++++}$	**100%**	**1553**	**45.13%**	**6720**	**100%**

In both the corpora, Research-oriented bundles are the most common bundles, indicating that both the native and the non-native students focus on describing research. However, the non-native students have used significantly more types and tokens than the native students. Figure 4.6 shows the distribution of bundles functions in native and non-native student corpora.

Figure.4.6 Types & tokens of bundle functions in the native student & the non-native student corpora



4.7.3.1 Research-oriented bundles

In the native student corpus, research-oriented bundles represent 47% of the total bundles, whereas in the non-native corpus, these bundles represent 55% types and tokens of the total bundles (see Table 4.88).

Table 4.88 Frequency & % of Research-oriented bundles (types & tokens) in the native student & the non-native student corpora

Functions	Types				Tokens			
	Natives	**%**	**N-Natives**	**%**	**Natives**	**%**	**N-Natives**	**%**
Research-oriented bundles	**43**	**46.73%**	**138^{++++}**	**54.95%**	**635**	**45.13%**	**3857^{++++}**	**58.86%**
Location	3	3.26	6	2.47	45	2.89	154^{++++}	2.29
Procedure	16	17.39	94^{++++}	36.36	184	18.73	2729^{++++}	41.91
Quantification	13	14.13	19	7.85	207	13.32	547^{++++}	8.13
Description	11	11.95	19	8.26	199	10.17	427^{++++}	6.51

Following are the main features of Research-oriented bundles used in native and non-native student corpora:

- In both the corpora, Research-oriented bundles are the most common bundles representing 47% of the total bundles in the native student corpus, and 55% of the total bundles in the non-native student corpus.
- In both the corpora, Research-oriented bundles represent the majority of the bundles, however, the non-native students have used twice as many Procedure bundles (38% of the total bundles) than were used by the native students (17% of the total bundles).
- So, the non-native students have used far more Research-oriented bundles for describing procedures of research.

4.7.3.2 Text-oriented bundles

In both the corpora, text-oriented bundles represent 40% types and tokens of the total bundles (see Table 4.89).

Table 4.89 Frequency & % of Text-oriented bundles (types & tokens) in the native student & the non-native student corpora

	Types				Tokens			
Functions	**Natives**	**%**	**N-Natives**	**%**	**Natives**	**%**	**N-Natives**	**%**
Text-oriented bundles	**37**	**40.20%**	**91**	**39.66%**	**739**	**43.33%**	**2421^{++++}**	**34.55%**
Transition signals	9	9.78	12	4.54	164	10.56	455^{++++}	6.60

Resultative signals	5	5.43	15	6.19	108	6.95	440^{++++}	6.54
Structuring signals	5	5.43	21^{+}	13.63	74	2.18	428^{++++}	9.22
Framing signals	18	19.56	43	15.28	393	25.30	1098^{++++}	12.17

Following are the main features of Text-oriented bundles used in native and non-native student corpora:

- In both corpora, Text-oriented bundles are equally common representing 40% types and tokens of the total bundles in each corpus.
- In both the corpora, Framing signals are the most common Text-oriented bundles.
- The non-native students have used significantly more Structuring signals than the native students, used for referring to the study, representing 14% types and tokens of the total bundles as were used by the native students representing only 2% types and tokens of the total bundles.

4.7.3.3 Participant-oriented bundles

These are the least common bundles in the native and the non-native student corpora representing 13% types and tokens of the total bundles in the native student corpus and 5% types and tokens of the total bundles in the non-native student corpus (see Table 4.90).

Table 4.90 Frequency & % of Participant-oriented bundles (types & tokens) in the native student & the non-native student corpora

Functions	Types				Tokens			
	Natives	%	N-Natives	%	Natives	%	N-Natives	%
Participant-oriented bundles	**12**	**13.04%**	**13**	**5.37%**	**179**	**11.52%**	**442^{++++}**	**6.57%**
Stance features	6	6.52	10	4.13	113	5.40	382^{++++}	5.68
Engagement features	6	6.52	3	1.23	66	6.11	60^{--}	0.89

Following are the main features of Participant-oriented bundles used in native and non-native student corpora:

- In both corpora, the Participant-oriented bundles are the least common bundles.
- The native students have used twice as more of these bundles (13% of the total bundle) as have been used by the non-native students (5% of the total bundles).

- The non-native students have used significantly fewer Engagement features than the native students.

Following are the main features of bundles use in native and non-native student corpora:

- The non-native students have used significantly more bundle types and tokens than the native students.
- The native and the non-native students have used equal proportion of Research-oriented bundles.
- The non-native students have used significantly more types and tokens of Procedure, Description and Quantifying bundles than the native students.
- The non-native students have used significantly more types and tokens of Transition signals, Resultative signals, Structuring signals, and Framing signals.
- The native students have used significantly more tokens of Engagement features than the non-native students.

4.8 Conclusion

In this chapter, the results and analysis of the expert corpus, the native student corpus, and the non-native student corpus were presented. The main findings of the analysis are as follows:

- In the use of bundle types and tokens, the native and the non-native students used significantly more bundle types and tokens than the expert writers. And the non-native students used significantly more bundle types and tokens than the expert writers and the native students.
- In the expert and the native student corpora, the Noun-based and the Preposition-based bundles are two-third of the total bundles, whereas in the non-native student corpus, these bundles represent half the total bundles.
- In the use of Verb-based bundles, the native and the non-native students have used significantly more bundle types and tokens than the expert writers. And the non-native students have used significantly more Verb-based bundles than the expert writers and the native students.
- The analysis of the three corpora reveals that the expert writers have used far more bundles for organizing the text than the native students and the non-native students. The

text-organizing framing bundles represent nearly half of the total bundles in the expert corpus.

The expert writers focused on organizing the text and contextualizing new information in the text, the non-native students focused on describing research, in particular describing the procedures of research. The expert writers used far more Preposition-based bundles representing 50% types and tokens of the total bundles, the majority (34% of the total bundles) of which were used for organizing text. The expert writers also used far more Text-oriented bundles representing 50% of the total bundles. In contrast, the native and the non-native student corpora shows that both the student groups have focused on describing research rather than organizing text. The most frequent bundle in the native student corpus, *the use of the,* was used for describing research. In the native student corpus, Noun-based bundles representing 35% of the total bundles, and Preposition-based bundles representing 33% of the total bundles are equally the most common bundles. The native student used a majority of Verb-based bundles (12% of the total bundles) and Noun-based bundles (25% of the total bundles) for describing research. The non-native students also used the majority of Verb-based bundles (28% of the total bundles) and Noun-based bundles (19% of the total bundles) were used for describing research.

In both student corpora, Research-oriented bundles were the most common bundles representing 47% of the total bundles in the native student corpus, and 55% of the total bundles in the non-native corpora. So, both the native and the non-native students have used bundle structures and functions for describing research rather than organizing text.

To conclude, the expert writers have used lexical bundles differently from the native and the non-native students. The expert writers have used far more Preposition-based bundles and Text-oriented bundles and have given importance to contextualizing information in the text. In contrast, the native and the non-native students have used more Noun-based and Verb-based bundles and have used bundles for describing research rather than organizing the text. But among the three corpora, the non-native students have used bundles more frequently than the experts and the native students, and the non-native students have used significantly more bundles for describing research and the procedure than the expert writers and the native students.

.

Chapter 5

Discussion

5.1 Introduction

The aim of this research was to compare the use of lexical bundles in the academic writing of expert writers, native English students, and non-native English students in the field of Applied Linguistics. This chapter discusses the distribution and role of lexical bundles across the three corpora. The discussion of the findings has been organized according to research questions of this study. The introductory paragraph is followed by a summary of the main findings and features of three corpora. There are three sections of this chapter: the first section (5.1) is based on the main findings. Second section (5.2) answers the research question no.1 and compares the frequencies of structural characteristics of lexical bundles across the three corpora. Section 5.3 answers the research question no.2, that is based on the comparison of discourse functions of lexical bundles across the three corpora.

The following are the findings and the main features of the three corpora:

Phrasal bundles

The Phrasal bundles that are characteristic of academic writing were found to be the most common bundles across the three corpora (Biber,2006; Biber & Barbieri, 2007). However, these bundles were far more common in the expert and the native student corpora representing 70% and 67% types and tokens respectively, whereas these bundles only represented 47% of types and tokens in the non-native corpus.

Verb-based bundles

The verb-based bundles were far more common in the non-native student corpus than in the expert writers and the native student corpora. However, the expert writers and the native students used the majority of verb-based bundles for hedging, whereas the non-native students used these bundles for describing research as well as hedging. The expert and the native student writers used similar bundle types for hedging and showed better control of hedging devices. The non-native students used different bundles for hedging.

Research-oriented bundles

Research-oriented bundles were far more common in the non-native student corpus than the native students and the expert writers' corpora. The non-native students used significantly more

bundle types and tokens for describing the data collection procedures, and the use of different aspects of research. They also used a large number of bundles for highlighting the importance of the study.

Text-oriented bundles

Text-oriented bundles are far more common in the expert corpus, whereas the distribution of these bundles was similar in the native and the non-native student corpora. The expert writers and the native students used similar bundle types for contextualizing new information, whereas the non-native students used different bundles for this purpose. The non-native students seem to struggle with the use of idiomatic bundles, e.g., *with respect to the, in line with the, in the face of* etc. used in the expert and the native student corpora. The expert and the native student writers used a small number of bundles for referring to the section or chapter of the study, whereas the non-native students used significantly more types for this purpose. The majority of these types had the repetition of the words ‘study’ and ‘present’.

Participant-oriented bundles

Participant-oriented bundles were the least common bundles across the three corpora, however, the non-native students used different Participant-oriented bundle types than the expert writers and the native students.

Frequency of bundles

The non-native students used significantly more bundle types and tokens than the expert writers and the native students, though there was no significant difference in the use of bundle types in most of the sub-categories. This indicates the tendency to repeat a limited number of bundles in their texts, which marks the non-native writing as different from formal academic writing.

There were other aspects of bundle use in which the students, both native and non-native, were similar:

- Quantifying

The native and the non-native students used significantly more quantifying bundles than the expert writers.

- Resultative signals

The native and the non-native student writers used more Resultative bundles than the expert writers.

- Use of informal bundles

Moreover, they also used some informal Quantifying bundles, e.g., *a lot of time, between two or more, the vast majority of the, all over the world, it is difficult to.*

To conclude, the results and the analysis of the three corpora reveals that the expert writers and the native students are generally similar in using bundles, which might suggest that nativeness does play a role in the use of lexical bundles in academic writing. At the same time, in some respects, both the native and the non-native students use bundles differently from the expert writers, which might suggest that the use of lexical bundles in academic writing is not acquired naturally, rather it has to be taught to the native and the non-native students alike.

