A COMPARATIVE ANALYSIS OF RESEARCH ABSTRACTS WRITTEN BY NOVICE AND PROFESSIONAL WRITERS: A SYNERGY OF GENRE-BASED AND CORPUS-BASED APPROACHES

VANLEE SIRIGANJANAVONG

A thesis submitted to Newcastle University for the degree of

(Integrated) Doctor of Philosophy

in Educational and Applied Linguistics

School of Education, Communication and Language Sciences

Faculty of Humanities and Social Sciences

October 2019

Abstract

Research abstracts can serve as both informational and promotional tools to attract readers' attention. Although numerous studies have confirmed that there are disciplinary differences in the rhetorical patterns commonly found in research abstracts, little attention has been paid to the comparison of abstracts written by professional and novice writers. In this study research abstracts by professional and novice writers in the Education and Applied Linguistics fields were analyzed. The English for Specific Purposes (ESP) approach to textual analysis was used to investigate the rhetorical move patterns of the abstracts, while a corpus-based approach was adopted to examine lexical choices. Two corpora were compiled: a Novice Abstract Corpus (NAC) consisting of 150 abstracts written by Master's degree students, and a Professional Abstract Corpus (PAC) consisting of 378 abstracts taken from four international peer-reviewed journals.

By using the five-move analysis framework developed by Hyland (2000), similarities and differences in the abstracts written by the two groups were examined in relation to three different issues: (1) the various combinations of moves that can be involved in the composition of an abstract; (2) the different structural patterns or orders of the moves; and, (3) the status of the various kinds of moves, in terms of whether they are optional, conventional or obligatory. The findings revealed that a greater variety of move compositional types were found to occur in the PAC. The result of a chi-square test showed that there is a statistically significant difference between the two corpora in the frequency of the seven compositional types found in both corpora**.** Similarly, the results showed that there is a greater variety of structural patterns found in the PAC than in the NAC (163 patterns: 33 patterns). In addition, variation in the frequency of occurrence of the five types of moves across the two corpora demonstrated that the NAC and the PAC exhibited differences in the status of only one move; that is, the Product Move was found to be obligatory in the NAC, but only conventional in the PAC.

After the move analysis, all the abstracts in both corpora were tagged according to their moves. The lexis of the two corpora and their constituent moves was analyzed, using corpus software to generate wordlists, keyword lists, and lists of recurring n-gram clusters (or lexical bundles). The word frequency lists revealed some similarities and differences between the top twenty most frequent lexical words in each corpus. The degree of similarity and difference

was evaluated using an effect size measure (log-ratio). In comparing the frequency of words in the two corpora in order to identify keywords in each, NAC keywords with zero occurrence in the PAC were identified that represent characteristic aspects of the research focus of the novice writers. The results of the lexical bundles analysis revealed a number of distinctive features in each specific move, such as how novice and professional researchers differently motivate their research in the Introduction Move. Pedagogical implications are discussed, in terms of how the results of the study can be used to raise teachers' and students' awareness in relation to conventional structural patterns and the use of lexis in abstract writing, to help prepare novice researchers for scholarly publications.

Declaration

I declare that the work presented in this thesis is my own. I have correctly acknowledged the work of others, and no part of the material has been previously submitted, either in whole or in part, for the award of a higher degree elsewhere.

Name: *Vanlee Siriganjanavong*

Date: ……

Acknowledgements

The thesis would not have been completed without the assistance and support from many people. First and foremost, I would like to express my heartfelt gratitude to my supervisors, Dr. Mei Lin and Dr. Adam Mearns for all their continuing and relentless support, encouragement, insightful feedback, and greatest of all, patience. Their kindness and their sense of humor always made the supervisory meetings more memorable than stressful. Their thorough comments and feedback tremendously helped in shaping and improving my thesis. Dr. Mei Lin has assisted me with her critical and thought-provoking ideas. With her keen eye, she was able on numerous occasions to pinpoint details that I had overlooked. Dr. Adam Mearns has provided me with his time, generous support and constructive suggestions, especially on technical issues regarding corpus linguistics and statistics. His guidance on move tagging, text cleaning by NotePad++, and many more are truly appreciated. I feel blessed and fortunate to have had them as my supervisors.

In addition, my appreciation and thanks go to the internal examiner, Professor Paul Seedhouse, and the external examiner, Dr. Nicholas Groom, for examining the thesis and providing me with insightful comments. Their feedback and suggestions greatly helped improve my thesis.

Also, I am greatly beholden to Thammasat University, which provided me with a full scholarship to undertake my PhD studies. Without this funding, it would have been impossible for me to pursue my studies at Newcastle University. In addition, I would like to extend my gratitude to Associate Professor Dr. Supong Tangkiengsirisin and Associate Professor Dr. Pornsiri Singhapreecha for their encouragement, both before and during my studies. My special thanks also go to my PhD friends, Chanuwan Phetkla and Yuwaree Tongvivat, who assisted me in being the inter-coders of my pilot study.

Two Intensive Summer Schools and a Future Learn online course in corpus linguistics provided by Lancaster University have widened my knowledge in corpus linguistics. They also provided me with rare opportunities to study and meet with world-renowned experts in the field. I would not have known such wonderful summer schools without a recommendation from Dr. Rachelle Vessey, my year-long former supervisor. Moreover, I am indebted to Professor Lawrence Anthony, who created the user-friendly corpus software, AntConc, which was used in the study. The continuous development of the

software and the time devoted for answering queries that users, including me, may have in Google Forum proved invaluable.

My gratitude also goes to Dr. Gareth Williams, always an empathetic listener acting as a consultant, in addition to his role as my tutor. Not to forget, I also would like to mention my beloved friend in Thailand, Sunee Treetanakorn, who helped me when I was in need. No matter how busy you were, I realized that whenever I needed help, you were there for me unquestionably. My PhD journey would not have been so memorable without many lovely friends in the UK, who shared our thoughts, experiences, and meals as well as helped to tackle the challenges that we came across. The happiness and fun we shared will always be cherished.

Last but not least, I would like to thank my family: my dad, my mum, my sister (Jamjun) and my cousin (Kulaporn) for their support throughout the long journey of my PhD studies. Big thanks to my sister. Despite her busy working schedule, she took great care of mum and dad so that I could focus on my studies. The perseverance I have learned from them backed me up right to the finish line. My parents always told me when I was young that the only treasure they could give me is education. Therefore, this thesis is dedicated to them, my first teachers.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

BNC	British National Corpus
BAWE	British Academic Written English Corpus
CL	Corpus Linguistics
CR	English for Careers Program
EAP	English for Academic Purposes
ELF	English as a lingua franca
EL	English Language Teaching Journal
ELT	English Language Teaching
em	embedded moves
ES	English for Specific Purposes Journal
ESP	English for Specific Purposes
L1	first language
L2	second language
LB	lexical bundle
LITU	The Language Institute, Thammasat University
LL	Language Learning Journal
Log-L	Log-likelihood
Log-R	Log-ratio
m1	Introduction Move
m2	Purpose Move
m3	Method Move
m4	Product Move
m5	Conclusion Move
MA	Master of Arts
NAC	Novice Abstract Corpus
NNS	Non-native speaker
NR	New Rhetoric
NS	Native speaker
PAC	Professional Abstract Corpus
pmw	per million words
RA	Research article
RAB	Research abstract
SFL	Systemic Functional Linguistics
TE	Teaching English as a Foreign Language Program
TQ	TESOL Quarterly Journal

Chapter 1. Introduction

1.1 Overview

This introductory chapter aims to provide a background to the thesis in order to contextualize the study. Section 1.2 provides background information on the rationale for the study and also on the importance that the Thai educational system places upon the English language. Section 1.3 presents the aims of the study and research questions. Section 1.4 outlines the significance of the study by pointing out a number of research gaps in the existing literature and the expected contributions of the study. Section 1.5 presents the outline of the thesis.

1.2 Background and Rationale of the Study

Section 1.2.1 provides brief details regarding the importance of the English language in the Thai educational system and recent economic and education reforms before addressing the two primary motivations for this research: (1) The English for Specific Purposes (ESP) approach as a response to the education and economic reforms reviewed in Section 1.2.2; and, (2) The role of research abstracts as a requirement for the dissemination of knowledge in Section 1.2.3.

1.2.1 English in Thailand and the changes in economic and education reforms

English is acknowledged as a world language, or lingua franca. Although it is not the world's top spoken language in terms of total number of native speakers, Kachru and Nelson (2006) assert that English is the most widely taught, learned, and spoken language in the world. The estimation of the total number of English speakers is approximately 1.75 billion people worldwide, a number expected to reach two billion people by 2020 (The British Council, 2013). The rapidly growing importance of English as a means of communication and commerce in Southeast Asia was already well recognized before the inauguration of the ASEAN Economic Community (AEC)[1] in 2015. The Association of

[1] AEC consists of 10 member states in Southeast Asia as follows: (1) Brunei, (2) Cambodia, (3) Indonesia, (4) LAO People's Democratic Republic, (5) Malaysia, (6) Myanmar, (7) Philippines, (8) Singapore, (9) Thailand, and (10) Vietnam.

ASEAN Charter, signed in February 2009, officially declared English to be the sole working language of communication (Kirkpatrick, 2014).

Compared with citizens in postcolonial countries, where English is regarded as (one of) the national language(s), it is apparent that people in Thailand, which was the only non-colonized country in the Southeast Asia region, generally have a lower level English proficiency. The change in the National Education Act (NEA) of 1999 illustrates how Thailand has attempted to prepare its citizens to respond to the economic developments associated with globalization. In the 1999 Act, apart from ensuring at least 12 years of mandatory basic education for all, the government also initiated the use of English as a medium of teaching and learning in higher education (Kaur, Young and Kirkpatrick, 2016). This led to an increasing number of universities offering international programs, and by 2004 some 465 programs were offered by public and private higher education institutions at undergraduate and postgraduate level (Lao, 2015). A decade later, figures from the Office of Higher Education Commission (OHEC, 2016) showed a remarkable internationalization of Thailand's higher education institutions, with a total number of 769 international programs using English as the medium of instruction. Among these programs, 290 programs are at Master's level and 224 at PhD level.

1.2.2 The ESP approach as a response to the economic and education reforms

In response to the economic and education reform, it is apparent that ESP, such as English for Academic Purposes (EAP) and English for Business, has become a major means of improving student literacy in preparation for the changing and complex world of knowledge. Hyland (2009b, p. 4) points out that the growing interest in academic discourse, which has partly evolved into a developing research area during the last few decades can be attributed to three reasons: (1) widening educational access policies, which have brought about an increasing diversity of the students entering higher education; (2) the increased attention paid to instructional process by funding bodies; and, (3) the recognition of English as the lingua franca in research and scholarship. The benefit of using English as the 'hegemonic' language in academic research is widely held to be that it allows the rapid dissemination of knowledge on a worldwide scale and allows the formation of a cross-cultural understanding of peoples and nations (Flowerdew, 1999, p. 244). However, at the same time, an emphasis on using English poses considerable

challenges for non-native scholars of English, while (inadvertently) offering advantages to their native counterparts (ibid., 1999).

1.2.3 The role of research abstracts for the dissemination of knowledge

Recently, an increasing number of universities in Thailand have established a new requirement for the degree completion of postgraduate students; that is, in order to reach a wider listenership or readership, their research must be disseminated through either publications or conferences (see highlighted in Appendixes A-B for graduation requirements for MA programs targeted for data compilation). Only a small number of MA students publish their work (or even plan to publish) in leading international peer-reviewed journals. However, at the Language Institute, where the dataset of the novice researchers was compiled, an annual graduate student conference has been organized to provide an opportunity for all the MA students to present their research findings and published their proceedings since 2012. It should be noted that in a number of other countries such as Iran, China, and the Netherlands, the requirement for the graduation are even more rigorous, with the only option being to have parts of a doctoral dissertation published in an academic journal (Paltridge and Starfield, 2016, p. xii). The publication policy, however, is still not a requirement for degree completion in a number of countries such as the U.K. and the U.S.