Table 5.1 presents the main features of the lexical bundle used across the three corpora.

Table 5.1 Main characteristics of bundles in the expert, native student & the non-native student corpora

Characteristics	**Experts**	**Non-experts (students)**	
		Native	**Non-native**
Structural characteristics	Phrasal	Phrasal	Phrasal + clausal
Structural characteristics	Varied and formal use of hedging	Varied and formal use of hedging	Varied and informal use of hedging
Main focus	Organising text	Describing research	Describing research
Research procedures	Less common	Less common	Highly Common
Quantifying	rare	Varied and frequent	Varied and frequent
Referring to results	Rare	frequent	highly frequent
Referring to study	common	rare	Varied and frequent
Contextualizing new information	varied, frequent	varied, frequent	varied, frequent
Stance	Reader friendly	Reader friendly	Less reader friendly

In the next section, I will discuss the answer to research question no.1:

RQ.1 (a) What are the most frequent bundle structural categories in the expert writer's corpus, native students' corpus, and non-native students' corpus respectively?

(b) How does frequency of structural categories compare across the three corpora?

5.2 Research Question no.1: Frequency of structural categories

In this section I will discuss the distribution and role of structural characteristics of lexical bundles (phrasal and clausal bundles) across the three corpora.

5.2.1 Phrasal and clausal bundles in academic writing

The Phrasal bundles, i.e., Noun-based and Preposition-based bundles (by contrast with clausal bundles, i.e., Verb-based bundles), are the most common bundles in written academic discourse (Biber et al., 1999; 2004; Chen & Baker, 2010). This is also found across the three corpora in the current study. However, Phrasal bundles were far more common in the expert and the native student corpus. In the expert corpus, phrasal bundles represent 75% types and 80% tokens of the total bundles. In the native student corpus, Phrasal bundles represent 67% types and 71% tokens of the total bundles, whereas in the non-native corpus, Phrasal bundles represented 48% types and 58% tokens of the total bundles. So, Phrasal bundles are far more common in the

expert and the native student corpora than in the non-native student corpora. One of the reasons for this difference might be the larger size of non-native student dissertations that warrant more explanation and description that leads to the use of more clausal bundles. The average size of the non-native student dissertations is twice as big as the native students' dissertations, and three times as big as the research articles. Therefore, non-native students have a lot more space to report on their study than the other two groups; they spend more time in describing the research process in detail.

The results of Bychkovska and Lee (2017) are similar to the findings of this study. They also found that native English students used more Phrasal bundles than the non-native English students, and conversely, used less Clausal bundles. Pan et. al. (2016) found that the majority of the bundles used in the native expert writing were phrasal, whereas in the non-native expert writing these bundles were not as frequent. This shows that the non-native writers tend to use more clausal bundles even at the more advanced level. This means that the nativeness and expertness both might be at play with respect to use of clausal bundles in the non-native writing.

Previous research has shown that the majority of bundles used in native expert writing are Phrasal (Chen & Baker, 2010; Pan et.al., 2016). However, with respect to native and non-native student writing, previous research has presented mixed findings. There are studies that showed

that both native and non-native students used more clausal bundles than Phrasal bundles (Chen & Baker, 2010; Shin, 2019). In terms of the use of Phrasal bundles by native students, the findings of Chen and Baker (2010) are different from the current study. One of the reasons for this difference might be that in Chen and Baker (2010), the student corpora consist of samples from a wide range of disciplines, whereas in the current study the native student data is based on the samples from the field of Applied linguistics only. Moreover, Chen and Baker (2010) looked at BAWE texts which are predominantly undergraduate assignments, not MA dissertations. The previous research has shown that the use of bundles is discipline specific (Cortes, 2004; Hyland, 2008a). For example, Cortes (2004) found that the students of History made a frequent use of Clausal bundles, whereas the students of Biology made frequent use of Phrasal bundles. So, the more frequent use of Clausal bundles in native and the non-native student writing might be due to data based on the disciplines that make more frequent use of clausal bundles.

Shin (2019) also found very different results from the current study. She compared the use of bundles in argumentative essays written by native and non-native English students. She found that in both the native and the non-native student corpora, the majority of the bundles were clausal. The difference again might be due to different genres used in both studies. The student

essays are more explanatory and descriptive in nature, whereas the dissertations used in the current study require more contextualization, organization and coherence, therefore, more Phrasal bundles are used in this genre.

So, previous research shows that native expert writers use more Phrasal bundles, and the non-native students use more clausal bundles. As has been discussed before, one of the reasons for this less frequent use of Phrasal bundles might be the different genre and the larger size of the non-native student dissertations. Previous research has shown that some disciplines use a very small number of clausal bundles (Cortes, 2004).

Another reason might be that the expert writers are more focused on giving information rather than explaining it, therefore they use more phrasal bundles which provide information focus (Pan et.al., 2016). On the other hand, the novice students are under pressure to prove their academic credentials, so they tend to focus on describing and explaining the research, which requires more frequent use of Verb-based bundles (Hyland, 2008a; 2008b).

5.2.2 Verb-based bundles

The non-native students have used significantly more types and tokens of Verb-based bundles than the expert writers and the native students. In the non-native corpus, Verb-based bundles

were the most common bundles representing 38% bundle types and 42% tokens of the total bundles. The native students also used significantly more Verb-based bundles than the expert writers.

Some Verb-based bundles are used for hedging which is an important technique used in academic writing to present author's opinion with a tone of neutrality (Pan et.al., 2016). Academic writers in English use hedging to present their evaluation while sounding more neutral, and to appear more polite and indirect, e.g., *it is possible that, it should be noted* etc. Hedging is carried out through the use of nouns, e.g., *there is no evidence, there is little evidence* etc., or through verb-based bundle structures such as Anticipator it + verb/adjective phrase, e.g., *it is possible to, it is worth noting* etc., or Copula be + noun/adjective phrase bundles, e.g., *are more likely to, is evident from the* etc. (Chen & Baker, 2010).

In the following section, I will discuss the sub-types of Verb-based bundles that have been used for hedging in the expert, native and the non-native student corpora.

5.2.2.1 Copula be + noun/adjective

In the expert, native, and the non-native student corpora, similar types of Copula be + noun/adjective phrase bundles were used for hedging. In the expert corpus, there were only two

bundle types used for hedging, e.g., *there is a need, are more likely to*. Both these types were also found in the native and the non-native student corpora. However, the expert writers and the native students used bundles to present likelihood, e.g., *is likely to be, are likely to be*. These bundle types were not used by the non-native students. Instead, the non-native students used the bundle, *is evident from the*. The use of bundles like *is likely to be, are likely to be*, are important as they present the proposition with a neutral tone (Adel & Erman, 2012; Chen & Baker, 2010). The absence of these bundles in the non-native corpus might suggest that they might use limited variety of lexical bundles for providing their own evaluation in their writing, as they are students, and it might therefore not be appropriate for them to provide their evaluation and thus pretend that they are experts (Hyland, 2008b). This limited use of hedging devices might also be linked with the teaching practices in the Pakistani context which does not encourage the students to use their own evaluation (see Section 2.4.5).

The educational background of the Pakistani students might affect the use of hedging devices. As has been mentioned that the use of hedging devices reflects the evaluation and critical thinking of the writers. Khan (2011, p.111) observed that the Pakistani students are not encouraged to use their creative as well as critical thinking in their academic writing. She notes 'It is evident that in Pakistani schools, the pedagogy, the curriculum and the assessment system

do not provide freedom for self-expression which is a prerequisite for creativity.' She believes that the English language testing system in Pakistani education system does not encourage the critical thinking in the students. This might be one of the reasons that the Pakistani postgraduate students use fewer lexical bundles that reflect their critical thinking.

The linguistic background of the Pakistani students might also affect the use of lexical bundles in the writing of Pakistani students. Hussain et al. (2013) have noted the impact of Urdu language, the official language in Pakistan, in the academic writing of Pakistani postgraduate students. This impact has been observed in the use of collocations, e.g., *fatal misunderstanding*, *the excitement was drowned*. Both these collocations have been directly translated from Urdu language. In the use of lexical bundles, an omission of the preposition was observed, e.g., *(in) the other hand*, which is also due to direct translation from Urdu language (Hussain et al., 2013)

The previous research on lexical bundles has noted the impact of culture and L1 on the use of lexical bundles in the writing of Chinese student (Bychkovska & Lee, 2017; Huang, 2015; Pang, 2009). The Chinese students also used some bundles like *more and more, a lot of people* that reflect their collective culture (Bychkovska & Lee, 2017; Huang, 2015; Pang, 2009). So, examples from the Pakistani and the Chinese students' academic writing provide a hint that the nativeness, and the cultural background might affect the use of lexical bundles.

The results of Chen and Baker (2010) are different from the current study. They found that non-native English students were different from native English students and native expert writers in the use of bundles for hedging. They found that native English expert writers and native English students used a greater variety of lexical bundles, such as Copula be + noun/adjective phrase bundles e.g., *are likely to be, is likely to be, are more likely to* etc. In contrast, the non-native students used only one bundle type, *are more likely to*. But in the current study, it was found that the non-native students did not use this bundle but instead used a variety of bundles for hedging. However, they did not use bundles that show likelihood which were used by both the experts and the native students. This might be due to the fact that the non-native students use limited variety of bundles for presenting their evaluation in research writing. The other possibility is that the non-native students might not be well equipped with the convention of hedging used in academic writing. This is also supported by the fact that the expert writers used Copula be + verb/adjective phrase bundles only for hedging, whereas the non-native students used these bundles for describing research as well (see Table 5.2). This might suggest that the native students are still learning the important role of these bundles for hedging purposes in academic writing.

Table 5.2 Frequency of ***Copula be + noun/adjective phrase bundles*** *(types & tokens) in the expert & the native student corpora*

Expert writers	Tokens	Native Students	Tokens	Non-native students	Tokens
there is a need	13	***there is also a***	13	there is a need (to)	20
are more likely to	11	***is likely to be***	12	is evident from the	11
		are likely to be	11		

5.2.2.2 Anticipatory it + verb/adj phrase' bundles

Anticipatory it + verb/ adj. phrase bundles were also used for hedging across the three corpora. In the expert and the native student corpora, 4 bundle types were shared, however, the non-native students shared only one bundle type, *it is important to*. The non-native students have used different types of bundles that were not used by the expert writers and the native students. The following bundles are unique to the non-native student writing: *it is necessary to, it is needed to, it is believed that, it has been observed that*. The reason for this difference might be

that the non-native students use these bundles for different purposes, i.e., giving recommendations, showing the need for something etc. *it is necessary to* and *it is needed to*. They tended to use these bundles for emphasis rather than for hedging, and also for presenting common beliefs and observations (*it has been observed that, it is believed that).*

In the use of 'Anticipatory it' bundles, the results of the previous studies are somewhat different from the current study as these studies found that the native expert writers and the native students used a greater variety of 'Anticipatory it' bundles, whereas these bundles were rare in the non-native student corpus (Adel & Erman, 2012; Chen & Baker, 2010). But these studies agree that the non-native students used a limited number of bundle types of the 'Anticipatory it' structure. Chen and Baker (2010) found that the non-native students use bundles like *it has been suggested that, it has been believed that*. Adel and Erman (2012), found three such bundles, e.g., *it is hard to, it is easy to, it is clear that*, which were used by the non-native students only. Pan et.al. (2016) found two such bundles, e.g., *it is difficult to, it is easy to,* used in the non-native expert writing.

This shows that the non-native students and the expert writers used these bundles to state perceptions, beliefs and evaluations, rather than to give their own interpretation by introducing possibility or likelihood. Bundles such as *it is possible to, it is possible that, is likely to be* are

used for giving interpretation, but the non-native students tend to use evaluative bundles instead (see Table 5.3). These types of bundles make non-native writing appear more judgemental.

The use of evaluative 'Anticipatory it' bundles might be due to the fact the non-native students are not well trained in giving their interpretation, therefore they end up being too judgemental. The lack of the knowledge of research conventions might be the reason behind using these types of bundles. The other reasons might be, as has been mentioned before, that the non-native students might think that it is not their place to interpret as they are not experts yet.

Table 5.3 Frequency of Anticipatory it + verb/adj phrase' bundles (types & tokens) in the expert & the native student corpora

Expert writers	**Tokens**	**Native students**		Non-native students	**Tokens**
it is important to (note that)	39	**it is important to**	29	it has been observed (that)	34
it is possible that	19	**it is possible that**	25	it can be seen	32

it should be noted	12	it is possible to	12	it was observed that (the)	29
it can be argued + that	10	it is worth noting	12	it is important to	24
it is possible to	10	it is difficult to	11	it refers to the	23
		it is interesting to	11	it is believed that	21
		it should be noted	10	it is necessary to	21
				it is also a	11
				it is needed to	11

5.2.2.3 Passive verb with prepositional phrase fragment

The bundles with Passive verb + prepositional phrase fragment are considered important in academic writing as they are used for adopting a neutral tone in the text. (Biber et al., 2004) Similar bundle types have been used across the three corpora in this category. However, the native and the non-native students have used these bundles significantly more frequently than the expert writers (see Table 5.4). This might be linked to the non-native students' general tendency to use more Verb-based bundles. The larger size of the non-native students' dissertations might be another potential reason for more frequent use of these bundles as the non-native students have more space to describe research. It could also be because these bundles fit to their tendency to maintain authorial anonymity. So, by using these bundles more frequently, the non-native students might want to appear more neutral.

Table 5.4 Bundles with Passive verb + Preposition-based bundles fragment in expert, native and the non-native corpora

Expert writers	**Tokens**	**Native Students**	**Tokens**	**Non-native students**	**Tokens**

can be seen in	18	as can be seen (from/in)	18	can be used to	24
can be seen as	16	can be found in (the)	15	can be seen in (the)	17
(that) can be used to	16	can be used to	13	can be used in	15
		can be used as	10	can be used for	12
				can be divided into	10

Previous research also finds that non-native students use more Verb-based bundles with Passive verb + prepositional phrase fragment (Chen & Baker, 2010; Hyland, 2008a; 2008b; Lu & Deng, 2019). Pan et.al. (2019) also found that the non-native experts used more Verb-based bundles, especially the Passive verbs + prepositional phrase fragment, than the native English writers.

5.2.2.4 Bundles with Active Verb

The non-native students have used 11 types and 755 tokens of Verb-based bundles with active verb, whereas the expert writers and the native students used none of these bundles. The absence of these bundles in the expert and the native student corpora might be because these bundles are used for explaining procedures, which is not very common in the expert and the native student corpora. The much larger size of the non-native student dissertations might be another possible reason, as they have more space to explain research.

The findings of previous research are mixed with regards to the use of active verb in native and non-native students. There are studies that concur with the findings of the current study (Bychkovska & Lee, 2017; Pan et.al., 2016). Bychkovska and Lee (2017) found that the native students did not use bundles with active verbs. Similarly, Pan et.al. (2016) found that the non-native expert writers used none of the bundles with active verbs. On the other hand, there are studies that found that non-native students used significantly more types and tokens of bundles with active verbs than were used by native students and expert writers (Chen & Baker, 2010; Lu & Deng, 2019). But the difference in the findings of the current study and these studies might be due to different research designs. For example, the size of texts and the genre used in

Chen and Baker (2010) is different from the current study. In Lu and Deng (2019) the corpus is based on PhD abstracts, in which the native students might tend to use more active verbs.

So, the overall findings suggest that the native students and the expert writers either use a very small number of bundles with active verbs or they do not use these bundles at all in their writing. But the non-native students make use of bundles with active verb as they might need to use multiple bundle types for explaining and describing research.

In the next section, I will discuss the answer to research question no.2:

RQ.2 (a) What are the most frequent bundle functional categories in expert writers' corpus, native students' corpus, and non-native students' corpus respectively?

(b) How does frequency of functional categories compare across the three corpora?

5.3 Research Question no.2: Frequency of functional categories

In this section I will discuss the distribution and role of bundle functions across the three corpora.

5.3.1 Research-oriented bundles across the three corpora

Research-oriented bundles are used for describing research. In academic writing, these bundles are not as frequent as text-oriented bundle (Biber et al., 1999; Cortes, 2004; Hyland, 2008a;

2008b). The use of Research-oriented bundles is different across the three corpora in this study. The native and the non-native students have used significantly more types and tokens of Research-oriented bundles than the expert writers. Additionally, the non-native students used significantly more bundle types and tokens than the expert writers and the native students. In the expert corpus, these bundles represent 41% types and tokens of the total bundles, whereas they are 47% of the total bundles in the native student corpus. The non-native students have used far more Research-oriented bundles (58% of the total bundles). One of the reasons for more frequent use of these bundles in the native and the non-native student corpus might be that the students have to use more bundles for describing research as the length of dissertation is much larger than the research articles. The other reason might be that the native and the non-native students want to make things clear by describing and explaining the research in detail, therefore they use more Research-oriented bundles. Yet another reason is that the non-native Pakistan students have less engagement with writing as social interaction. The non-native Pakistani students are not encouraged to use their critical thinking in their academic writing. That is why they tend to be descriptive in their writing (Haider, 2012). There is also a possibility that the use of these bundles might be a requirement for getting a good mark (see Section 2.4.4). As the Postgraduate thesis is written for assessment purposes, the students do not see

themselves as experts, and tend to be reluctant in making strong claims in their writing. That is why they are under pressure to display their understanding of the field, and to show that they can appropriately apply the research methods in their area of research. Thus, they tend to focus on providing procedural details. As a results, they end up using high proportion of research-oriented bundles in their writing. (Hyland, 2008a; 2008b)

The results of previous studies are mixed with respect to the use of Research-oriented bundles by experts, native non-native students. There are studies that agree with the findings of the current study (Chen & Baker, 2010; Hyland, 2008a). For example, Hyland (2008a) found that non-native Masters students used twice as many Research-oriented bundles as expert writers. Chen and Baker (2010) found that novice students used twice as many Discourse organizers (most of which correspond to Research-oriented bundles) than native expert writers.

On the other hand, Bychkovska and Lee (2017) found that native students used more Discourse organizers than non-native students.

One of the reasons for these differences might be differences in genre. student dissertations warrant more elaboration and explanation than research article. The results of Pan et.al. (2016) lend credence to the likelihood of this reason. They found that the distribution of Research-oriented bundles was almost similar in the native English expert writers and the non-native

English expert writers, representing 42% types and 48% tokens of the total bundle in the native expert corpus, whereas 38% types and 48% tokens of the total bundles in the non-native expert corpus. This study was based on the same genre, research articles.

In the next section, I will discuss the findings in the different sub-categories of Research-oriented bundles.

5.3.1.1 Procedure bundles

The non-native students have used significantly more types and tokens of Procedure bundles than the expert writers and the native students. The non-native students used far more bundles for referring to the data collection, participants, and their responses, e.g., *agreed with the statement. participants were asked to, to collect the data, the sample of the,* etc. For the same purpose, only 1 bundle, *participants in this study*, used by the expert writers and 1 bundle, *participants were asked to*, was used by the native students. The non-native students also used many bundles for presenting the importance of the research, whereas only 1 bundle, *an important role in*, was used by the expert writers and none by the native students for this purpose. Similarly, many bundles were used by the non-native students for referring to the procedures of the study, e.g., *the use of the, by the use of, about the use* of etc., whereas only 2 bundle types were used in the expert and the native student corpora.

The potential reason for the significantly more frequent use of these bundles might be the larger size of the non-native student dissertation for which they need more bundles for explaining and describing the procedures, thus resulting into more varied and frequent use of Procedure bundles. The other possible reason might be that the non-native students tend to present very minute details of their research, therefore have to use far more procedure bundles.

Previous studies agree with the findings of the current study (Hyland, 2008a; Lu & Deng, 2019). Hyland (2008a) found that 25% of the total bundles in the Masters dissertations were focused on describing research objects or context, e.g., *the structure of the, an important role in, in order to maintain* etc. These findings are in line with the findings of the current study, which found that the non-native students used many Procedure bundles for describing the importance of research, e.g., *play an important role in, a vital role in the* etc., and about the use of procedure, e.g., *the use of the, through the use of* etc.

5.3.1.2 Description bundles

The non-native students used significantly more Description bundle types and tokens than the expert writers and the native students. The non-native students used more Verb-based bundles as Description bundles, e.g., *is the case with, that they are not, that it is the, he is of the, it is also a, in which they are, used in this study*. The native students only used two such bundles,

e.g., *this is not the, is that it is*, and the non-native students used bundles such as *in real life situation*, which is considered to be characteristic of informal speech rather than academic writing (Chen & Baker, 2010).

One different feature of these bundles is the use of Verb-based bundles in the non-native corpus, whereas there was no Verb-based bundle was used in the expert corpus. The non-native students used bundles like, *is the case with, that they are not, that it is the, he is of the, it is also a, in which they are, used in this study*. More variety of Description bundles in the non-native corpus suggests that the non-native students describe research more than the expert writers, and for this purpose they use variety of bundles.

The findings of the previous research are similar to these findings. (Hyland, 2008a; Pan et.al. 2016). Hyland (2008a) found that the non-native Masters students used far more bundles for description of research. Pan et.al. (2016) found that the non-native expert writers used Description bundles significantly more frequently than the native expert writers. So, the results of the current study and the previous studies show that the non-native students used far more Description bundles than were used by the native expert writers and the students.