It is common practice that the contents of a research paper or a research article (RA henceforth) must be preceded by an abstract. As English is regarded as 'the language of international scholarship and research' (Flowerdew, 1999, p. 243), postgraduate students in most universities in Thailand are required to have their research abstracts (RAB hereafter) written bilingually, in English and Thai. The practice of having bilingual abstracts (students' native language and in English) is also common in other countries, such as China (Ren and Li, 2011) and the Czech Republic (Klimova, 2013). During the screening process, there is a tendency for editors or journal reviewers to make a preliminary decision on publication from reading the RABs.

Many researchers (e.g., Kanoksilapatham, 2005, 2009; Khedri, Heng and Ebrahimi, 2013; Ren and Li, 2011; Saeeaw and Tangkiengsirisin, 2014; Salager-Meyer, 1992; Supatranont, 2012; Swales and Feak, 2009) have confirmed that what is written in a RAB plays a crucial role in readers' judgment with respect to whether to read the full content

of that article. With the aim of discovering relevant sources to support their work and keeping abreast of current research trends, scholars are likely to select RAs as their preferred reference option because they provide more concise and up-to-date information than books and other materials. During the article selection process, after the title, the abstract of a journal article is likely to be the very first thing that readers take into consideration. Abstracts therefore serve as a 'hook' to attract the attention of readers. Further details regarding the importance of abstracts as an academic genre are provided in Section 2.91.

Cortes (2011) asserts that writing for scholarly publications requires special attention and endeavor, especially for novice researchers, because reporting a study that can convey the researcher's viewpoints is especially challenging. Despite the recognition of its significance, it is surprising that so few student researchers, who are defacto novice researchers, are specifically trained how to produce quality abstracts. One possible explanation might be that RABs are probably viewed as a general overview or merely a summary of the research, and not as a specific genre that warrants special attention. Some students, therefore, have to self-study or follow exemplars by reading previous years' graduate work without developing a systematic and comprehensive understanding as to how to compose an effective, high-quality abstract.

To date, few studies have compared the similarities and differences between novice and professional writers with regard to their RABs. The majority of these studies compile data on RA abstracts from renowned international journals (e.g., Hu and Cao, 2011; Kanoksilapatham, 2009; Pho, 2008; Saeeaw and Tangkiengsirisin, 2014; Samraj, 2005; Supatranont, 2012; Tankó, 2017) and make comparisons between disciplines (e.g., Kanoksilapatham, 2009; Pho, 2008; Samraj, 2002; 2005; Supatranont, 2012), between different scholarly text types (namely research papers, reviews, and case reports) (Salager-Meyer, 1992), or between different languages (Hu and Cao, 2011; Kanoksilapatham, 2007; Martín, 2003). One exception is Ren and Li' s (2011) study, which compared abstracts from Chinese Master's English theses with abstracts published in international journals. However, although this study compared novice writers with established academics, the lexical usage and rhetorical moves used by Chinese novice writers might differ from their Thai counterparts.

1.3 Aims of the Study

This study thus aims to (1) analyze the rhetorical patterns of English language abstracts written by novices and experts in the fields of Education and Applied Linguistics; and, (2) examine the keywords and lexical bundles used in those abstracts. The compilation of abstracts written by the two groups results in two corpora, referred to in this study as a novice abstract corpus (NAC) and a professional abstract corpus (PAC). Feak and Swales (2009) regard doctoral students as novice researchers, but in this study, the term 'novice' is defined as those postgraduate students undertaking a Master's degree at a university in Thailand. They are expected to have no experience of writing RABs for international peer-reviewed journal publications. Contrastingly, the term 'professional researchers' is used to define writers whose work has been accepted for publication in four international journals, namely English Language Teaching Journal, English for Specific Purposes, Language Learning, and TESOL Quarterly. It is assumed that these researchers have been trained in RA writing or at least have gone through a rigorous review and revision process, as described in Paltridge and Starfield (2016, pp. 62-64).

To achieve the aims of the study, the following research questions are addressed:

Overarching questions:

- Are there any differences between the rhetorical moves and linguistic patterns of the NAC and those of the PAC?[2] If yes, what are their pedagogical implications?

[2] The analytical methods in this thesis were based not only on move analysis but also on various corpus analytical techniques. Although rhetorical moves and the textual boundaries appear to be cognitively defined (Paltridge, 1994), the evidence shown in concordance lines from corpus analysis reveals the linguistic patterns of each move, which can be vertically and horizontally read and analyzed (Tognini-Bonelli, 2010).

Sub-questions:

1. What are the similarities and differences between the rhetorical moves found in abstracts written by novice researchers and those written by professional researchers in the fields of Education and Applied Linguistics?
2. To what extent are the most frequent lexical words in the NAC similar to those in the PAC?
3. What keywords are associated with each move in (a) the NAC and (b) the PAC?
4. What lexical bundles are associated with each rhetorical move in (a) the NAC and (b) the PAC?

1.4 Significance of the Study

The study is expected to contribute to a deeper understanding of the move compositional types[3] and the rhetorical move patterns[4] regarded as conventional practices among novice and professional researchers in their abstract writing. With regard to the number of the RABs, many studies have drawn on either very small samples, such as 12 RA introductions and abstracts, each from two different fields (Samraj, 2005), 25 expert-written and 25 novice-written abstracts (Ren and Li, 2011), inconsistent numbers of RABs across two or more disciplines (Supatranont, 2012), or inconsistent numbers of RABs across text types (Salager-Meyer, 1992). These shortcomings highlight the problem of representativeness, which this study seeks to remedy by investigating the largest possible number of RABs that the time constraints permit. The NAC consists of 150 abstracts written by Master's students in two disciplines (Education and Applied Linguistics), while the PAC consists of 378 abstracts published in four international peer-reviewed journals (see more details in Section 4.6). Because the PAC is used as a reference corpus in the study, the number of abstracts is thus approximately double the size of that in the NAC. The reason why other available corpora, such as the British Academic Written English Corpus (BAWE), is not used as a reference corpus in this

[3]Move compositions refers to how moves as constituents are combined in each abstract without considering and counting the recurrence of the same move if that specific move occurs more than once. The samples of how move compositional types were classified are shown in Section 5.3.

[4]Rhetorical move pattern refers to the pattern in which all the sequence of moves, including the recurrence of moves in each abstract, are identified. The samples of how rhetorical move pattern were classified are shown in Section 5.3.

study because it contains different genres, mainly student assignments (e.g., essays, critiques, and case studies). In contrast, the focus of this study is to compare the linguistic features found in RABs written by novice and professional writers, not to compare and contrast academic writing in general.

With respect to embedded moves[5], some studies (e.g., Dos Santos, 1996; Hyland, 2000; Pho, 2008; Saeeaw and Tangkiengsirisin, 2014) have pointed out that embedded moves are often found in abstracts. However, a dearth of evidence has been found with respect to how and where these embedded moves commonly occur. Therefore, the study is intended to bridge the gap by finding the answers to these queries in the process of move analysis.

The second aim of the study is to investigate the similarities and differences of lexical choices found in the NAC and the PAC using corpus tools, namely a wordlist, a keyword list, and cluster/n-gram functions. With a wordlist list function, the findings reveal the most frequently found function and lexical words. With a keyword list function, the findings might shed some light on lexical choices either exclusively found or unusually frequent in the NAC, reflecting the small culture of the group. The lexical choices appearing exclusively in the PAC and those frequently found in the PAC relative to the NAC might somehow reflect the features that are commonly found and deemed 'acceptable' within a specific professional research community. With cluster/n-gram functions, lexical bundles in each move can be found.

As pointed out by Creswell (2012) that improvements for practice can be guided through research, the findings of the comparative analysis are applicable in several ways. For example, they can be used to raise awareness of different move compositions and patterns in writing an abstract and in training postgraduate students how to write English language abstracts with appropriate lexical choices and the accepted organization of rhetorical moves. Additionally, those who aspire to disseminate their papers in peer-reviewed journals in Education and Applied Linguistics will be able to learn how to adjust their

[5] Embedded moves refer to the adherence or combination of two or more moves in the same sentence. For example, while the main clause may signify Introduction (m1), the subordinate clause or the attached prepositional phrase may represent another communicative purpose such as Purpose (m2) or Method (m3).

abstract writing to conform to international standards, thereby increasing the chance that their papers will be accepted for publication.

1.5 Outline of the Thesis

The thesis consists of eight chapters. After the introduction in **Chapter 1, Chapter 2** defines *genre* and discusses interrelationships among *genre*, *register* and *discipline*. Three theoretical approaches to genre theory, namely English for Specific Purposes (ESP), Systemic Functional Linguistics (SFL) and New Rhetoric (NR) will be explored with a special emphasis on genre and ESP, where the current study is situated. The chapter also touches on the notion of '*culture*', which is embedded in the rhetorical writing patterns of English non-natives. The chapter then discusses why Hyland's (2000) analytical framework has been adopted for the move analysis. The final part looks at the importance of abstracts as an academic genre and related previous studies on abstract writing.

Chapter 3 discusses key issues in corpus analysis: (1) the use of general versus specialized corpora; (2) corpus-based versus corpus-driven approaches; and, (3) keyword versus key keywords. The chapter then provides the theoretical background to corpus approaches. The chapter describes how keywords can represent specific characteristics of each corpus. The importance of lexical bundles in abstract writing will also be discussed. The chapter will finally draw on related studies on keywords and lexical bundles.

Chapter 4 details the study's research methodology, which combines genre-approach and corpus-based approach. The chapter begins with the overview of three research paradigms, followed by the ontological and epistemological stance of the study. Then, the theoretical framework and research design are highlighted. The research context of novice researchers is provided before turning to the details of the NAC and the PAC and the procedures in textual and corpus analyses.

Chapters 5 addresses RQ 1 by examining the similarities and differences between the rhetorical moves found in abstracts written by novice researchers and those written by professional researchers in the fields of Education and Applied Linguistics by adopting Hyland's (2000) framework. A number of samples of how the classifications of move

patterns and move compositional types are presented. The analyses reveal the move compositional types and move patterns found in the NAC and the PAC.

Chapter 6 addresses RQs 2-3 by firstly looking at the frequency lists of the NAC and the PAC to identify the similarities and differences in the use of frequently found lexical words. Then, some of these lexical words are examined in-depth by exploring their concordance lines in order to investigate their meaning in context. The keywords found in the NAC and the PAC (with zero occurrence relative to another corpus) are also examined.

Chapter 7 addresses RQ 4 by identifying lexical bundles in each move of the NAC and the PAC. Then, these lexical bundles are further classified into functional and structural categories.

Chapter 8 summarizes and discusses the key findings as well as addressing certain methodological issues. The implications for further research and teaching are also offered.

Chapter 2. Issues in Genre Analysis

2.1 Introduction

This chapter has four main aims: (1) to provide an overview of how the English for Specific Purposes (ESP) approach has been used in textual analysis and research; (2) to discuss why the ESP approach to genre analysis was employed for the analysis in comparison with two other approaches, namely Systemic Functional Linguistics (SFL) and New Rhetoric (NR); (3) to review the notion of culture in second language writing and how the term should be interpreted in the study; and, (4) to discuss why Hyland's framework for move analysis was adopted. The chapter also reviews the literature related to using move analysis for abstract and research article writing.

To begin with, Section 2.2 explores the various definitions of *genre* that have been presented in the literature. Section 2.3 describes certain issues related to discourse community and its membership. Section 2.4 clarifies the interrelationships between genre, register, and discipline, and explains how they overlap or are used interchangeably. Section 2.5 discusses genre theory and three different approaches to genre analysis: ESP, upon which the present study is based, SFL, and NR. The discussion of the cultural aspects of genre in Section 2.6 demonstrates how those aspects play a role in second language writing. Section 2.7 describes how genre-based approaches can benefit language teaching and learning. Section 2.8 looks at different frameworks that can be applied to move analysis as a version of textual analysis and specifically at Hyland's (2000) analytical model. Section 2.9 starts by explaining how important research abstracts are as an academic genre before examining the literature on the use of move analysis in research on abstract writing.

I will argue that despite situating within the ESP approach to move analysis, the cultural aspect, one of the features in SFL and NR, should be included in the analysis.

2.2 Definitions of Genre

Different definitions of the term *genre* are provided by influential genre theorists such as John Swales, Ken Hyland, J.R. Martin, and others. For example, the classic definition provided by Swales (1990, p. 58) describes a genre as:

> a class of communicative events, the members of which share some set of communicative purposes. These purposes are recognized by the expert members of the parent discourse community, and thereby constitute the rationale for the genre.

In addition to the communicative purposes that are highlighted in this definition, Swales further explains that what individual genres tend to share is a set of structural patterns, style, content, and a target readership. A somewhat similar definition is provided by Hyland (2009a, p. 245), who describes *genre* as 'a set of texts that share the same socially recognized purpose and which, as a result, often share similar rhetorical and structural elements to achieve this purpose'. Influenced by the work of SFL, Martin (1984, cited in Paltridge, 2012, p 64) defines genre as 'a staged, goal-oriented, purposeful activity in which speakers engage as members of our culture'. It is noticeable that what these three definitions have in common is the notion that genre studies should address at least the following major components: rhetorical and structural patterns, communicative purposes, and the target discourse community. However, Martin's definition involves the debatable term *culture*, raising some questions as to what it actually means or refers to. The notion of *culture*, therefore, will be discussed later in Section 2.6.3. In this study, genre is defined as a set of conventionalized rhetorical and structural patterns as well as lexis which are realized and used by discourse members to convey their communicative purposes in accordance with the target readership. Both novice and professional abstracts are defined as the same genre which is an academic genre. This means that their target readers are those in academia. The rhetorical and structural patterns of the abstracts can also reflect the level of expertise in the field.

Having mentioned that some definitions of *genre* involve the features of discourse community and its membership, the next section will look at some characteristics and the relationship between them.

2.3 Discourse Community and its Membership

Certain characteristics have been proposed by Swales (1990) in order to identify a discourse community—the term which was first coined by Martin Nystrand (Swales, 2016). Two of these characteristics deserve attention here as they are relevant to this study. First, a discourse community, in Swales' (1990, p. 24) conceptualization, must have 'a broadly agreed set of common public goals'. These goals may be open to the public by being formally written in documents, or they may be tacit knowledge. If an analogy between Swales' conceptualization of discourse community with the publication process is drawn, it is apparent that both novice and professional writers should acknowledge the agreed set of rules written in the submission guidelines. However, some tacit (hidden) rules are recognized exclusively by those established members of the community, namely the professional researchers.

Second, a discourse community contains certain specific lexis (Swales, 1990). The specialized terminology can also be seen in the creation of community-specific abbreviations or acronyms such as EFL, ESL, EAP, etc. While using these specific terms promotes effective communication among well-established members of the community, they can also create a certain amount of bafflement for outsiders. As Hyland (2009b, p. 48) puts it, 'community conventions are therefore also a means of fostering group mythologies, solidarity and social control, helping to ring-fence communities by identifying users as insiders and excluding others'.

Swales (2016, p. 7) introduced three types of discourse community, namely, 'local', 'focal', and 'folocal'. The first one refers to people in the same area who work at the same place or at the same occupation. They create and acquire some abbreviations, acronyms and special words to facilitate their communication. However, these words are probably incomprehensible to the outsiders. The behavior conventions and expectations are only realized among group members. The second one, 'focal', refers typically to the regional, national or international associations. Examples of such focal communities include BAAL (the British Association of Applied Linguistics) and TESOL International Association. The groupings can be either formal or informal groupings, where membership fees might be required. The members of these associations receive any updated news or latest developments through websites or participate in a conference. The last type of discourse community is the mixture of both, and their members have a double

commitment. A clear example can be seen from members of a university who understand the requirements and conventions of their own department, while they are also recognized as scholars who participate in conference presentation. A conflict may arise between the local and focal demands on their time, especially for those who work in higher education organizations where there has been an increasing pressure for publications.

Paltridge (2012, pp. 16-17) points out a number of ways to differentiate between novice and professional members of a discourse community. One way is to look at how long a member has engaged with the community and how much knowledge of the particular genre a member has. It is very likely that the degrees of membership develop over time. Another way of judging the degrees of membership is by looking at how close the relationships are among the members. Two types of relationships are identified: close-knit groups and loose-knit groups. Moreover, Paltridge claims that one member might belong to multiple groups which are overlapping. For example, a person who has been working for a university for some time (a senior member of a faculty) may be starting a doctoral degree (a new member of a community). As a result, the linguistic repertoire associated with the two communities s/he is involved with might differ.

Bhatia (2014) proposes a four-space model of genre analysis which looks at language as text, language as genre, language as professional practice and language as social practice. Despite different orientations among these four perspectives towards language, Bhatia concludes that they share certain characteristics. The highlighted ones relevant to the present study are the two descriptions regarding the differentiation between novice and professional members of the community. First, he affirms that the members of a particular discourse community possess greater knowledge and understanding of the genre associated with their discourse community. They are more competent and flexible in the use and exploitation of that specific genre when compared to the novices, newcomers or outsiders. The second description is related to the issue of conventionalized structures of genres. Interestingly, Bhatia points out that the exploitation of genres tends to occur with expert members who intentionally break the conventional practices in order to create ‘private’ constructs. In this respect, convention integrity is viewed as ‘contestable, negotiable and developing’ (Bhatia, 2014, p. xii). It

is therefore intriguing to see whether Bhatia's observation regarding the violation of conventional practices is shown in the professional group in this study.

In order to provide a clearer understanding relating to genre, the distinctions between the key terms *genre*, *register*, and *discipline* should also be clarified. The interrelationships among these three key terms in genre studies will be discussed in the section that follows.

2.4 Genre, Register, and Discipline

The three terms—*genre*, *register*, and *discipline*—though interrelated and sometimes overlapping, are not the same, and might create confusion for some language teachers and researchers. Historically, the development of the term *genre* evolved from the concept of *register* (Swales, 1990). Hyland (2009a) further elucidates that *genre* represents the way language is used to respond to recurring situations. Each genre consists of a number of characteristic features, making one genre different from another. Those features include a specific purpose, an organizational pattern, and specific linguistic features. More importantly, legitimate members of the community are able, of course, to recognize, understand, and use those features appropriately.

Textual analysis can take the form of either *register* or *genre perspectives*, and there may be variations according to different academic disciplines. Biber and Conrad (2009) claim that the two perspectives are alike in that both include the communicative purposes and situational contexts of the texts in the analysis. However, a number of distinctions can be made between the two perspectives. One of the differences lies in the focus of their linguistic analysis. ***The register perspective*** encompasses an analysis of linguistic properties that are found to be pervasive and conventional in a text variety, such as the pronouns and verbs that are specific to the relevant communicative purposes and situations. ***The genre perspective***, in contrast, concentrates on the conventional structures employed in a text construction—that is, how the text begins and ends—which 'conform to the culturally expected way of constructing texts belonging to the variety' (ibid., p. 16). The linguistic properties found in the genre perspective are not pervasive because they sometimes occur only once but in a certain position of the text, such as *Dear*… and *Sincerely yours* in business letters. This means the features are not distributed throughout the texts, but are distinctive, normally formulaic (ibid., 2009).

Another difference between the register and genre perspectives is found in the textual focus. While an excerpt of a text can be analyzed with ***the register perspective***, it cannot be analyzed with ***the genre perspective*** as some linguistic features occur only once in a certain position in the text, just like the case of business letters. Therefore, any texts can be analyzed from the register perspective, but only complete texts can be analyzed from the genre perspective (Biber and Conrad, 2009, pp. 17-18). Considering that the main aim of the present study is to analyze the complete abstracts of novice and professional writers in terms of their move compositional types and move patterns, it can therefore be regarded as being situated in the genre perspective. However, the complete abstract is also divided into moves for the examination of the lexical choices found in each move. This aspect of the study can thus be viewed through the register perspective. Taken together, it can be argued that these two perspectives are situated along a continuum and cannot be entirely seperated from each other. In conclusion, the present study should be viewed from both perspectives.

It is understandable that a blurred line exists between the terms *register* and *discipline* because they are closely related and occasionally used interchangeably. For example, students who are studying at a business school (*business discipline*) should learn how to write a business report (*academic*/*professional genre*) by employing business terminology (*business register*). In this respect, the term *discipline* is used to divide the fields of study, for instance, into medical, business, and scientific.

Bhatia (2014, p. 36) argues that 'genres cut across disciplines', as shown in Figure 2.1 (overleaf) because genres are used to differentiate specific types of social activities and communicative purposes within the same register or discipline. Research abstracts (RABs) as an academic genre are likely to contain a number of both specific and shared features across disciplinary boundaries. For instance, shared features might include the organization of sections in terms of an introduction, literature review, methodology, results, discussion and conclusion. Although the titles of the chapters or sections of an article may differ (such as *guiding literature* instead of *literature review*), the contents of articles that have broadly similar communicative purposes might be arranged in similar ways, or with similar structures, despite belonging to different disciplines. Despite these sorts of shared characteristics across disciplines, other features are found to be exclusive to specific disciplines (Hyland, 2008a, 2012a; Peacock, 2012).

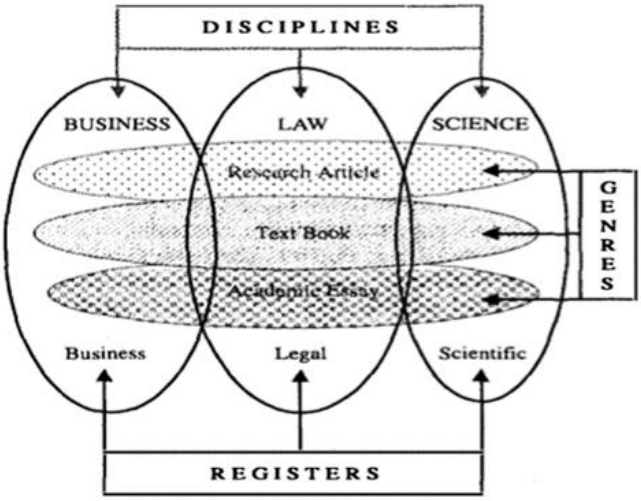

Figure 2.1 The interrelationships between registers, genres and disciplines in academic discourse (Bhatia, 2014, p. 36)

Halliday *et al.* (1964, cited in Bhatia, 2014) describe the relationship between the text and the contextual elements below as being the key to the language patterns analysis, which can be conducted in terms of text structure and lexico-grammatical patterns (Paltridge, 1997, p. 23). Halliday *et al.* (1964, cited in Bhatia, 2014) suggest that these three contextual aspects can be divided into:

- ***field*** - the theme or topic of the text;
- ***tenor*** - the text-participants' relationship as well as the formality level; and
- ***mode*** - the medium used in communication.

Given the description above, it should be noted that the current study focuses attention on the academic written genre, and specifically the genre of research abstracts (RABs). Other studies (e.g., Ren and Li, 2011; Samraj, 2005; Swales and Feak, 2009) may classify RABs as a part-genre. However, to avoid any confusion that might occur, this study uses the word 'genre', not 'part-genre', to refer to RABs. The target disciplines are the educational and applied linguistics fields of study, although the differences between these two closely related disciplines are not the focus of the study.

The next section highlights three theoretical approaches to genre analysis in terms of their origins, their distinctive characteristics, and the target learners. Comparisons among the three approaches have also been made.

2.5 Three Approaches to Genre Analysis

Approaches to genre analysis can be divided into three perspectives: (1) English for Specific Purposes (ESP), (2) Systemic Functional Linguistics (SFL), and (3) New Rhetoric (NR). This section describes how these approaches have developed and discusses some issues related to each approach. I will argue that, although the present study is situated within the ESP approach to genre analysis, the notion of 'culture', which is a feature of both the SFL and NR approaches, should be included in the analysis.

2.5.1 English for Specific Purposes (ESP)

Despite its inception during the 1960s, the ESP approach was not widely recognized until John Swales' (1990) publication of *Genre Analysis: English in Academic and Research Settings*. Since then, the approach has been extensively adopted in teaching and research (Bawarshi and Reiff, 2010). Triggered by Swales' (1990) concepts of genre analysis, three main areas have been very much at the forefront in subsequent research: discourse/rhetorical genre theory, methods of discourse analysis, and approaches to English language teaching. Over two decades since the publication of his book on genre analysis, the topic still remains one of the key issues that have been revisited, expanded, and challenged, as demonstrated by the fact that, as of the last quarter of 2018, his book has over 14,000 citations recorded on GoogleScholar.