5.3.1.3 Quantification bundles

The native and the non-native students have used significantly more types and tokens of quantification bundles than the expert writers. There are three bundles that were used only by the expert writers. These bundles are *in the number of, the amount of time, the total number of.* There are quantification bundles that were used only by the native students. These bundles are *a large number of, a small number of, is one of the, for each of the, each of the three, the vast majority of, the size of the, the majority of the*.

The difference in the use of Quantification bundles might be due to the difference of size between the research articles and the student dissertations. The native student corpus consists of Masters dissertation that are almost double the size of research articles in the expert corpus, and the size of the non-native student dissertations is three times more than the size of the research article. In the native and non-native dissertations, the results section is also much larger, and results and figures need explaining. Therefore, the native students might have to use more quantifying bundles than the expert writers.

The native and the non-native students have not only used more bundle types and tokens but also used bundles that are not characteristic of formal academic writing, e.g., *a large number of, the vast majority of.* These types of quantifying bundles are used for presenting a general

sense of quantity and lack precision. These types of bundles are found in novice student writing (Chen & Baker, 2010).

The non-native students used significantly more types and tokens of quantifying bundles than the expert writers. The non-native students used bundles containing the word 'majority', *majority of the respondents, the majority of the, that the majority of the, by majority of the;* 'most' *that most of the, one of the most, as one of the, a lot of time*. Bundles such as *a lot of time* are characteristic of speech rather than academic writing. None of these quantifying bundles were used by the expert writers. Therefore, the use of these bundles makes the student writing different from formal academic writing.

The findings of the current study are in contrast with Chen and Baker (2010) who found that the native expert and native student writers made a good use of quantifying bundles, whereas the non-native students used a small number of Quantification bundles. They observed that that both the native expert and the native student writers used a bundle type, extent/degree modifiers, e.g., *the extent to which, the degree to which* etc. They found that there were 4 types of these bundles in the expert corpus and 2 types in the native student corpus, however, no such quantifying bundle was found in the non-native student corpus. These findings are quite different from the findings of the current study because in the current study the native and the

non-native students used significantly more bundle types and tokens than the expert writers. However, it is important to note that the size and nature of the corpora in Chen and Baker (2010) is different from the corpus size and type of corpus in the current study. In the current study, the native and the non-native student corpus consists of Masters dissertations with long results chapter, therefore students have to use quantifying bundles. But in Chen and Baker (2010), the native and the non-native student corpus is based on term papers that are much smaller in size than dissertation.

In contrast to the findings of Chen and Baker (2010), Bychkovska and Lee (2017) found that in the use of quantifying bundles, the non-native students used significantly more bundle types and tokens. However, the non-native students repeatedly used bundles that are considered informal, e.g., *more and more people*, *nowadays more and more, a lot of people, a lot of time* etc. Bychkovska and Lee (2017) showed that these types of bundles were not used by the native English students.

These findings are in line with the findings of the current study. Two such informal bundle types of Quantifying bundles, e.g., *a lot of time, between two or more,* were found in the non-native student corpus. This indicates that the non-native students tend to use these types of informal quantifying bundles that are characteristic of informal speech. This also suggests that

as second language learners the non-native students have not yet fully mastered different levels of formality and different registers.

5.3.1.4 Location bundles

In the expert, native and non-native student corpora, similar location bundles were used, however there were bundle types that were not shared. The location bundles, *at the university of, in the middle of,* were used only in the expert corpus, whereas the bundles *all over the world* and *with the passage of time,* were used only in the non-native student corpus. Bundles such as *all over the world* are considered to be part of informal speech (Chen & Baker, 2010).

This shows the non-native students' tendency to use bundles which lack precision, as already observed in their use of Quantifying bundles. Previous research has also found the use of bundles of this type in non-native students (Bychkovska & Lee, 2017; Chen & Baker, 2010).
In the next section, I will discuss the role of Text-oriented bundles in the expert, native and the non-native student corpora.

5.3.2 Text-oriented bundles across the three corpora

The use of text-oriented bundles is one of the main characteristics of academic writing (Hyland, 2008a; 2008b). These bundles are used for organizing text, contextualizing new information, linking parts of discourse, and establishing textual coherence.

In the current study, text-oriented bundles were used differently across the three corpora. In the expert corpus, text-oriented bundles were far more frequent, representing 55% types and tokens of the total bundles. In the native and the non-native corpora, these bundles represented 40% of the total bundles in each corpus.

One of the reasons for this difference might be the difference in genre as the research articles are very precise and follow a strict word limit, therefore the expert writers have to use more bundles for the organization of the text.

The findings of Hyland (2008a) are in line with the findings of the current study. He found that text-oriented bundles were the most common bundles representing nearly two-third of the total bundles in research articles, whereas in non-native student writing these bundles represented 43% of the total bundles.

But there are other studies which found that the Text-oriented bundles were equally the most common bundles in native and the non-native student corpora. (Bychkovska & Lee, 2017; Lu & Deng, 2019). Pan et.al. (2016) also found that the Text-oriented bundles were the most common bundles in the native and the non-native expert corpus. Bychkovska and Lee (2017) found that native students used 70% of text-oriented types and tokens, and the non-native students used 59% types and 64% tokens of the total bundles. Lu and Deng (2019) found that both native and non-native students used Text-oriented bundles with similar proportions. However, it is important to note that these two studies are based on different genres from the genres used in the current study. Bychkovska and Lee (2017) is based on a corpus of student essays, and Lu and Deng (2019) is based on dissertation abstracts. Pan et. al. (2016) found a similar distribution of Text-oriented bundles in native and non-native English corpora, but they also compared the use of bundles in the same genre for the native and the non-native expert writers.

So, it appears that genre might play an important role in the characteristics of expert, native and non-native student writing. Research shows that Text-oriented bundles are equally common in the native and non-native student writing when they are used in the same genre, but they are less common in expert writers as the expert writing is based on a different genre,

the research article. The difference of genre might make a difference because the use of bundles is genre specific as has been shown in previous research (Cortes, 2004). The research articles are smaller in size than dissertations, moreover, the information in research articles has to be precise and concise, which needs more bundles for organization of the text. So, this might lead to more frequent use of Text-oriented bundles.

In the following section, the use of text-oriented bundles for organizing text will be discussed across the three corpora.

5.3.2.1 Framing signals

In the expert corpus, Framing signals are the most common text-oriented bundles representing 36% types and tokens of the total bundles, whereas in the native and the non-native student corpuora, Framing signals are less than twice as common representing 15% types and tokens of total bundles. Previous research has found that non-native students used significantly fewer framing bundles than native students (Byckhovska & Lee, 2017; Chen & Baker, 2010; Shin, 2019). Similarly, non-native expert writers also used framing bundles significantly less frequently than native expert writers (Pan et.al., 2016). In the current study, it has been found that framing signals are twice as common in the expert corpus as in the native and the non-native student corpus, however, the non-native students used significantly far more bundle

tokens than the expert writers and the native students. This difference might be linked with the students' general tendency to use text-oriented bundles less frequently and to use more research-oriented bundles. At the same time, the non-native students have used different framing bundles from the expert writers and the native students.

For example, the expert writers and the native students used a wide variety of framing signals that were not used by the non-native students. The bundles that were found only in the expert corpus are *in this case the, in this way the, on the part of, as a means of, as a way of, at the time of, at the level of, at the expense of, is based on the, when it comes to, with respect to the, in the light of* (see Table 5.5). These bundles are important for the organization of the text as they help contextualizing the information. The use of these bundles helps in making the text well organized and easy to follow.

Table 5.5 *frequency of the Framing signals (types & tokens) in the expert and the non-native student corpora*

Expert writers	**Tokens**	**Native students**	**Tokens**	**Non-native students**	**Tokens**
in terms of the	39	in line with the	28	as compared to the	35
in relation to the	25	in relation to the	28	regarding the use of	32
with respect to the	24	in the context of	24	in such a way	27

in response to the	17	the way in which	24	for the sake of	25
the way in which	17	in the case of	23	keeping in view the	25
as a way of	16	in terms of the	22	it refers to the	23
at the expense of	16	in the same way	20	is based on the	20
in line with the	16	in order to answer (the)	19	such a way that	19
with regard to the	16	on the basis of	19	is related to the	15
as a means of	15	with regard to the	18	same is the case	15
in this way the	15	the fact that the	16	the context in which	15
the relationship between the	15	with the exception of	15	in accordance with the	14
in terms of their	14	whether or not the	14	in this regard the	13
in the same way	13	in terms of their	13	when they try to	12
in the face of	10	in order to provide	12	are based on the	10
the context in which	10	as a starting point	11	in the presence of	10
the degree to which	10	the ways in which	11		
		the context in which	10		

The expert writers and the native students have used some of framing bundles that are idiomatic, e.g., *over the course of, in the course of, in relation to the, with respect to the, in response to the, at the expense of, in line with the, with regard to the, as a means of, in the face of.* The use of these bundles makes academic writing sound formal and native like. Moreover, all of these bundles in the expert corpus are Phrasal bundles. Interestingly, the non-native students did not use any of these bundles. The non-native students also used two idiomatic bundles. e.g., *from the perspective of, in accordance with the*. This difference between the expert writers, native students, and the non-native students indicates that the non-native students might find difficulty in using bundles that are idiomatic in nature. (Chen & Baker, 2010) The non-native students also used more clausal bundles for contextualizing new information, e.g., *it refers to the, is based on the, is related to the, same is the case, when they try to, are based on the*. Pan et.al. (2016) also found the similar results in the use of bundles in the non-native expert writing. Pan et.al. (2016) found that the majority of the text-oriented bundles in the non-native expert writing were Clausal bundles.

Previous research has also found that non-native students and non-native expert writers use different framing signals than were used by native students and native expert writers (Lu & Deng, 2019; Pan et.al., 2016). Lu and Deng (2019) found that native English students used

Preposition-based bundles beginning with *in* (*in the context of, in the face of*), whereas these bundles were rarely used by the non-native students. The non-native students used bundles beginning with 'with' (*with the help of*) and Verb-based bundles beginning with a past participle, e.g., *based on the analysis*. This is in line with the results of the current study which found that the non-native students used the bundle *with the help of*, but they did not use *in the face of*, that was used by the expert writers. One of the reasons of this difference might be that the non-native students find it difficult to learn formulaic bundles that are opaque, e.g., *in the face of, with respect to the, in terms of the, in line with the* etc. Therefore, they use bundles that are not opaque, e.g., *as compared to the, in the presence of* etc. (Paquot & Granger, 2012)

A different use of framing signals seems to exist between the native and the non-native writers even at the expert level. Pan et.al. (2016) found that Framing signals was the only category in which the non-native expert writers used bundle tokens less frequently than the native expert writers. Moreover, in their use of framing signals, the native expert writers used a wide variety of bundles for specifying the condition, e.g., *in the context of*, for focusing readers on a given case, e.g., *in the case of*, to emphasize an aspect of an argument, e.g., in terms of the. In contrast, a limited number of framing signals were used by the non-native expert writers for the same purpose, e.g., *in the case of, with respect to the*.