A number of similarities and differences between ESP and the other two approaches can be highlighted. In terms of similarities, the ESP and SFL approaches are alike in their view that linguistic features are bound up with social context and function. Both approaches take a linguistic view to deal with grammar and discourse; therefore, they concentrate more on rhetorical realization and lexico-grammatical features in a genre (Flowerdew, 2002). In addition, both are motivated by the pedagogical motivation that attempts to offer non-native students 'an explicit understanding of how target texts are structured and why they are written the way they are' (Bawarshi and Reiff, 2010, p. 43). However, the two approaches are different in terms of their target learners. Bawarshi and Reiff (2010) clarify that while the SFL approach has been targeted at disadvantaged young school learners (primary and secondary levels), the ESP approach has initially been adopted in UK and US universities for the teaching and learning of more advanced

international students who are non-native English speakers. These non-native English speakers (NNES) are considered linguistically disadvantaged learners.

In contrast to the NR approach, the ESP approach focuses on the teaching and learning of specialized varieties of English to non-native English speakers in academic and professional settings, while the NR approach places more emphasis on first language (L1) learners than second language (L2) learners, (Bawarshi and Reiff, 2010; Charles and Pecorari, 2016; Hyon, 1996). Moreover, NR places less emphasis on lexico-grammar and rhetorical patterns, but focuses more on situational context (Flowerdew, 2002, p. 91).

2.5.2 Systemic Functional Linguistics (SFL)

Systemic Functional Linguistics (SFL) was developed by Michael Halliday and his followers. The theory was influenced by Firth's 'notion of language as a set of systems' (Flowerdew, 2013, p. 10) and Malinowski's conception of cultural context (Paltridge, 2012). Distinctions between 'contexts of situation' and 'contexts of culture' have been made by Firth, Malinowski and later Halliday (van Dijk, 2009, p. 154). The first type of context is regarded as local, involving face-to-face communication within a particular setting. The latter one is regarded as global, involving the whole community and basic properties such as knowledge, norms and values. Bawarshi and Reiff (2010, p. 30) clarify that the SFL approach is 'systemic' in that a system of choices is available to speakers and writers for realizing the interactions between social actions and contexts. 'Functional' refers to the work that language does and the way to achieve it within a particular context.

Bawarshi and Reiff (2010) claim that concerns over the efficacy of student-centered, process-based literacy teaching partially gave rise to SFL. The argument from critics of that kind of teaching is that a process-based approach places an emphasis on learning-by-doing, but at the same time withdraws learners from the social process of literacy acquisition. Consequently, the absence of a systemic, patterned textual repertoire leads to the learner's inability to perform appropriately within different social contexts. A 'schematic structure' (Paltridge, 2012, p. 65) is therefore necessary to provide students with some background knowledge in organizing their texts. Charles and Pecorari (2016, p. 51) explain that in the SFL approach a detailed genre-based pedagogy emerges to serve educational needs. It consists of three main steps:

1. **Deconstruction**: models of a genre are presented and analyzed.
2. **Co-construction**: a new sample of the genre is constructed by the input from teachers and learners.
3. **Independent construction**: learners write another piece of the genre on their own.

The teaching-learning-cycle model of SFL has been criticized because of its linear pedagogy transmission, in which generic models and structures are presented to show learners how to write so that they can emulate those models (Bawarshi and Reiff, 2010). However, it can be argued that these explicit linear model serves its pedagogical purposes and the motivation behind the SFL model; that is to offer those disadvantaged students who might need more language input and observe from sample models as the very first step. Hyland (2004, p.11) put forward that the explicit teaching model enables learners to have a clear understanding of the text structures and the reason why learners construct the text in such a way. The SFL teaching model is therefore considered as a better method than those 'hit-or-miss inductive methods'.

2.5.3 New Rhetoric (NR)

The NR approach, emerging from a group of North American scholars from various disciplines, focuses on aspects of L1 teaching, such as rhetoric, composition studies, and professional writing (Hyon, 1996). The central notion of the approach lies in perceiving 'genre as social action' (Miller, 1984), with less emphasis placed upon the content and form (Charles and Pecorari, 2016). The priority is thus given to the action that is used to accomplished rather than the form of discourse. This contrasts with both the SFL and ESP approaches, in which the first priority is given to texts. Miller (1984, p. 165) argues that 'genres serve as keys to understanding how to participate in the actions of the community', and genres are 'stabilized-for-now' (Miller, 2015, p.61), which means it changes over time. Her view has important practical implications for genre-based teaching in that the teaching and learning activities should reflect what is practical as social actions, not merely designing the texts to fit formal requirements (Miller, 1984). In NR approach, genres are both socially embedded and constructed (Paltridge, 2012). Comparing to the ESP approach in which genres are understood as 'communicative tools situated within social contexts', genres in the NR approach are viewed as 'sociological concepts embodying textual and social ways of knowing, being, and interacting in

particular contexts'(Bawarshi and Reiff, 2010, p. 54). The methodology that followers of this approach have used in text analysis tends to involve ethnographic rather than linguistic methods, thereby offering rich descriptions regarding surrounding contexts and the situational actions that texts perform (Hyon, 1996).

As can be seen in the discussion above, the underlying need for the development of genre theoretical approaches was primarily related to the application of genre analysis in answering pedagogical issues in ESP classrooms. However, Bhatia (2012) highlights that in the last three or four decades, the attention has turned to the use of genre theory as a tool for textual analysis.

Taking into consideration the target of the present study (NAC in comparison with PAC), its aims (to analyze the internal organizational patterns and examine some linguistic features of abstracts), and its methods of analysis (move analysis and corpus analysis), it can be argued that this study is situated under the ESP approach. In the ESP approach, the target learners are those NNESs and the technique used in analyzing texts is move analysis in which the internal organization patterns and lexico-grammar of texts are under scrutiny.

2.6 Second Language Writing and Contrastive Rhetoric

This section is divided into three parts. Section 2.6.1 addresses the issues of second language writing and academic writing. Section 2.6.2 discusses Kaplan's model of contrastive rhetoric, which was once the central theme of research studies that focused on the comparison of writing in L1 and L2. In Section 2.6.3, the term '*culture*' is problematized and interpreted within the context of the present study.

2.6.1 Second language writing and academic writing

Academic writing is considered one of the most challenging and demanding tasks for second/foreign language learners, especially at postgraduate level, because written work is commonly used as a primary form of assessment and a degree is awarded upon the completion of a final project or a dissertation. In the words of Hyland (2009b, p. 2):

> Only through language, whether in the form of a dissertation, viva, essay assignment or unseen exam, can students consolidate and display their learning to university gatekeepers and so progress to graduation and beyond.

At present, the situation is even more demanding since many universities now encourage students to disseminate their research either through peer-reviewed journal publications or conference presentations. The dissemination of research at postgraduate level, has become one of the requirements for the award of a degree in many universities, especially at the doctoral level.

When it comes to journal publications, especially in the leading journals with high impact factors,[6] the rejection rate is usually higher than the acceptance rate. Swales and Feak (2009) point out that rejection might happen on the basis of a reading of the abstract alone. Although a well-written coherent abstract does not guarantee that a paper will be accepted, it helps in passing the manuscript to the next step of external review. Mahboob and Paltridge (2015), the editors of *TESOL Quarterly*, state that the journal had an acceptance rate of only 10.7% in 2014. However, the acceptance figures varied across different genres in the journal (general research articles, brief reports and summaries), with the highest acceptance rate for book reviews. Considering the acceptance per rejection ratio, writing a research article of publishable quality is an extremely challenging and even intimidating task for postgraduate students.

Various factors contribute to the difficulty associated with second language writing. Hu and Cao (2011) indicate that unawareness of different rhetorical conventions between L1 and L2 is a major factor in L2 writing difficulty. Other problems include the limitations of lexical and grammatical knowledge, and a lack of familiarity with specific written genres. Consequently, as Matsuda *et al.* (2003) point out, there is a growing recognition that L2 students should have a clearer understanding of the typical structure of the kind of texts that they intend to write. However, some questions remain as to which method of dealing with these issues is most efficient.

Hyland (2009a) identifies two ways of approaching writing. Historically, writing was perceived as a textual product, disembodied and context-free. Therefore, the personal experiences of readers and writers were not taken into account in the process. Good writing was therefore judged in terms of grammatical accuracy and clear composition. The second approach views texts as discourse, looking beyond their surface structures.

[6] Impact factor is the average number of citations the journal received in a certain year calculated from the number of the articles in the journal that have been cited in the two preceding years.

In this approach, writers who have a certain goal in mind attempt to write in a way that their goal will be accomplished. Meaning-making is bound up with situated contexts and social actions. The teaching-learning method puts more emphasis on the interrelationship between language forms, purposes and contexts. In this study, abstract writing is approached from the second view, regarding texts as discourse and analyzing them by investigating their rhetorical compositions and organizational patterns. With regard to novice researchers, their meaning-making through written abstracts is associated with the academic institution where they study and the guidelines for written work that their institutions provide. This can be regarded as a 'small culture'. In relation to professional writers, their meaning-making is intertwined with the guidelines of each journal. These are broadly the same in terms of the language used, the length of abstracts, and so forth, and can be conceived of as a facet of the 'professional/academic culture' in which the writers work. Further details regarding the relationships between a 'small culture' and a 'professional/academic culture' can be found in Section 2.6.3 below.

In an academic context where English is used as a medium for communication, the norms of Anglophone rhetorical patterns are apparently employed as a benchmark for evaluation. Although some students are able to produce a well-structured essay in L1, it should be noted that their writing ability might not be transferable to the L2 context. Texts written by non-native students that do not conform to English rhetorical patterns tend to receive negative feedback, with suggestions that they exhibit issues such as poor organization or problems with cohesion (Kaplan, 1966). Some contributory factors that have made the texts seem incomprehensible or incoherent to native speakers are believed to be different writing patterns stemming from culture (ibid., 1966). Indeed, a growing body of research comparing texts written by native and non-native English speakers has revealed evidence of various cultural diversities, such as move structures (Dong and Huan, 2010; Martín, 2003; Soler-Monreal *et al.*, 2011), the degree of hedge and booster usage (Hu and Cao, 2011; Hyland and Milton,1997; Pérez-Llantada, 2013), choice of active or passive construction (van Bonn and Swales, 2007), the expression of stance and voice (Hyland and Guinda, 2012), and so forth.

The issue of contrastive rhetoric deserves attention in the context of this study because the study focuses on the analysis of abstracts written by Thai novice writers and international professional writers. The writing patterns of Thai novices might be expected

to differ somehow from those international professional writers whose writing patterns are expected to conform to Anglophone publication standards. In the sections that follow, the discussion will focus on the notion of rhetorical patterns which appear to be unique to each language, based on the frequently-claimed-models of contrastive organizational patterns introduced by Robert Kaplan in 1966.

2.6.2 Kaplan's models of contrastive rhetoric

According to the American linguist Robert Kaplan (1966, p. 12), the logic underpinning rhetoric is by no means universal, and neither is the notion of rhetoric itself. The sequence of thought and rhetorical patterns of any two cultures may vary to some extent. His notion of contrastive rhetoric hypothesizes that each language possesses rhetorical patterns specific to that culture, and that L1 interference is likely to occur in second language writing, accounting for the deviation from the first language's norms. Errors that students produce in L2 writing are often claimed to be caused by cultural patterns embedded in their L1.

Kaplan's ground-breaking study on text analysis found that the essays produced by the five groups of his participants (Anglo-European, Semitic, Oriental, Romance, and Russian) differed in their organizational patterns (see Figure 2.2). For example, the Anglo-European essays were developed using a linear sequence, beginning with a topic statement followed by a series of subdivisions such as exemplifications, illustrations, and so on to support a core argument. As such, a well-organized paragraph in English should have the flow of ideas arranged in a straight line without any digression. However, it should be noted that the reverse organization, starting with details and ending with the topic statement, was also possible although not as prevalent.

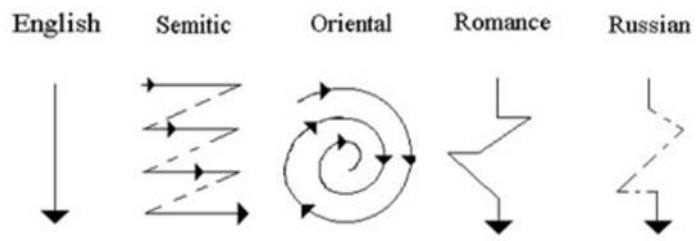

Figure 2.2 Diagram of cultural differences in paragraph organizational patterns (Kaplan, 1966, p. 21)

The typical paragraph organization in Semitic languages was found to be based upon a series of parallel patterns. English native speakers interpret writing that follows this organizational pattern as containing irrelevancy in the content, as the English language favors subordination rather than coordination (Kaplan, 1966, p. 16). In contrast with the Anglo-European and Semitic rhetorical styles, the way in which Oriental writers composed their paragraphs seemed to reflect a circular movement, called an 'indirect method'. Like the Semitic style, the Oriental style of turning around the subject was also criticized by English native speakers as awkward.

Since 1966, when Kaplan began his pioneering work on the notion of language and culture, it has become a well-recognized theoretical model, used in many research studies to explore rhetorical patterns across cultures. Despite its popularity, however, his model has also come under considerable criticism over its ethnocentrism, stereotyping and overgeneralization (W. Baker, 2013). For example, his notion of culture was influenced by the Sapir-Whorf hypothesis of linguistic relativity, which suggests that our worldview is shaped by the language that we speak and our social group. Therefore, it is notable that Kaplan's notion of culture was heavily tied with national entities.

Atkinson (2004), for example, categorizes the interrelationship between culture and nationality as reflected in Kaplan's model as '*received culture*'. It should be acknowledged that this categorization overlooks aspects of cultural complexity and heterogeneity (W. Baker, 2013). Moreover, Kaplan (1987, p. 9) also acknowledged in one of his later articles the criticism of the model as an example of 'reductionism', which simplifies the whole of linguistics to a single issue. However, it should be noted that Kaplan's model was initially created to raise teachers' awareness of various atypical

argumentative patterns embedded in different languages and manifested in the compositions of non-native language users of English.

It is widely acknowledged that language and writing are inseparable from culture (Connor, 1996). In addressing this issue in the next section, the present study will not attempt to provide a definition, but instead to explain how the notion of '*culture*' can best be interpreted.

2.6.3 Interpreting the term 'culture'

As stated in the previous section, language is closely tied up with culture (Connor, 1996). Matsuda and Atkinson (2008, p. 297, quoted in W. Baker, 2013, p. 22) criticize the fact that 'contrastive rhetoric (CR) talks about *culture* without ever telling us what *culture* is'. Indeed, many researchers focusing on contrastive rhetoric have failed to conceptualize the key term 'culture' in their studies (W. Baker, 2013; Risager, 2006; Wellein, 2008). Given that defining the term 'culture(s)' is problematic, this section will not attempt to provide any generally applicable definition of the term, but rather to focus in particular on how it should and will be categorized and interpreted specifically in the context of this study.

Different ideas have been proposed to categorize 'culture'. For example, Atkinson (2004, p. 279) divides the concept of culture into four sets of oppositions, as follows: (1) *received culture* versus *postmodern culture* versus *cultural studies culture*; (2) *culture as product* versus *culture as process*; (3) *culture in the head* versus *culture in the world*; and, (4) *big culture* and *small culture*. These four sets of oppositions demonstrate how culture can be framed in order to increase readers' awareness of when cultural phenomena are discussed. In general, the most widely used of these concepts in contrastive rhetoric is the concept of 'received culture', where notions of cultures are intertwined with nation states or ethnic communities, as in the terms American culture, Japanese culture, etc. (ibid., p. 280). In this approach to the concept, the internal complexity *within* cultures has been neglected.

While caution in interpreting the term 'culture' as 'received culture' is urged, the binary opposition between 'big culture' and 'small culture' is noteworthy here. The concrete example of small cultures in an educational setting, as depicted in Figure 2.3, owes much to the work of Holliday (1994, as cited in Atkinson, 2004). In the figure, the smallest of

the cultural spaces in an educational setting is considered to be the 'student culture', which –like any culture –will be shaped by its own particular internal norms and practices. Some of these norms and practices may overlap with (1) other small cultures that are relevant to the setting, such as a youth culture, a classroom culture, and a professional-academic culture, and (2) the big culture, here represented as the national culture.

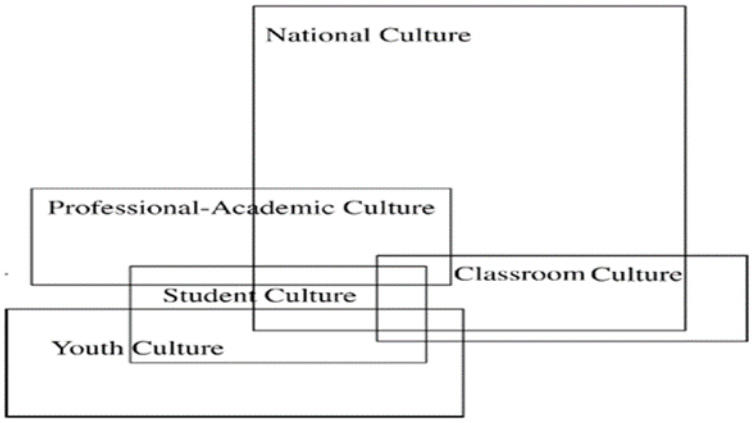

Figure 2.3 The interaction complexity of small cultures in an educational setting (Holliday, 1994 as cited in Atkinson, 2004, p. 286)

In the context of the present study, culture should be viewed as a small culture in which novice writers represent the smallest entity and professional writers represent the bigger culture. Both groups possess their own norms and practices. These norms and practices can easily be recognized by members of the group when they engage in the discourse community. Some of these may differ from those of the other group, while others may overlap. The overlapping characteristics are likely to inform the norms and practices of the bigger culture, probably the professional-academic culture in general.

As can be seen in Section 1.4, the thesis is motivated by pedagogical concerns and the findings are expected to be beneficial to those learners who aims to disseminate their research, but still need some improvements in their abstract writing. The section that follows will thus discuss the advantages of genre-based approaches and how they can be used in classroom to better prepare students in terms of abstract writing.

2.7 Genre-Based Approaches and Pedagogies

The notion of genre has been studied quite extensively, leading to a better understanding of how discourse is used in various contexts: academic, professional and institutional (Bhatia, 2014). Evidence also shows that a greater repertoire of lexical and structural knowledge of a particular genre can facilitate learners in producing texts appropriate to their discipline (Pérez-Llantada, 2014; Tessuto, 2015).

Hyland (2004, pp. 10-11; 2009a, pp. 17-18) summarizes seven advantages of genre-based instructions for writing: they are explicit, systematic, need-based, supportive, empowering, critical, and consciousness raising. Firstly, the explicitness of genre-based instruction enhances the clarity of what is to be learned, which in a way facilitates the acquisition of student writing skills. Secondly, it is systematic because it involves a coherent framework concentrating not only on language but also its contexts. Thirdly, course objectives and content are designed according to target needs. Fourthly, with genre-based instruction, students' learning and creativity are supported by their teachers. The last three benefits seem to be interrelated in that by providing exemplary patterns with possible variations, students are firstly empowered. When they have learned from those available resources, teachers can give their students an opportunity to critique and challenge those patterns. The samples of texts available as resources can raise the consciousness of both teachers and students regarding the variety of writing patterns that can be used in a particular genre.

Despite the theoretical, methodological, and pedagogical contributions that genre-based approaches offer, debate continues over certain issues, such as 'genre integrity' (Bhatia, 2014, p. 142) or the extent to which practitioners should follow the conventionalized structures of a specific genre, both in terms of rhetorical moves and lexical choices.

Previous research has confirmed that different disciplines do have their own preferences for the use of some lexical choices and organizational patterns (see the summary in Hyland, 1994, 2000). However, text analysts (e.g., Bhatia, 2014; Guinda, 2015) claim that an increasing number of hybrid genres with mixed communicative purposes are found in professional contexts outside the classroom, such as promotional discourses in academic introductions, abstracts and the integration of persuasive and informative features in advertising. This raises a pedagogical concern as to whether one specific genre

appropriate to a learner's discipline should be introduced or explicitly taught in class, or whether a variety of genres should be covered. In contrast, other researchers (e.g., Hyland, 2007; Saeeaw and Tangkiengsirisin, 2014) claim that raising students' awareness of salient recurring discourses and the norms of rhetorical organization of particular genre(s) is important, especially for novices.

Another pedagogical concern is related to criticisms of the focus on genre rather than discipline variations, and similarity rather than difference (Hyland, 2000). This situation can typically be found in the teaching of English for academic purposes, where the course is carried out by teaching students generic skills, and the reproduction of these skills is deemed necessary. Texts are thus perceived as 'objective, rational, and impersonal' (ibid., p. 4). It is worth noting that despite the importance of the genre-based approach to teaching and learning, the emphasis on disciplinary variations should also be acknowledged, depending on the availability of time and resources.

2.8 Move Analysis and the ESP Approach

The ESP approach to genre analysis, which focuses on the analysis of text organizational patterns, has given rise to a text analytical approach called move analysis. The concept of move analysis usually involves the units of analysis called 'move' and 'step'. Both terms are generally used to explain the mechanisms and organizational patterns found in academic genres (Kanoksilapatham, 2009) that writers use to convey their social purposes (Hyland, 2012b). While 'move' can be defined as 'a rhetorical or discoursal unit in a text that performs a coherent and distinctive communicative function' (Paltridge, 2012, p. 244), 'step' is considered as 'submoves' or 'strategies' used to achieve a particular communicative purpose. For example, move 1 of the research abstracts (Introduction) involves different steps or submoves, such as providing background information about the topic, identifying research gaps, and indicating significance of the study. The study, however, focuses only on the identification and analysis of moves, not on steps because the division of moves into steps would lessen the number of samples in each category and make the tagging process more complicated.

In addition, Connor *et al.* (2007, p. 24) claim that moves generally have distinct linguistic boundaries that can be objectively analyzed. The idea of the text being made up of separate and identifiable moves can be used to categorize the main communicative

purposes of the text – i.e. there are different kinds of moves, and within different moves, different sub-components, or steps, can be identified, each with their own purpose. It is important to note here that the aim of this study focuses attention only on the macro-structure or moves, not on the micro-structure or steps. Various researchers have attempted to create their own model of move analysis for abstracts (e.g., Bhatia, 1993; Hyland, 2000; Swales, 1990) in the hope that an established framework can be used in learning, teaching, and research.

2.8.1 Swales' IMRD framework

Swales (1990, p. 52) claims that 'established members of discourse communities employ genres to realize communicatively the goals of their communities'. He also stresses that there is a tendency that only established members of the discourse community are fully aware of the shared objectives of a particular genre; apprentice members might be partially aware of them; and non-members might be either aware or unaware of them (ibid., p. 53). In light of this, training in abstract writing should be offered to student writers with the aim of providing them with the repertoire of relevant rhetorical moves and lexical choices, as part of the process of preparing them for professional academia. The training can be regarded as a way to 'socialize students into the expectations of their particular academic discourse communities' (Northcott, 2013, p. 4).

The IMRD model consists of four moves: (1) Introduction, (2) Method, (3) Results, and (4) Discussion. It was originally used to examine research articles in their entirety, not just abstracts. However, it has also been used for abstract analysis by some researchers (e.g., Lorés, 2004; Supatranont, 2012) and for introduction analysis by Samraj (2005). As an abstract is viewed as a summary of the whole RA, it should not be problematic to apply Swales' IMRD model for analysis. The applicability of the model for analyzing other individual parts of the RAs, such as the Introduction, is arguably rather dubious because we would expect the rhetorical organization of those parts to be somewhat different.