So, the findings of previous studies and the current study show that non-native students and non-native expert writers do not use bundle types that are used by expert writers and native students. The possible reasons for this might be non-native students' difficulty in coping with opaque formulaic bundles. The other reason might be that the non-native students are not well versed with the use of framing signals in academic writing. (Adel & Erman, 2012; Bychkovska & Lee, 2017; Chen & Baker, 2010; Hyland, 2008a; Pan et.al., 2016;)

5.3.2.2 Transition signals

The non-native students have used significantly more bundle tokens of Transition signals than the expert writers and the native students. There is no significant difference in the use of bundle types of transition signals across the three corpora, though the non-native students used bundle types that were not used by the expert writers and the native students. These bundle types are *as well as in, in other words the, in spite of the, on the one hand, and at the same, and its use in, on the other side, and they do not*.

The possible reason for significantly more frequent use of Transition signals by the native and the non-native students might be that the native and non-native students use more Transition signals to contextualize new information. In the process, the non-native students might also make mistakes. For example, Pan et.al. (2016) found that non-native expert writers use

Transition signals, *as well as the*, as a conjunction. Although no such evidence of mistakes in the use of bundles was found in the current study.

The results of previous research are in line with the findings of this study (Lu & Deng, 2019; Pan et.al. 2016). Both, Lu and Deng, (2019) and Pan et.al. (2016) found that non-native students used Transition signals significantly more frequently than native students.

5.3.2.3 Structuring signals

The non-native students used significantly more types and tokens of Structuring signals than the expert writers and the native students. The native students used only one bundle type of Structuring signals, e.g., *in the present study*. The findings of Pan et.al. (2016) are in line with the finding of the current study. They also found that non-native expert writers used Structuring signals significantly more frequently than native expert writers. However, there were some bundles in the native corpus referring to the section, e.g., *in the next section, in the previous section, in this section we*, that were not found in the non-native corpus.

The non-native students used Structuring bundles that show repetition. These bundles are *of the present study, for the present study, in this study the, the present study is, the present study was, for the present research, this research was to, in the present research, in this section the,*

this chapter deals with. There is a repetition of the noun ‘present’ and ‘study’ in these bundles. One of the possible reasons for this repetition might be due to a lack of lexical sources in non-native students, characteristic of non-native student writing (Adel & Erman, 2012; Chen & Baker, 2010). The other possible reason might be differences in genre. The non-native student dissertations are much larger in size than the research articles and the native student dissertations. The non-native students might have to refer the study multiple times, thus resulting in more Structuring signals. On the other hand, Lu and Deng (2019) found that the native English student used twice as many bundles as were used by the non-native students. However, Lu and Deng (2019) is based on the Doctoral abstracts, which is a different genre from the dissertations used in the current study.

5.3.2.4 Resultative signals

The non-native students have used significantly more types and tokens of Resultative signals than the expert writers and the native students. One of the reasons of this might be the larger size of the corpora in the non-native student corpus (see Section 3.2.1). As the non-native student corpus is based on Masters dissertations that contain a much larger results section than the research article and also the native student masters dissertation, therefore, the non-native

students might have to refer to the results many times, thus resulting into use of more resultative signals.

Hyland (2008a) found different results from the current study. He found that expert writers used resultative signals more frequently than the non-native students. One of the reasons for more use of resultative bundles in the expert corpus might be the larger size of the expert corpus in Hyland (2008a) which consisted of 730,000 words, whereas in the current study the size of the expert corpus is 505945 words. The other possible reason might be the design of the corpus in Hyland (2008a), in which the data has been taken from four different disciplines. In contrast, the data of the current study has been taken from the field of Applied Linguistics only. So, disciplinary differences might have played a role in differences in the use of resultative bundles.

On the other hand, the findings of Pan et.al. (2016) are similar to the findings of the current study. They found that non-native expert writers used the resultative signals significantly more frequently than native expert writers. So, research shows mixed findings. In the next section, I will discuss the role of participant-oriented bundles in the expert, native and non-native student corpora.

5.3.3 Participant-oriented bundles across the three corpora

Participant-oriented bundles are used for presenting writers' evaluation and engaging the readers. These are the least common bundles, across the three corpora representing 8% types and tokens of total bundles in the expert corpus, 13% types and tokens of total bundles in the native student corpus, and 5% types and tokens of total bundles in the non-native corpus. However, both the native and the non-native students used these bundles significantly more frequently than the expert writers.

5.3.3.1 Stance bundles

The non-native students have used Participant-oriented Stance bundles differently from the expert writers and the native students. The expert writers and the native students used similar Stance bundles, e.g., *it is possible to, it is possible that, it is likely to* etc. These bundles showing the possibility and likelihood were not used by the non-native students. As has been mentioned before that the reason for this might be that the non-native students use limited variety of lexical bundles for giving their evaluation as they do not see them as expert writers. The other reason might be that the non-native students might be asked by their teachers not to present their evaluation. The previous research has also shown that the non-native students tend to use limited number of Stance bundles whereas the native students used a wide variety of Stance

bundles. (Adel & Erman, 2012; Chen & Baker, 2010; Hyland, 2008a) In contrast to these studies, there are studies which found that the non-native students used significantly more Stance bundles than the native students. (Bychkovska & Lee, 2017; Shin, 2019). These different findings might be due to different research designs in these two studies. For example, these two studies are based on the student argumentative essays which require writers' evaluation, therefore the frequent Stance bundles would have been expected by the native and the non-native students.

5.3.3.2 Engagement features

The non-native students used the Participant-oriented engagement features significantly less frequently than the native students. These bundles are used for engaging with the reader in the text. The engagement features were only 1% types and tokens of the total bundles in the non-native corpus. The possible reason for using a very small number of engagement features might be non-native students' general tendency to maintain authorial anonymity. As they do not see themselves as experts and do not feel the need to engage with the reader by using bundles like, *it is interesting to note, it is worth noting* etc. This tendency might also be linked with classroom instruction as students might be asked to avoid providing their evaluation to maintain subjectivity in their writing.

5.4 Conclusion

The following are the main features discussed above:

- The findings of the current study and previous studies show that the genre affects the use of lexical bundles in academic writing. In different genres, the frequency, and the bundle functions might change dramatically. Therefore, it is important to compare the use of lexical bundles in similar genres rather than comparing different genres.

- The non-native students seem to be quite different from expert writers and native students in their use of lexical bundles. On the other hand, in some aspects of bundle use, both the native and the non-native students used bundles similarly. Following are the main feature of the expert writing:

- The expert writers use more bundles for organizing text and established coherence in the text, whereas the native and the non-native students seem to use more bundles for describing research. The tendency to explain the procedures and describe research was found to be the most prominent feature of bundle use in the non-native corpus.

- The expert and the native students showed good control of techniques for presenting interpretations of the results, and for this they successfully adopted a tone of neutrality.

They used techniques to detach themselves while giving interpretation. On the other hand, the non-native students either tried to maintain authorial anonymity or they used bundles that made them appear subjective.

- The expert and the native students used varied bundles which made their bundle use less repetitive, whereas the non-native students tend to be repetitive in their use of lexical bundles. But the role of much larger size of the non-native student dissertations cannot be ignored in this regard.

- The expert writers and the native students used quite similar bundle types, especially for the purpose of contextualizing the text. They made a good use of idiomatic formulaic bundles for this purpose. In contrast, the non-native students seem to use different and non-idiomatic bundles for this purpose. This might suggest that the non-native students have not yet mastered the use of idiomatic bundles in their writing.

- The expert writers used bundles that made their writing more precise and formal, whereas both the native and the non-native students tended to use bundles that made their writing less precise and more informal.

Chapter 6

Conclusion

This chapter summarises the study, highlighting its most important findings. Limitations of this study are then presented before outlining some recommendations for future research.

6.1 Summary

Chapter 1 presented the background underpinning this study examining the use of lexical bundles in the academic writing of expert writers, native and non-native students. Chapter 2 introduced the concept of formulaic language and different approaches to its investigation were discussed, before focusing on lexical bundles, i.e., frequency-based corpus-driven formulaic word sequences. Studies on the use of lexical bundles in academic writing in expert writers, native students and non-native students were then critically evaluated. Chapter 3 presented the methodology used in this study and the different criterion such as frequency, dispersion, and size of lexical bundles were discussed. This chapter also discussed various tools of the corpus software, AntConc, which has been used for corpus analysis in this study. Chapter 4 presented the results for each corpus and compared them. Chapter 5 discussed the results of this study in

the light of previous literature, and revisited the answers to the research questions set in this study, which can be summarised as follows:

R.Q.1 (a) What are the most frequent bundle structural categories in the expert writers' corpus, native students' corpus, and non-native students' corpus respectively?

(b) How does frequency of structural categories compare across the three corpora?

Answer: The distribution of the bundle structural categories presented a number of differences across the three corpora. Preposition-based bundles were found to be the most common bundles in the expert corpus, Noun-based bundles were the most common bundles in the native corpus, whereas Verb-based bundles were the most common bundles in the non-native corpus. Moreover, Phrasal bundles that are characteristic of academic writing were found to be the most common bundles across the three corpora. However, these bundles were far more common in the expert and the native student corpora than in the non-native student corpus. These differences had an impact on the quality of academic writing used in the three corpora. In the expert writing, the use of bundles was more text-oriented, i.e., more bundles were used for organizing the text, whereas the native and the non-native student writing was more research-oriented, i.e. more bundles were used for describing research. This was a marked

difference between the expert writing and the novice writing. This difference was more prominent in the non-native student writing who used far more verb-based bundles for describing research and explaining procedures.

Verb-based bundles were far more common in the non-native student corpus than in the expert and the native student corpora. Moreover, the expert writers and the native students used the majority of verb-based bundles for hedging, whereas the non-native students used these bundles for describing research as well as hedging. The expert and the native student writers used similar bundle types for hedging and showed better control of hedging devices. The non-native students used different bundles for hedging.

R.Q.2 (a) What are the most frequent bundle functional categories in expert writers' corpus, native students' corpus, and non-native students' corpus respectively?

(b) How does frequency of functional categories compare across the three corpora?