2.8.2 Bhatia's framework

A somewhat similar model to Swales' framework was subsequently created by Bhatia (1993). In Bhatia's model, the Introduction Move was replaced by a Purpose Move, and the Discussion Move was substituted by a Conclusion Move. Therefore, this version of

the four-move model consists of Purpose, Method, Results, and Conclusion Moves. In suggesting that the first move should be classified as the Purpose Move, it seems that Bhatia was attempting to fill a gap in Swales' framework, since the features of this move are regarded as a priority in some disciplines. The opening of the text is not considered simply as lead-in information or background for the study, but is seen as having an important role, just like the other moves. However, it is worth noting that 'introduction' is a broader term that covers the purpose of the study, especially when used with the whole research paper. On the other hand, the use of the term 'purpose' is more limited, and it is arguably seen as focusing more on the objective of the study than on the context or the rationale that motivates the research. With respect to the change from 'discussion' to 'conclusion', this is probably caused by the fact that the word limitations of an abstract generally prevent writers from engaging in lengthy discussion. As such, 'conclusion' is considered a more appropriate term in the context of abstract writing.

2.8.3 Hyland's analytical model

A five-move analysis for abstracts was subsequently proposed by Hyland (2000). Unlike Bhatia's model, which introduces the Purpose Move in place of the Introduction Move, Hyland's framework consists of Introduction, Purpose, Methods, Product, and Conclusion Moves. The framework is applicable to abstracts from various disciplines, rather than specific disciplines. Disciplinary differences can be reflected in whether the Introduction or Purpose Moves occur in the abstracts. For example, in Physics and Biology the percentage ratio of Introduction per Purpose was found to be 13:77 and 13:57 respectively. However, in Philosophy both moves are found with a similar proportion (46:50). It is with this applicability to a broader range of disciplines in mind that the present study will employ Hyland's framework for RAB analysis.

Hyland (2000, p. 67) indicates that the Introduction Move and the Purpose Move were distinguished from each other in his study because they perform different roles. While the Introduction Move describes the context of the paper and explains the motivation behind the research, the Purpose Move points out its aims and assumptions and outlines the intention behind the paper (see Table 4.2 in Chapter 4 for a description of Hyland's model). In creating his framework for move analysis, however, Hyland did not explain why the term 'Product' was adopted instead of 'Result'. He clearly states in the objective of his study that his intention was not to 'provide definitive description of the move

structure of features of this genre', but to provide an account of abstract writing focusing on how the writers negotiate the significance of their research (ibid., p. 65).

2.9. Previous Studies Related to Move Analysis and Research Abstracts

Before discussing in detail previous studies related to the use of move analysis on research abstracts, the following section points out the importance of abstracts as an academic genre.

2.9.1 The importance of abstracts as an academic genre

The importance of abstracts as an academic genre has been reported in the literature. For Hyland (2000, p. 63):

> Abstracts are worthy of study because they are significant carriers of a discipline's epistemological and social assumptions, and therefore a rich source of interactional features that allow us to see how individuals work to position themselves within their communities.

There seem to be different views among researchers regarding the main communicative purpose of RA abstracts, as reflected in the different ways in which they classify and subcategorize them. For example, in Hyland's (2000) study, abstracts are viewed as a promotional discourse in which researchers promote their professional credibility by engaging in the topics they have written on as insiders. Such promotion can be achieved through 'marketization of oneself and one's paper through discursive means' (ibid., p. 63). It is worth noting that among the five moves, the promotional function is likely to be found in m1 (Introduction) or m4 (Purpose). Swales and Feak (2012, p. 331) claims that the aims of the introduction are to strive for readers and research space, which is somewhat similar to abstracts. In addition to m1, m4 or Product Move can also be used to emphasize the main argument, which are found to be essential in a promotional genre (Hyland, 2000).

Lorés (2004) points out that written abstracts can be identified as indicative and informative, while Tankó (2017, p. 43) mentioned another type which is indicative-informative (a mix of both). The indicative abstract focuses on the subject and the main findings without having a detailed account of the methodology section. The informative

abstract acts like a summary of the RA. The indicative-informative type has the characteristics of the descriptive type, but with the addition of conclusions. Sala (2015), on the other hand, distinguishes between informative and indicative abstracts, indicating that the second type excludes the results and focuses instead on the scope and the purpose of the study.

2.9.2 Previous studies using move analysis to analyze RA abstracts

The existing literature regarding the move analysis of RA abstracts has confirmed disciplinary differences in rhetorical patterns (Hyland, 2000; Kanoksilapatham, 2009; Samraj, 2005), in move compositions (Hyland, 2000; Saeeaw and Tangkiengsirisin, 2014), certain linguistic features (Hu and Cao, 2011; Salager-Meyer, 1992) and in lexical choices (Supatranont, 2012; Saeeaw and Tangkiengsirisin, 2014). The communicative purposes of abstracts have also been examined in some studies (e.g., Hyland, 2000; Ren and Li, 2011). However, despite the fact that previous studies acknowledge the existence of embedded moves, few of them have investigated this aspect in detail.

Hyland (2000)'s study examined 800 abstracts from the 1997 issues of 10 journals across eight different disciplines ranging from humanities to sciences. In analyzing these abstracts, the five-move model - Introduction-Purpose-Method-Product-Conclusion - was developed. With respect to move compositions, the results reveal that a Product Move was found in almost 95% of all the abstracts, a Purpose Move was found in 81%, a Method Move was found in 49%, and a Conclusion Move was found in 22%. Less than 5% of the abstracts contained all the moves. An explicit Introduction Move was absent in a high proportion in some disciplines (as discussed in Section 2.8.3), with the overall of approximately 55% without this move. Abstracts with only two-move structures, such as Purpose-Product and Introduction-Product, were also found. With regard to move structures, the most frequent structures in the corpus (25%) were Purpose-Method-Product, followed by Intro-Purpose-Product (15%). It was also found that longer abstracts, which contained the repetition of some moves, emphasized a series of results as outcomes of different aims and methods. This repetition (recycled) moves are often found in the sciences.

Ren and Li (2011) compared the rhetorical moves of abstracts written by Master's students and professional writers in applied linguistics. The corpus consists of 25

published RA abstracts from five leading journals in 2007 and 25 thesis abstracts from the same year. Hyland's (2000) framework was initially used to analyze the data, but subsequently three moves were added to the analysis - Structure, Promotion and Limitation. In terms of length, the RA abstracts were more concise, with an average of 199 words, while the thesis abstracts were lengthy, averaging 434 words. The most frequently found move patterns in the RA abstracts were Purpose-Method-Product (32%), followed by Introduction-Purpose-Method-Product (28%) and Introduction-Purpose-Product (28%). The most prevalent move pattern in the thesis abstracts was a six-move-pattern, with an addition of a move termed 'structure'. Ren and Li conclude that experts seem to be more confident in the choices of move they select to serve the goal of their writing, whereas novices tend to be cautious by including all the moves. In addition, expert-written abstracts show a more persuasive role than an informative role, but student-written abstracts reveal the opposite trend.

Saeeaw and Tangkiengsirisin (2014) examined the variation in RA abstracts from two different fields (Environmental Sciences and Applied Linguistics). Two hundred empirical RA abstracts (100 from each field) were selected from the four leading journals in each discipline. Hyland's (2000) model was employed for the analysis. Criteria were adopted to distinguish moves as obligatory, conventional and optional. The Introduction Move was found to be optional in both disciplines. Tenses used in this move were found to be present simple and present perfect. A Purpose Move was conventional in both fields, with the distinctive linguistic features being the use of reporting verbs preceded by determiners and nouns. Some embedded moves (the embedding of an Introduction with a Purpose Move and a Purpose Move with a Method Move) were noticeable. However, it should be noted here that the study did not provide a detailed account of the percentage of the occurrences of these embedded moves. A Product Move was found to be obligatory (100%) in Environmental Sciences abstracts but merely conventional (95%) in Applied Linguistics. The formulaic pattern of a *that*-complement clause was dominant in this move in both fields. A Conclusion Move was conventional in both disciplines, with an occurrence of approximately 70% in each field. The most frequently found move pattern in Environmental Sciences was Introduction-Purpose-Method-Product-Conclusion (23%), while the most commonly found pattern in Applied Linguistics was Purpose-Method-Product-Conclusion (20%).

In her study of the generic structure of RA abstracts in four academic disciplines (Biochemistry, Microbiology, Civil Engineering, and Software Engineering), Kanoksilapatham (2009) drew data randomly from larger corpora compiled at different points in time (in 2003, 2007 and 2008). The findings confirmed that a generic structure of five rhetorical moves was found. Although different terms were used to describe each move (1. Background Information, 2. Purposive Statement, 3. Methodological Description, 4. Result Announcement, and 5. Discussion, Conclusion and Implication), they closely correspond to Hyland's five-move-model. In Kanoksilapatham's study, Result Move and Discussion Move are found to be prevalent across all four disciplines. She also claims that when the Discussion Move is present, it normally functions as the conclusion of an abstract.

It is worth noting that comparisons between the findings of previous studies need to be made with care when different analytical frameworks have been used. The discrepancies in move classifications with regard to models that separate the Introduction and Purpose Moves and models that combine them as one might affect the proportion reported for all moves.

2.10 Summary

This chapter has addressed a number of prominent issues regarding genre analysis, beginning with definitions of the term *genre*. It has also pointed out a number of important features of *genre*, *register*, and *discipline*, terms which are interrelated and sometimes overlapping. While a whole text can be analyzed from either the genre or register perspectives, partial sections of texts can be analyzed only from register perspectives. This is due to the fact that the register perspective focuses on the pervasive and conventional linguistic characteristics that may occur in a variety of texts, but the genre perspective focuses on the organizational patterns of the text (Biber and Conrad, 2009). The present study is thus situated somewhere along the continuum of both the genre and register perspectives. Although the primary focus of this study is on move compositions, rhetorical patterns and linguistic features, a secondary aim focuses on the lexical choices found in written abstracts.

Three approaches to genre theory (ESP, SFL and NR), were briefly discussed in terms of their similarities and differences. Move analysis, which is associated with the ESP

approach, will be used for analyzing the rhetorical patterns of abstracts in this study. The conceptualization of the discourse community was then explored in relation to its membership and their levels of involvement. Turning to the relationship between second language writing and academic writing, some concepts and issues were illustrated and discussed, specifically Kaplan's model and the term 'culture', which tends to be overgeneralized and/or equated with the notion of national entities. The more specific notions of 'small culture' versus 'big culture' have been adopted to analyze the phenomena occurring in the present study, as opposed to using the idea of 'received culture', which is bound up with national entities. This chapter also looked at how genre studies have played an important role in terms of learning and teaching ESP/EAP. Then, the development of three move analysis models employed in genre analysis and ESP was explored, leading to an explanation of why Hyland's analytical framework was selected for the analysis in this research project. In the final part of the chapter, relevant studies related to the move analysis of abstract writing were discussed, demonstrating how Hyland's model was employed in each of them, sometimes with modifications.

Chapter 3. Issues in Corpus Analysis

3.1 Introduction

The chapter examines literature related to issues in corpus linguistics. Section 3.2 briefly discusses the distinction between general and specialized corpora, corpus-based and corpus-driven approaches, and keywords and key keywords. Section 3.3 starts with the definition of lexical bundles (LBs) and discusses the two approaches to collocations; that is, phraseological and frequency-based approaches. Section 3.4 explores previous studies in relation to LBs in academic writing.

In this chapter, I will argue that adopting a pure English for Specific Purposes (ESP) approach to move analysis, though useful in identifying the organizational patterns of abstracts, is not adequate for textual analysis. The combination of the move analysis and the corpus analysis offers a more complete picture by providing empirical evidence in that (1) the texts are authentic and (2) the generalization with regard to lexical choices and other linguistic features of the two corpora can be achieved by observing the repeated patterns from concordance lines. Apart from being empirical, a large collection of texts, known as corpus, together with the multi-techniques of corpus software, enable researchers to explore lots of evidence in linguistic patterns and usage at once.