Answer: The distribution of functional categories was also different across the three corpora. The native and non-native students used significantly more Research-oriented bundles than the expert writers, especially the non-native students who used far more bundles for describing research procedures than the experts and the native students. The non-native students also used far more bundles for referring to the data collection procedure, the importance of the study, and

the general procedures of research. The native and the non-native students rarely used or did not use bundles for these purposes. In their use of Research-oriented Quantification bundles, both the native and the non-native students used significantly more bundles than the expert writers. One of the features of the Quantifying bundles used by both groups of students was the lack of precision and overgeneralization.

Text-oriented bundles were far more common in the expert corpus than the native and the non-native student corpora, especially the bundles that are used for contextualizing new information, such as Framing bundles. The expert writers and the native students used more varied framing bundles and they used Framing bundles that were idiomatic. The non-native students also used bundles differently from both the expert writers and the native students. They did not use idiomatic Framing bundles. In their use of Structuring bundles that are used for referring to the study, the non-native students used significantly more types and tokens than the expert writers and the native students. On the other hand, in the use of Resultative signals that are used for referring to the results, the native and the non-native students used significantly more bundle types. These differences were expected in the native and the non-native student corpora as the student dissertations are twice as big as the research articles, therefore the student writers have a much longer results sections which results in more Resultative bundles. In the

use of Transition signals, no significant differences were found across the three corpora, and the bundle types used across the three corpora were very similar.

Participant-oriented bundles were the least common bundles across the three corpora. The expert writers and the native students used similar bundle types, however, the non-native students used different bundle types. They also used significantly fewer tokens for engaging the readers in the text.

6.2 Limitations of the study

There are some limitations to the current study. The major limitation is that the size of the native student corpus is smaller than both the expert and the non-native student corpora. This is one of the most challenging aspects of corpus studies (Biber & Barberi, 2007). Although efforts have been made in this study to ensure comparability of different size corpora (see section 3.2.4), similar size corpora would make the findings more reliable. Another limitation is the comparability of the genres, as two different genres have been compared in this study: research articles and student dissertations. As the use of lexical bundles is genre specific, and some genres are quite different in their use of bundles, comparing two different genres might affect the results (Cortes, 2004; Hyland, 2008a). Research articles are the closest to Masters dissertations, as both involve the reporting and writing up of research, and published articles

are the model that students are supposed to emulate. Previous research has compared student writing to textbooks, but these are even more different from student writing than research articles. (Biber et al., 2004)

So, the research article is the only genre that is closely comparable to the student dissertations. In the following section, I will discuss some of the pedagogical implications of the current research.

6.3 Pedagogical implications

The current research has implications for English language teachers who teach academic writing to the leaners of English language. The current research has shown that the learners struggle with the use of certain lexical bundles (e.g., idiomatic bundles). Similarly, the learners used the different bundles for hedging in their writing. These results suggest that the learners need to be identify these types of bundles in expert writing. Therefore, this section will outline some suggestions that can be useful for teaching lexical bundles. Moreover, I will also describe some useful techniques like the use of personal corpora, which can be useful in learning lexical bundles. The pedagogical implication of the current study are as follows:

- The study shows that there are some features of lexical bundles in which both native and non-native students deviate from the norms of academic writing. For example, they tend to

use more Research-oriented bundles than Text-oriented bundles which makes their writing more descriptive. Novice students need to be taught the use of these bundles, especially the bundles that are used for contextualizing new information. The native students used significantly fewer bundles for referring to the study and the sections of the study. These aspects of bundle use are very important to make the text well organized and reader friendly. In addition to this, in the use of Quantifying bundles, both the native and the non-native students use imprecise bundles such as *a large number of*. This imprecise use of bundles was even more common in the non-native student writing, and the bundles they used were also more informal, e.g., *a lot of the*. The non-native students also used a location bundle, *all over the world*, that is also considered informal and shows a tendency to overgeneralize (Chen & Baker, 2010). The use of these types of bundles provides imprecise and overgeneralized information to the readers. Therefore, it might be useful to teach both native and the non-native students the importance of formality and preciseness in academic writing.

- The non-native students might need to be trained in the use of some bundle functions. They may need to be encouraged to use more Phrasal bundles which make writing more information focused and help organise the text. They might also need to be shown how to

use bundles for hedging, as they used a variety of bundles for hedging, but not for showing e.g., the likelihood or possibility of a proposition. This might be because non-native students generally use limited variety of lexical bundles for giving their evaluation. The use of these bundles would make their academic writing more neutral.

- Idiomatic lexical bundles present another challenge for non-native students. In their use of framing bundles, both expert writers and the native students used a variety of framing bundles that are idiomatic. Non-native students, on the other hand, used bundles that are non-idiomatic. They might therefore need to be taught these idiomatic bundles, in order to make their academic writing more native-like.
- Non-native students also need to improve their use of bundles for engaging the readers, as they use significantly fewer bundle tokens for that purpose. This is important for making student writing reader friendly, making the reader more involved and interested. It can also bring more clarity to a text.
- The students can build their own corpus (personal corpus) e.g., a corpus of research articles from their field (Charles, 2012; 2014). The use of personal corpora can be of great help in making student learn and improve the features of academic writing, especially at the university level. The students can learn lexico-grammatical features and discourse

functions by using their personal corpus. For example, the students can learn the collocates of different words by using the cluster tools; they can use the corpus tools to search for the frequency and sentence position of different words, e.g., linking adverbials, *however* (Charles, 2014). Research has proved that the students who used their personal corpus have improved their academic writing (Charles, 2014).

- The teacher can ask the students to develop a corpus of research articles in the field of linguistics and applied linguistics. This corpus can be used for teaching different discourse functions, e.g., hedging by using tools like KIWI. For example, the teacher might ask the students to search for the words like, *possible, believe, seems, suggest, should be, would be, might be* in the corpus. By finding these words the students would be able to observe and examine the use of these hedging devices. The learners can then search for the same hedging devices in their personal corpora.
- The teachers can use corpora in the classroom to teach the bundles that are most frequent and idiomatic. Different excerpts from research articles might be used for teaching these bundles. The teacher can encourage the students to assess the meanings of those bundles in the context of their usage.

- The teacher can also ask the students to use a small corpus of research articles in groups and to find out various discourse features like hedging, reader-orientation, organization of the text. The students can then discuss differences and similarities of these features that they have found out (Anthony, 2020).
- The findings of the current study might be helpful for EAP teachers as it highlights areas of bundle use where the non-native students have not yet learnt important techniques of academic writing, such as using more bundles for organizing the text, using formal bundles and using more bundles for referring to the study itself. In addition to this, there are other aspects of bundle use that the non-native students might need to improve, such as a greater use of Phrasal bundles, and to use of bundles for hedging, for engaging the reader, as well as more idiomatic use of bundles.

6.4 Recommendations for future research

Some suggestions can be made for future research on the basis of the findings of this study. As has been mentioned in the limitations of this study, future researchers might endeavour to find ways to ensure more comparable size of corpora. In this study, the expert and the non-native student corpora were of an almost similar size, but the native student corpus was smaller. A larger corpus from native students would help to ensure better comparability of the three

corpora. Moreover, the similar size of the individual texts in each corpus would be more helpful for making the corpora comparable. Additionally, the use of two different genres, research articles and student dissertations, might have had an impact on the findings of this study, and future researchers might explore more similar genres for the expert and the student academic writing. More comparable genres might provide more robust findings.

An important area of research on lexical bundles can be the role of personal corpora on the learning of lexical bundles. In this area, research has shown that by using personal corpora postgraduate students improved their academic writing (Charles, 2014). However, there is a need to do more research in this area to see how this type of DDL techniques can help in teaching lexical bundles to the students at various levels. The influence of second language learners' culture and L1 background on the use of lexical bundles is another interesting and important area of research. Some research has been conducted in this area as well, but more detailed research in this area will further help to understand the differences between native and non-native students in their use of lexical bundles.

References

Ädel, A., & Erman, B. (2012). Recurrent word combinations in academic writing by native and non-native speakers of English: A lexical bundles approach. *English for specific purposes, 31(2),* 81-92.

Altenberg, B. (1998). On the phraseology of spoken English: The evidence of recurrent word combinations. In A. Cowie (Ed.), *Phraseology: Theory, analysis and applications* (pp. 101–122). Oxford: Oxford University Press.

Andrews, R. (2007). Argumentation, critical thinking and the postgraduate dissertation. *Educational Review*, *59*(1), 1-18.

Anthony, L. (2014). AntConc (Version 3.4. 3) [Computer Software]. Tokyo, Japan: Waseda University.

Anthony, L. (2016). Introducing corpora and corpus tools into the technical writing classroom through Data-Driven Learning (DDL). *Discipline-Specific Writing* (pp. 176-194). Routledge.

Anthony, L. (2019). AntConc (Version 3.5. 8). *Tokyo: Waseda University*. Available from: http://www.laurenceanthony.net

Anthony, L. (2020). AntConc (Version 3.5. 9) [Software]. Waseda University. Available from: http://www.laurenceanthony.net

Biber, D. (2006). Stance in spoken and written university registers. *Journal of English for Academic Purposes, 5(2),* 97-116.

Biber, D. (2009). A corpus-driven approach to formulaic language in English: Multi-word patterns in speech and writing. *International journal of corpus linguistics*, *14*(3), 275-311.

Biber, D. (2010). Corpus-based and corpus-driven analyses of language variation and use. In B. Heine & H. Narrog (Eds.), *The Oxford handbook of linguistic analysis* (pp. 159–191). Oxford: Oxford University Press

Biber, D., & Barbieri, F. (2007). Lexical bundles in university spoken and written registers. *English for specific purposes, 26(3),* 263-286.

Biber, D., Conrad, S., & Cortes, V. (2004). If you look at…: Lexical bundles in university teaching and textbooks. *Applied linguistics, 25(3),* 371-405.

Biber, D., Conrad, S., & Reppen, R. (1998). *Corpus linguistics: Investigating language structure and use*. Cambridge: Cambridge University Press.

Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (1999). Longman grammar of spoken and written English. London: Longman

Boulton, A. (2012). Corpus consultation for ESP: A review of empirical research. *Corpus-informed research and learning in ESP: Issues and applications*, pp. 261-291. Amsterdam/Philadelphia: John Benjamins.

Bychkovska, T., & Lee, J. J. (2017). At the same time: Lexical bundles in L1 and L2 university student argumentative writing. *Journal of English for Academic Purposes, 30*, 38-52.

Charles, M. (2012). 'Proper vocabulary and juicy collocations': EAP students evaluate do-it-yourself corpus-building. *English for Specific Purposes, 31*(2), 93-102.

Charles, M. (2014). Getting the corpus habit: EAP students' long-term use of personal corpora. *English for Specific Purposes, 35*, 30-40.