3.2 Issues in Corpus Analysis

In this section, the distinctions between general and specialized corpora are firstly discussed. Then, the rationale why the present study employed specialized corpora to compare the move patterns and lexical choices of the novice abstract corpus (NAC) and the professional abstract corpus (PAC) is presented. Next, the section discusses the dichotomy between corpus-based and corpus-driven approaches. Following this, the differences between keywords and key keywords are discussed.

3.2.1 General and specialized corpora

This section describes a number of distinctions between general and specialized corpora and the argument why it is necessary to use specialized corpora in the present study. Distinctions between the two can be made in terms of their coverage. General corpora are generally a large collection of texts, intended to be representative and balanced for a

language as a whole, whereas specialized corpora contain a smaller amount of data, restricted to a specific variety, register or genre (Gries, 2009). Well-known general corpora include the British National Corpus (BNC), which contains 100 million words of both spoken and written language, Corpus of Contemporary American English (COCA), which consists of more than 560 million words divided according to different registers, and the British Academic Written English Corpus (BAWE), which is a collection of approximately 3000 academic assignments containing 6,506,995 words. A specialized corpus, in contrast, aims to be used as a representative of a specific text type and used for investigating a specific type of language (Hunston, 2002). Although Hunston (2002, p. 14) states that 'there is no limit to the degree of specialization', the parameters such as, a time frame, a text type, and some social settings can be set before the compilation.

Arguably, there are some reasons why large-scale corpora do not serve the aims of the present study. Firstly, those large-scale corpora contain a more variety of texts in the collection that are beyond the scope of the study. Secondly, because analyzing the move patterns and lexis used in academic abstracts is the main aim of the study, those corpora are not considered as a representative of this specific genre. Although BAWE might be considered as a more relevant corpus in comparison to the BNC or the COCA, the data of students' assignments still represent a different genre. Therefore, there is a need to compile specialized corpora in order that the comparisons between abstracts written by novice writers and professional writers can be made. With the compilation of specialized corpora, a specific genre, academic research abstracts (RABs), can be analyzed. Additionally, the qualitative analysis of move compositions and patterns can also be achieved through a more manageable size of data. Moreover, the move patterns and lexical choices used by Thai novice writers in comparison with those international scholars can be acknowledged in order that the results can be used for pedagogical purposes. In this study, the two specialized corpora are known as the novice abstract corpus (NAC) and the professional abstract corpus (PAC). The details of these corpora can be found in Section 4.6.

3.2.2 Corpus-based and corpus-driven approaches

The distinction between corpus-based and corpus-driven approaches has been discussed in the literature. The differentiation was first initiated by Tognini-Bonelli (2001, p.65),

suggesting that a 'corpus-based' approach can be used as a general term to refer to 'all types of work that relate to or draw on a corpus'. In fact, the distinction can be based on the methodology, the term 'corpus-based' refers to the use of corpus data to explain, verify, or exemplify a theory or hypothesis. This means that the data stored in the corpus is used to confirm the pre-conceived theory or hypothesis by allowing the linguistic features to be explored and quantified as evidence. In contrast, in a corpus-driven approach, the data in the corpus is considered as more than a repository for theory or hypothesis testing, but rather for theory or hypothesis creation. In other words, the corpus-based approach begins with a pre-determined linguistic theory or an assumption in relation to linguistic forms or structures, and uses the theory or assumption to analyze them, while the corpus-driven approach explores the linguistic features and constructs that emerge from the data without pre-determined assumptions (Biber, 2012). Thus, the corpus-driven approach can be viewed as inductive knowledge creation, while the corpus-based approach can be seen as deductive knowledge one. McCarthy (2001, p. 129) informally states that 'one can assume one knows it all already and just go to a corpus to find texts that demonstrate the known facts', which is regarded as the corpus-based approach. In contrast, 'one can …go with a completely open mind to a corpus, willing to be guided, illuminated by it in ways one could not dream of' (ibid.), which is known as corpus-driven approach. It can also be argued that in comparison with corpus-based studies, researchers using the corpus-driven approach rely heavily on a large amount of data to study linguistic features. By using a large amount of data, the existence of linguistic constructs that are not described by linguistic theories can be documented with evidence (Biber, 2012). The corpus-based approach can also be referred to as a top-down approach where the analysis starts with the framework or theory, which is opposite to a bottom-up approach (Biber *et al.*, 2007; Upton and Cohen, 2009). It should be noted that despite the differences between the two approaches, some studies use the term 'corpus-based' to refer to studies that employ corpus techniques for data analysis. This means that any studies having a collection of data for analysis can be referred to as corpus-based studies.

In the present study, the internal organization of texts or linguistic structures beyond the sentence (move analysis) is the primary focus of the analysis. Numerous previous studies analyzed texts from this perspective have been carried out using qualitative approach, with detailed analyses of a small number of texts (e.g., Ren and Li, 2011; Salager-Meyer,

1992; Samraj, 2005). The synergy of a corpus approach and a move analysis of internal text structure is regarded as a challenge in corpus linguistics (Biber *et al.*, 2007). However, an increasing number of studies (e.g., Chen and Baker, 2010; Hu and Cao, 2011; Hyland, 2000, 2008a, 2008b, 2012a; Saeeaw and Tangkiengsirisin, 2014) have combined the two approaches together, which help increase the generalizability of results to a certain extent. In categorizing texts into moves according to their communicative purposes, corpus techniques such as keywords and cluster/n-grams functions can be used to analyze the lexical choices that researchers employ to achieve those purposes. In addition, these techniques can be adopted to investigate distributional differences in relation to language varieties and choices, 'based on the premises that language variation is functional: that we choose to use particular linguistic features because those forms fit the communicative context of the text' (Biber, 2012, p.3).

Biber *et al.* (2007) describe the procedures of corpus-based (see Table 3.1) and corpus-driven approaches in detail. Table 3.1 shows how corpus-based analyses of discourse structure can be performed. Examining the procedures in Table 3.1 more closely, it can be argued that step 2 (segmentation) and step 3 (classification) are occurring at the same time, not in a sequential order as described in the table because the segmentation cannot be achieved if the function of communicative purposes in each move is not realized.

The corpus-based approach is considered as a top-down process, where the analysis starts with the development of analytical framework. In some cases, the available framework is adopted instead of the self-developed framework. In the present study, Hyland's (2000) five-move analytical framework for abstracts was adopted before dividing the sentences/clauses/phrases into moves (segmentation).

McEnery and Gabrielatos (2006, p. 35) point out that one of the important features of the corpus-based approach is an annotation or markup determined by a researcher to test the theory. In the present study, corpus annotation, in particular move tagging, was performed by determining its communitive purposes or moves. The advantage of annotated corpora over unannotated ones is that the implicit information has been made explicit through annotation (McEnery and Wilson, 1996). Tagging helps the researcher to explore lexical choices used specifically in each move.

Table 3.1 Corpus-based analyses of discourse organization (Biber *et al.*, 2007, p. 13)

Required step in the analysis	Realization in this approach
1. Communicative/Functional Categories	Develop the analytical framework: determine set of possible functional types of discourse units, that is, the major communicative functions that discourse units can serve in corpus
2. Segmentation	Segment each text into discourse units (applying the analytical framework from Step 1)
3. Classification	Identify the functional type of each discourse unit in each text of the corpus (applying the analytical framework from Step 1)
4. Linguistic analysis of each unit	Analyze the lexical/grammatical characteristics of each discourse unit in each text of the corpus
5. Linguistic description of discourse categories	Describe the typical linguistic characteristics of each functional category, based on analysis of all discourse units of a particular functional type in the corpus
6. Text structure	Analyze complete texts as sequences of discourse units shifting among the different functional types
7. Discourse organizational tendencies	Describe the general patterns of discourse organization across all texts in the corpus

Table 3.2 Corpus-driven analyses of discourse organization (Biber *et al.*, 2007, p. 14)

Required step in the analysis	Realization in this approach
1. Segmentation	Segment each text in the corpus into discourse units, based on shifts in vocabulary or other linguistic features
2. Linguistic analysis of each unit	Analyze the full range of lexical / grammatical characteristics of each discourse unit in each text of the corpus
3. Classification	Identify the set of discourse units types that emerge from the corpus analysis, based on linguistic criteria; that is, group all discourse units in the corpus into linguistically-defined categories or 'types'
4. Linguistic description of discourse categories	Describe the typical linguistic characteristics of each discourse category, based on analysis of all discourse units of a particular type in the corpus
5. Communicative/functional categories	Describe the functional bases of each discourse category, based on post-hoc analysis of the discourse units identified as belonging to a particular type
6. Text structure	Analyze complete texts as sequences of discourse units shifting among the different functional types
7. Discourse organizational tendencies	Describe the general patterns of discourse organization across all texts in the corpus

3.2.3 Keywords and key keywords

In this section, the terms *keywords* and *key keywords* will be differentiated and discussed. Keywords are neither determined by only the frequencies they occur (Evison, 2010) nor based on the subjective view of importance (Baker, 2004). However, keywords are those words that are statistically frequent (positive keywords) or infrequent (negative keywords) in a target corpus in comparison to a reference corpus (Flowerdew and Forest, 2009; McEnery *et al.*, 2006). A test of keyness is usually based on either a log-likelihood (log-L) test or a chi-squared test to determine a p-value set by a user. With regard to key keywords, Scott (1997) describes them as those frequent words occurring in several texts, rather than those occurring in a single or a few texts in a corpus. According to Scott, three types of keywords are usually identified: (1) proper nouns; (2) 'aboutness' words -- keywords that describe the characteristics of a particular text; and, (3) high-frequency words signifying the author's style.

Although proper nouns provide general characteristics of what the corpus focuses on, these words might not serve pedagogical purposes. In addition, the problem of keyword retrieval occurs when some words are absent or have zero occurrence in a target corpus, while having a high frequency in a reference corpus. These keywords usually appear at the top of the keyword list. The interpretation of these keywords should be carefully considered because the absence of these keywords denotes a speaker or writer's limited linguistic knowledge or his/her choice of production (Gablasova *et al.*, 2017, p. 147). It can be argued that the issue of zero occurrence is likely to occur if the texts from different cultures are compared.

3.3 What are Lexical Bundles (LBs)?

Research studies have used various terms to refer to multi-word expressions. For example, they are referred to as *formulaic language* (Schmitt & Carter, 2004; Wray, 2000; 2002), *formulaic sequences* (Simpson-Vlach and Ellis, 2010), *multi-word constructions* (Liu, 2012), *lexical bundles* (e.g. Biber and Barbieri, 2007; Cortes, 2013; Hyland, 2008a), and *clusters* (Hyland, 2008b).

Wray (2000, p.465) compiled a list of more than 50 terms that can be used to describe a smaller or larger set of strings found in the literature (see Figure 3.1). It is apparent that some of these words do not mean exactly the same. In Wray's idea, the neutral term that

can be used without imposing any theoretical positon is ‘formulaic language’ (Wray, 2002, p. 8). Wray (2000, p. 465) defines the term as:

> A sequence, continuous or discontinuous, of words or other meaning elements, which is, or appears to be, prefabricated: that is, stored and retrieved whole from memory at the time of use, rather than being subject to generation or analysis by the language grammar.

amalgams	gambits	preassembled speech
		prefabricated routines and patterns
automatic	gestalt	ready-made expressions
chunks	holistic	ready-made utterances
clichés	holophrases	recurring utterances
co-ordinate constructions	idiomatic	rote
collocations	idioms	routine formulae
composites	irregular	schemata
conventionalized forms	lexical(ized) phrases	semi-preconstructed phrases that constitute single choices
F[ixed] E[xpressions] including I[dioms]	lexicalized sentence stems	sentence builders
fixed expressions	multiword units	stable and familiar expressions with specialized subsenses
formulaic language	non-compositional	stereotyped phrases
formulaic speech	non-computational	stereotypes
formulas/formulae	non-productive	stock utterances
fossilized forms	non-propositional	synthetic unanalysed chunks of speech
frozen metaphors	petrifications	
frozen phrases	praxons	

Figure 3.1 Terms for aspects of formulaicity in literature (Wray, 2000, p. 465)

Still, the term *lexical bundles* firstly introduced by Biber *et al.* (1999), which is used in the study, has not been on Wray’s (2000) list. Biber *et al.* (1999, p. 990) define the term as ‘recurrent expressions, regardless of their idiomaticity, and regardless of their structural status’. These recurring sequences of words commonly co-occur in a natural language use. Hyland (2008b, p. 41) defines LBs as ‘extended collocations which appear more frequently than expected by chance, helping to shape meanings and contributing to our sense of coherence in a text’. According to Biber *et al.* (1999), the most frequent bundles are three-word bundles, while the longer bundles (four- to five-word bundles) are more phrasal and less common. The bundles must frequently recur in a given register at a minimum frequency. However, it should be noted that there is no agreed minimum LB threshold; on the contrary, it is totally arbitrary and varies from study to study.