Chen, P., & Zhao, C. (2022). The treatment of academic lexical bundles in online English monolingual learners' dictionaries. *International Journal of Lexicography*. https://doi.org/10.1093/ijl/ecab032

Chen, Y. H., & Baker, P. (2010). Lexical bundles in L1 and L2 academic writing. *Language learning & technology, 14(2),* 30-49.

Chen, Y. H., & Baker, P. (2016). Investigating criterial discourse features across second language development: Lexical bundles in rated learner essays, CEFR B1, B2 and C1. *Applied Linguistics, 37(6),* 849-880.

Conrad, S. M., & Biber, D. (2005). The frequency and use of lexical bundles in conversation and academic prose. *Lexicographica,20,* 56-71.

Cortes, V. (2004). Lexical bundles in published and student disciplinary writing: Examples from history and biology. *English for specific purposes, 23(4),* 397-423.

Cortes, V. (2015). Situating lexical bundles in the formulaic language spectrum: Origins and functional analysis developments. In *Corpus-based research in applied linguistics* (pp. 197-216). John Benjamins.

Cowie, A. P. (Ed.). (1998). *Phraseology: Theory, analysis, and applications*. Oxford. OUP

Fareed, M., Ashraf, A., & Bilal, M. (2016). ESL learners' writing skills: Problems, factors and suggestions. *Journal of Education and Social Sciences, 4*(2), 81-92.

Fazal, M. A. M. H. Z., & Moavia, H. (2019). Formulaic Language in Social Sciences: A Functional Analysis of Lexical Bundles in Native and Non-Native Academic Discourse. *Pakistan Social Sciences Review, 3(1),* 234-249

Fillmore, C. J., Kay, P., & O'connor, M. C. (1988). Regularity and idiomaticity in grammatical constructions: The case of let alone. *Language, 64(3),* 501-538.

Gablasova, D., Brezina, V., & McEnery, T. (2017). Collocations in corpus-based language learning research: Identifying, comparing, and interpreting the evidence. *Language learning, 67*(S1), 155-179.

Gilquin, G., & Paquot, M. (2008). Too chatty: Learner academic writing and register variation. *English Text Construction, 1*(1), 41-61.

Gledhill, C. J. (2000). *Collocations in science writing* (Vol. 22). Tubingen: Gunter Narr Verlag.

Granger, S. (2002). A bird's-eye view of learner corpus research. *Computer learner corpora, second language acquisition and foreign language teaching, 6,* 3-33.

Granger, S., & Lefer, M. A. (2016). Towards more and better phrasal entries in bilingual dictionaries. In *Proceedings of the 15th EURALEX International Congress* (pp. 682-692). Oslo: University of Oslo.

Haider, G. (2012). Teaching of writing in Pakistan: A review of major pedagogical trends and issues in teaching of writing. *Journal of Educational and Social Research*, *2*(3), 215-215.

Halliday, M.A.K. (1994) *Functions of language*. 2nd edn. London: Arnold.

Hinkel, E. (2017). Teaching Idiomatic Expressions and Phrases: Insights and Techniques. *Iranian Journal of Language Teaching Research*, *5*(3), 45-59.

Huang, K. (2015). More does not mean better: Frequency and accuracy analysis of lexical bundles in Chinese EFL learners' essay writing. *System*, *53*, 13-23.

Huang, L. S. (2017). Has Corpus-Based Instruction Reached a Tipping Point? Practical Applications and Pointers for Teachers. *TESOL Journal*, *8*(2), 295-313.

Hunston, S. (2002). *Corpora in applied linguistics*. Cambridge, UK: Cambridge University Press.

Hussain, Z., Hanif, M., Asif, S. I., &Rehman, A. U. (2013). An error analysis of L2 writing at higher secondary level in Multan. *Interdisciplinary Journal of Contemporary Research in Business, 4 (11)*, 828- 844

Hyland, K. (1996). Writing without conviction? Hedging in science research articles. *Applied linguistics*, *17*(4), 433-454.

Hyland, K. (1998). *Hedging in scientific research articles* (Vol. 54). Amsterdam: John Benjamins Publishing Company.

Hyland, K. (2008a). As can be seen: Lexical bundles and disciplinary variation. *English for specific purposes, 27(1),* 4-21.

Hyland, K. (2008b). Academic clusters: Text patterning in published and postgraduate writing. *International Journal of Applied Linguistics, 18(1)*, 41-62.

Hyland, K. (2009). *Academic Discourse: English In A Global Context* (Vol. 7). Bloomsbury Publishing.

Hyland, K. (2012). Bundles in academic discourse. *Annual review of applied linguistics, 32,* 150-169.

Hyland, K., & Jiang, F. K. (2017). Is academic writing becoming more informal? *English for Specific Purposes, 45*, 40-51.

Khan, H. I. (2011). Testing creative writing in Pakistan: Tensions and potential in classroom practice. *International Journal of Humanities and Social Sciences, 1(15)*, 111-119.

Li, J., & Schmitt, N. (2009). The acquisition of lexical phrases in academic writing: A longitudinal case study. *Journal of second Language Writing, 18(2)*, 85-102.

Lu, X., & Deng, J. (2019). With the rapid development: A contrastive analysis of lexical bundles in dissertation abstracts by Chinese and L1 English doctoral students. *Journal of English for Academic Purposes, 39*, 21-36.

Mahlberg, M. (2005). *English General Nouns: A Corpus Theoretical Approach* (Vol. 20). John Benjamins Publishing.

Marco, M. J. L. (2000). Collocational frameworks in medical research papers: A genre-based study. *English for specific purposes, 19*(1), 63-86.

Montgomery, S. (1996) *The Scientific Voice*. New York: Guilford Press

Moon, R. (1998). *Fixed expressions and idioms in English: A corpus-based approach*. New York: Oxford University Press.

Moon, R. (2015). Multi-word items. In J. R. Taylor (Ed.), *The Oxford handbook of the word*, 1st ed. (pp. 120–140). Oxford: Oxford University Press

Myles, F. (2005). Interlanguage corpora and second language acquisition research. *Second Language Research*, *21*(4), 373-391.

Myles, F., & Cordier, C. (2017). Formulaic sequence (FS) cannot be an umbrella term in SLA: Focusing on psycholinguistic FSs and their identification. *Studies in Second Language Acquisition*, *39*(1), 3-28.

Nesi, H., & Basturkmen, H. (2006). Lexical bundles and discourse signaling in academic lectures. *International Journal of Corpus Linguistics*, *11*(3), 283-304.

Pan, F., Reppen, R., & Biber, D. (2016). Comparing patterns of L1 versus L2 English academic professionals: Lexical bundles in Telecommunications research journals. *Journal of English for Academic Purposes*, *100*(21), 60-71.

Pang, P. (2009). A study on the use of four-word lexical bundles in argumentative essays by Chinese English-majors: A comparative study based on WECCL and LOCNESS. *Teaching English in China, 32*(3), 25-4.

Paquot, M., & Granger, S. (2012). Formulaic language in learner corpora. *Annual Review of Applied Linguistics, 32*, 130-149.

Paquot, M. (2008). Exemplification in learner writing: A cross-linguistic perspective. In Meunier, F. & Granger, S. (Eds.), *Phraseology in foreign language learning and teaching*, 101–119. Amsterdam, the Netherlands: John Benjamins.

Partington, A. (1998). *Patterns and Meanings – Using Corpora for English Language Research and Teaching*. Amsterdam/Philadelphia: John Benjamins.

Partington, A., & Morley, J. (2004). From frequency to ideology: investigating word and cluster/bundle frequency in political debate. In B. Lewandowska-Tomaszczyk (Ed.), Practical applications in language and computers – PALC 2003 (pp. 179–192). Frankfurt a. Main: Peter Lang.

Peters, E., & Pauwels, P. (2015). Learning academic formulaic sequences. *Journal of English for academic purposes, 20*, 28-39.

Rayson, P. (2016). Log-likelihood calculator. Available at: http://ucrel.lancs.ac.uk/llwizard.html

Renouf, A. & J. Sinclair (1991). Collocational frameworks in English. In Aijmer, K. & B. Altenberg (eds.). *English Corpus Linguistics*: Studies in Honour of Jan Svartvik, 128–143. London & New York: Longman

Römer, U. (2011). Corpus research applications in second language teaching. *Annual review of applied linguistics, 31,* 205-225.

Salazar, D. (2011). *Lexical bundles in scientific English: A corpus-based study of native and nonnative writing* [Unpublished doctoral dissertation]. University of Barcelona.

Shin, Y. K. (2019). Do native writers always have a head start over nonnative writers? The use of lexical bundles in college students' essays. *Journal of English for Academic Purposes, 40,* 1-14.

Simpson-Vlach, R., & Ellis, N. C. (2010). An academic formulas list: New methods in phraseology research. *Applied linguistics, 31(4),* 487-512.

Sinclair J. (2004) *Trust the Text – Language, corpus and discourse*. London: Routledge.

Sinclair, J., & Sinclair, L. (1991). *Corpus, concordance, collocation*. Oxford University Press, USA.

Sinclair, John (1987). *Looking up*. London: Collins.

Siyanova-Chanturia, A. (2015). On the 'holistic'nature of formulaic language. *Corpus Linguistics and Linguistic Theory, 11(2),* 285-301.

Staples, S., Egbert, J., Biber, D., & McClair, A. (2013). Formulaic sequences and EAP writing development: Lexical bundles in the TOEFL iBT writing section. *Journal of English for academic purposes*, *12*(3), 214-225.

Stubbs, M. (2007). An example of frequent English phraseology: distribution, structures and functions. In Facchinetti, R. (ed.). *Corpus Linguistics 25 Years on*, 89–105. Amsterdam & New York: Rodopi.

Tognini-Bonelli, E. (2001). *Corpus Linguistics at Work*. Amsterdam: John Benjamins.

Wray, A. (2013). Formulaic language. *Language Teaching*, *46*(3), 316-334.

Yousaf, M., & Shehzad, W. (2018). Fixedness of expressions in doctoral research dissertations: A corpus-based analysis. *Kashmir Journal of Language Research*, *21*(2), 27-45.

Appendix

The Research Articles in the expert writers' corpus:

Abobaker, R. (2017). Improving ELL s' Listening Competence Through Written Scaffolds. *TESOL Journal*, *8*(4), 831-849.

Alonso-Almeida, F., & Luisa Carrio-Pastor, M. (2019). Constructing legitimation in Scottish newspapers: The case of the independence referendum. *Discourse Studies*, *21*(6), 621-635.

Andrus, J. (2019). Identity, self and other: The emergence of police and victim/survivor identities in domestic violence narratives. *Discourse Studies*, *21*(6), 636-659.

Berger, C. M., Crossley, S. A., & Kyle, K. (2019). Using native-speaker psycholinguistic norms to predict lexical proficiency and development in second-language production. *Applied Linguistics*, *40*(1), 22-42.

Butler, Y. G. (2017). Motivational elements of digital instructional games: A study of young L2 learners' game designs. *Language Teaching Research*, *21*(6), 735-750.

Caplan, N. A., & Farling, M. (2017). A dozen heads are better than one: Collaborative writing in genre-based pedagogy. *TESOL Journal*, *8*(3), 564-581.

Cardimona, K. (2018). Differentiating mathematics instruction for secondary-level English Language Learners in the mainstream classroom. *Tesol Journal*, *9*(1), 17-57.

Crossley, S. A., & McNamara, D. S. (2014). Does writing development equal writing quality? A computational investigation of syntactic complexity in L2 learners. *Journal of Second Language Writing*, *26*, 66-79.

Dang, T. C. T., & Seals, C. (2018). An evaluation of primary English textbooks in Vietnam: A sociolinguistic perspective. *TESOL Journal*, *9*(1), 93-113.

De Oliveira, L. C., & Lan, S. W. (2014). Writing science in an upper elementary classroom: A genre-based approach to teaching English language learners. *Journal of Second Language Writing*, *25*, 23-39.

Dinkin, A. J. (2018). It's no problem to be polite: Apparent-time change in responses to thanks. *Journal of Sociolinguistics*, *22*(2), 190-215.

Dionigi, A., & Canestrari, C. (2018). The role of laughter in cognitive-behavioral therapy: case studies. *Discourse Studies*, *20*(3), 323-339.

Dizon, G. (2017). Using intelligent personal assistants for second language learning: A case study of Alexa. *Tesol Journal*, *8*(4), 811-830.

Drew, P. (2018). Epistemics–the rebuttal special issue: an introduction. *Discourse Studies*, *20*(1), 3-13.

Eckert, P., & Labov, W. (2017). Phonetics, phonology and social meaning. *Journal of sociolinguistics*, *21*(4), 467-496.

El Naggar, S. (2018). 'But I did not do anything!'–analysing the YouTube videos of the American Muslim televangelist Baba Ali: delineating the complexity of a novel genre. *Critical Discourse Studies*, *15*(3), 303-319.

Fairclough, N., & Fairclough, I. (2018). A procedural approach to ethical critique in CDA. *Critical Discourse Studies*, *15*(2), 169-185.

Foley, J. (1991). A phycholinguistic framework for task-based approaches to language teaching. *Applied linguistics*, *12*(1), 62-75.

Gebril, A. (2018). Test preparation in the accountability era: Toward a learning-oriented approach. *TESOL Journal*, *9*(1), 4-16.

Gevers, J. (2018). Translingualism revisited: Language difference and hybridity in L2 writing. *Journal of Second Language Writing*, *40*, 73-83.

Graham, P. (2018). Ethics in critical discourse analysis. *Critical Discourse Studies*, *15*(2), 186-203.

Greco, S., Schär, R., Pollaroli, C., & Mercuri, C. (2018). Adding a temporal dimension to the analysis of argumentative discourse: Justified reframing as a means of turning a single-issue discussion into a complex argumentative discussion. *Discourse Studies*, *20*(6), 726-742.

Grossi, V., & Gurney, L. (2020). 'Is it ever enough?'Exploring academic language and learning advisory identities through small stories. *Discourse Studies*, *22*(1), 32-47.

Han, J., & Hiver, P. (2018). Genre-based L2 writing instruction and writing-specific psychological factors: The dynamics of change. *Journal of Second Language Writing*, *40*, 44-59.

Hansson, S. (2018). Analysing opposition–government blame games: Argument models and strategic maneuvering. *Critical Discourse Studies*, *15*(3), 228-246.

Hellermann, J. (2006). Classroom interactive practices for developing L2 literacy: A microethnographic study of two beginning adult learners of English. *Applied Linguistics*, *27*(3), 377-404.

Hirano, E. (2014). Refugees in first-year college: Academic writing challenges and resources. *Journal of Second Language Writing*, *23*, 37-52.

Izumi, S. (2003). Comprehension and production processes in second language learning: In search of the psycholinguistic rationale of the output hypothesis. *Applied Linguistics*, *24*(2), 168-196.

Jackson, D. O., & Cho, M. (2018). Language teacher noticing: A socio-cognitive window on classroom realities. *Language Teaching Research*, *22*(1), 29-46.

Keating, G. D. (2008). Task effectiveness and word learning in a second language: The involvement load hypothesis on trial. *Language teaching research*, *12*(3), 365-386.

Keck, C. (2014). Copying, paraphrasing, and academic writing development: A re-examination of L1 and L2 summarization practices. *Journal of Second Language Writing*, *25*, 4-22.

Kevoe-Feldman, H., & Pomerantz, A. (2018). Critical timing of actions for transferring 911 calls in a wireless call center. *Discourse Studies, 20*(4), 488-505.

Kidwell, M., & Kevoe-Feldman, H. (2018). Making an impression in traffic stops: Citizens' volunteered accounts in two positions. *Discourse Studies*, *20*(5), 613-636.

Kim, H., & Billington, R. (2018). Pronunciation and comprehension in English as a lingua franca communication: Effect of L1 influence in international aviation communication. *Applied linguistics*, *39*(2), 135-158.

Kim, M. S. (2017). The practice of praising one's own child in parent-to-parent talk. *Discourse Studies*, *19*(5), 536-560.

Lambert, C., Philp, J., & Nakamura, S. (2017). Learner-generated content and engagement in second language task performance. *Language Teaching Research*, *21*(6), 665-680.

Lawson, M. (2017). Negotiating an agentive identity in a British lifestyle migration context: A narrative positioning analysis. *Journal of Sociolinguistics*, *21*(5), 650-671.

Licoppe, C., & Morel, J. (2018). Visuality, text and talk, and the systematic organization of interaction in Periscope live video streams. *Discourse Studies*, *20*(5), 637-665.

Linan-Thompson, S., Degollado, E. D., & Ingram, M. D. (2018). Spelling it out, one por uno: patterns of emergent bilinguals in a dual language classroom. *TESOL Journal*, *9*(2), 330-347.

Little, A., & Fieldsend, T. (2018). Teaching the passive through semantically enhanced input. *TESOL Journal*, *9*(1), 138-159.

Macgregor, A., & Folinazzo, G. (2018). Best practices in teaching international students in higher education: Issues and strategies. *TESOL Journal*, *9*(2), 299-329.

Marshall, S., & Marr, J. W. (2018). Teaching multilingual learners in Canadian writing-intensive classrooms: Pedagogy, binaries, and conflicting identities. *Journal of Second Language Writing*, *40*, 32-43.

McArthur, A. (2019). Pain and the collision of expertise in primary care physical exams. *Discourse Studies*, *21*(5), 522-539.

McDonough, K., Crawford, W. J., & De Vleeschauwer, J. (2014). Summary writing in a Thai EFL university context. *Journal of second language writing*, *24*, 20-32.

Mendoza-Denton, N., Eisenhauer, S., Wilson, W., & Flores, C. (2017). Gender, electrodermal activity, and videogames: Adding a psychophysiological dimension to sociolinguistic methods. *Journal of Sociolinguistics*, *21*(4), 547-575.

Millar, N. (2011). The processing of malformed formulaic language. *Applied Linguistics*, *32*(2), 129-148.

Mondada, L. (2018). The multimodal interactional organization of tasting: Practices of tasting cheese in gourmet shops. *Discourse Studies*, *20*(6), 743-769.

Nagle, C. L. (2019). A longitudinal study of voice onset time development in L2 Spanish stops. *Applied Linguistics*, *40*(1), 86-107.

Neumann, H. (2014). Teacher assessment of grammatical ability in second language academic writing: A case study. *Journal of Second Language Writing*, *24*, 83-107.

Ong, J. (2014). How do planning time and task conditions affect metacognitive processes of L2 writers?. *Journal of Second Language Writing*, *23*, 17-30.

Payant, C., & Reagan, D. (2018). Manipulating task implementation variables with incipient Spanish language learners: A classroom-based study. *Language Teaching Research*, *22*(2), 169-188.

Pienemann, M. (1989). Is language teachable? Psycholinguistic experiments and hypotheses. *Applied linguistics*, *10*(1), 52-79.

Polat, R. K. (2018). Religious solidarity, historical mission and moral superiority: construction of external and internal 'others' in AKP's discourses on Syrian refugees in Turkey. *Critical Discourse Studies*, *15*(5), 500-516.

Polio, C., & Shea, M. C. (2014). An investigation into current measures of linguistic accuracy in second language writing research. *Journal of Second Language Writing*, *26*, 10-27.

Raymond, C. W. (2018). On the relevance and accountability of dialect: Conversation analysis and dialect contact. *Journal of Sociolinguistics*, *22*(2), 161-189.

Rolander, K. (2018). Family literacy: A critical Inquiry–Based approach to English language acquisition. *Tesol Journal*, *9*(1), 58-75.

Rott, S., Williams, J., & Cameron, R. (2002). The effect of multiple-choice L1 glosses and input-output cycles on lexical acquisition and retention. *Language teaching research*, *6*(3), 183-222.

Schubert, C. (2019). 'OK, well, first of all, let me say…': Discursive uses of response initiators in US presidential primary debates. *Discourse Studies*, *21*(4), 438-457.

Sharma, B. K. (2018). English and discourses of commodification among tourism workers in the Himalayas. *Journal of Sociolinguistics*, *22*(1), 77-99.

Siyanova-Chanturia, A., & Martinez, R. (2015). The idiom principle revisited. *Applied Linguistics*, *36*(5), 549-569.

Strid, J. E. (2017). The myth of the critical period. *TESOL Journal*, *8*(3), 700-715.

Svendsen, B. A. (2018). The dynamics of citizen sociolinguistics. *Journal of Sociolinguistics*, *22*(2), 137-160.

Talib, N., & Fitzgerald, R. (2018). Putting philosophy back to work in critical discourse analysis. *Critical Discourse Studies*, *15*(2), 123-139.

Tanaka, J., & Gilliland, B. (2017). Critical thinking instruction in English for academic purposes writing courses: A dialectical thinking approach. *TESOL Journal*, *8*(3), 657-674.

Van Leeuwen, T. (2018). Moral evaluation in critical discourse analysis. *Critical Discourse Studies*, *15*(2), 140-153.

Widdicombe, S. (2017). The delicate business of identity. *Discourse Studies*, *19*(4), 460-478.

Yamagata, S. (2018). Comparing core-image-based basic verb learning in an EFL junior high school: Learner-centered and teacher-centered approaches. *Language Teaching Research*, *22*(1), 65-93.