Cortes (2013) explains that there are two approaches to identify LBs. First, an intuition-based method, in which the lists of expressions are made up by the researcher's perception of what string of words occurs frequently in the language. With the corpus software, the empirically-based research method can be used to identify these multi-word constructions in a particular register.

Having a large repertoire of LBs does not guarantee that speakers and writers can employ them in any discourse functions. Different studies led by Douglas Biber (e.g., Biber, 2012; Biber and Barbieri, 2007) argue that different registers possess their own set of linguistic patterns that can be used to convey discourse functions, associated with communicative purposes of a specific register. For example, a greater number of lexical bundles are found in conversation than in academic writing, and the result is even more noticeable with the longer bundles (Biber *et al*., 1999). Therefore, the English for Specific Purposes (ESP) approach, which considers the importance of different registers in teaching and learning, is necessary to describe LBs usage in abstract writing.

3.4 Previous Studies on LBs in Academic Writing

A large body of research have been conducted on LBs in academic writing (e.g., Chen and Baker, 2010; Hyland, 2008a, 2008b, 2012b; Liu, 2012). While a number of studies place an emphasis on native or non-native learners (e.g. Wei and Lei, 2011; Ädel and Erman, 2012), numerous studies focus on different genres (e.g., El-Dakhs, 2018; Hyland, 2008b; 2012a; Sánchez, 2014) and different registers (Biber and Barbieri, 2007; Biber *et al.*, 2004).

Native speakers, of course, have no problems in processing the formulaic language (Wray, 2002). Language fluency can also be determined by how a skillful one can use these multiword constructions (Hyland, 2012a, p. 150). Some studies (Ädel and Erman, 2012; Wei and Lei, 2011) found that native speakers possess a greater variety of LB types. Although Cortes (2004) argues that a competent language use can be determined by the frequent use of LBs within a register, Hyland (2008b) found that less proficient learners often rely more on the usage of LBs. The result of Hyland's study is somewhat similar to Wray's (2002) observation in that a heavy reliance on formulaic language was found in learners during the early stages of both first and second language acquisition. Qin (2014) also states that writers' incompetency or lack of expertise in an academic

context may be shown by the absence of formulaic language usage. Therefore, these inconclusive observations and findings raise a further question in relation to which group of researchers (novice or professional) rely more heavily on the LB usage in this study.

Biber *et al.* (2004) develop a functional framework for describing LBs in discourse, which consists of three main categories: (1) stance expressions, (2) discourse organizers, and (3) referential expressions. Stance expressions include those impersonal attitudinal/modality stance bundles such as *it is necessary to* and *it is important to* and those epistemic stance markers such as *are likely to* and *the fact that the.* Discourse organizers include topic introduction bundles, such *as in the next section* and *the rest of the paper*, and topic elaboration bundles, such as *on the other hand* and *as well as the*.

Hyland's (2008b) study was geared towards different genres of academic discourse. Instead of examining two-word-collocations, his focus was on extended collocations, or 'clusters'. The aim of Hyland's study was to explore the degree to which 4-word clusters are vital to academic discourse and how their use differs across three different genres. His 3.5-million-word corpus consisted of research articles, doctoral dissertations and master's theses from four disciplines, namely electrical engineering, business studies, applied linguistics and microbiology. The research articles were selected from the leading journals in the field. All of the PhD and the Master's theses were written by Cantonese L1 students at five universities in Hong Kong. It should be noted that the similarities and differences among these genres can shed some light on the performance of amateurs, experts-to-be, and experts, providing us with valuable pedagogical insights.

Hyland first generated a wordlist for each genre and then the cluster lists by using the cut-off point of 20 times per million words and the range of 10% of text recurrence in the sample were set. Concordance lines were then generated for qualitative analysis; that is to examine the textual contexts of the clusters and identify their structures and functions. Apart from identifying the textual patterns, Hyland also classified the clusters into three broad functional types: (1) research-oriented clusters which include those indicating location (*at the beginning of, in the present study*), procedure (*the use of the, the role of the*), quantification (*a wide range of, one of the most*), description (*the structure of the, the size of the*); (2) text-oriented clusters which consist of transitional links (*on the other hand, in contrast to the*), resultative links (*as a result of, it was found that*), structuring signals (*in the present study, as shown in fig.*), and framing signals (*in*

the case of, with respect to the); and, (3) participant-oriented clusters which include stance indication (*are likely to be, may be due to*) and engagement indication (*it should be noted, as can be seen*) (see Hyland, 2008b, p. 49 for more details).

Overall, 130 different 4-word clusters from the whole corpus were found based on the previously mentioned criteria. The three most prevalent clusters found in the corpus were *on the other hand*, *at the same time* and *in the case of*. When retrieving the 4-word clusters in each corpus, the normalized figures of clusters found in Master's dissertations, PhD theses, and research articles constituted 5.1% (149 clusters), 3.8% (95 clusters) and 3.1% (71 clusters) respectively. In other words, many clusters found in the Master's or the PhD texts did not appear in the research articles. Hyland, therefore, suggests that these apprentice writers rely more on formulaic patterns or clusters in developing their arguments when compared with the PhD students and the expert writers. Among the eight structural patterns (noun phrase + *of*, noun phrase with other post-modification, prepositional phrase + *of*, other prepositional phrases, passive + prepositional phrase fragment, anticipatory *it* + verb/adjective, verb (be) + noun phrase, and others), prepositional phrases and noun phrases with *of* fragments were found to be the most common clusters; they are also common in the top 50 clusters in the list. Further classification was performed with the top 50 clusters.

With regard to **structural differences**, nearly 50% of the clusters in the doctoral theses were prepositional phrases with embedded *of-phrase* fragments such as *the use of the, in the case of,* and *at the end of*. Similarly, over half of all the clusters in the expert texts were *of-phrase* structures. Although the Master's students employed the noun phrase with *of-phrase* fragment the most, they also frequently used passive structures, such as *is based on the, is shown in figure*. Interestingly, the two corpora of student genres found far more examples of the *anticipatory-it* pattern, suggesting that the students tried to restrain their authorial stance. As for **functional differences** of the clusters among these three genres, the expert writers and the doctoral students used text-oriented clusters the most. However, the Master's students tended to use research-oriented clusters with a slightly higher frequency than text-oriented ones. The least frequently found clusters seem to be participant-oriented clusters, which requires writers to engage readers and express the writers' authorial stance. The study concludes that there were considerable

variations of the clusters employed by the three groups in terms of forms, structures and functions.

The disciplinary variation which was absent in Hyland (2008b) is discussed in relation to the issue of LBs in Hyland (2012a). It is important to note here that LBs in Hyland (2012a) are referred to as 'clusters' in Hyland (2008b). As pointed out by Hyland, parameters such as the threshold frequency, the number of distributions, and the length of strings should be taken into consideration before identifying LBs. Among the four disciplines (electrical engineering, business studies, applied linguistics, and biology), bundles found in electrical engineering were of greatest range and very specific to the discipline. Although some LBs, such as *on the other hand, as well as the,* are found to be shared across disciplines, many bundles are still discipline-specific. Another disciplinary difference lies in the distribution of lexical bundle types (see Hyland 2008b). Research-oriented bundles were more likely to be found in science/technology texts, whereas participant-oriented ones appeared more often in social science texts.

Chen and Baker (2010) investigated LBs in three different corpora: (1) published academic texts; (2) native student texts; and, (3) non-native student texts. The threshold used in retrieving four-word bundles in their study was set to a minimum of 4 occurrence which are equal to 25 times pmw. After the retrieval, a number of overlapping bundles with regard to a complete overlap and a complete subsumption were combined into a longer unit to avoid inflated results (see Section 7.2.4 for detailed discussion). The comparisons between structural and functional LBs among the three groups of writers of different writing proficiency were discussed. In terms of structure, LBs were classified into three main groups: (1) NP-based; (2) PP-based; and, (3) VP-based. NP-based refers to any noun phrases with post-modifier fragments. PP-based is defined as those LBs beginning with a preposition and a noun-phrase fragment. The last category includes any words that have a verb component. The results show that structurally, the three groups employed NP-based bundles most differently. When grouping NP-based LBs into those with *of* such as *in the context of* and those without *of* such as *the way in which*, it is notable that NP-based without *of* does not occur in the non-native corpus, and these LBs are shown to be a part of relative clauses, such *as the extent to which, the fact that this,* and *the way(s) in which.* When the NP-based and PP-based LBs are grouped into two productive frames ('*the* + Noun + *of the/a*', and '*in the* + Noun + *of*), it was found that

the importance of nominal or prepositional expressions were not recognized by both native and non-native students. However, both groups relied more on VP-based bundles than the native expert group.

In terms of function, three major categories were identified: referential bundles, discourse bundles and stance bundles. The study adapted the taxonomy from Biber and Barbieri, 2007 and Biber et al., 2004; 2004). The results reveal that while a higher proportion of referential LBs are found among the expert group, the other two groups relied more on discourse organizer LBs. In summary, the results reveal that LB usage in native and non-native student groups is similar in that both groups rely on more VP-based and discourse organizer bundles than the native expert group. In the present study, the structural taxonomy of Chen and Baker (2010) will be used with a number of adaptations (see Section 7.3.1).

In addition to Chen and Baker (2010), Liu's (2012) taxonomy was adopted to categorize LBs in terms of functions because of the detailed subcategories and comprehensive examples provided. In Liu's study (2012), these recurring strings of words are referred to as multi-word constructions (MWCs). His study aims to investigate the most prevalent MWCs in academic writing. The corpora used for identifying these MWCs were sub-corpora in the Corpus of Contemporary American English (COCA) and the British National Corpus (BNC). In selecting the most frequently found MWCs, the frequency and range cut-offs were considered. The frequency of 20 tokens PMWs were used as a frequency cut-off. In terms of the distribution, an MWC must appear in six out of the eight in the COCA and five out of six academic sections in the BNC. Liu notes that the semantic/functional categorization is much more complex because of these reasons: (1) the subjectivity in classification; and (2) the multi-functional of some MWCs. Although his categorization was based on Biber *et al.* (2003; 2004 as cited in Liu, 2012), a number of different findings were found. For example, it was found that a high proportion of '*NP + action verb* (e.g., *suggest*) can be identified in academic writing.

3.5 Summary

This chapter has discussed three issues in corpus linguistics: (1) general and specialized corpora; (2) corpus-based and corpus-driven approaches; and, (3) keyword and key keywords. Firstly, with regard to the differences between general and specialized

corpora, the rationale why specialized corpora were considered as more appropriate in analyzing a specific genre, or in particular research abstracts, was provided. Secondly, while a corpus-based approach is based on pre-determined theory, a corpus-driven approach relies on a large collection of data in order that the linguistic patterns can be observed and identified. Some benefits and drawbacks over corpus annotation according to the corpus-based approach were also provided. Thirdly, the terms *keywords* were firstly defined, followed by how the term differs from *key keywords*. Prior to the summary, previous studies on lexical bundles in academic writing were presented.

Chapter 4. Research Methodology

4.1 Introduction

This chapter describes the methodological approach adopted in the study. Section 4.2 begins with a brief overview of three main research paradigms along with the ontological and epistemological stance of the study before reiterating the aims and research questions of the study. Section 4.3 presents the theoretical framework and the research design. Section 4.4 provides the research context in terms of the institution, the programs of study and some academic background of the novice researchers. Following that, Section 4.5 describes the data collection procedures. Section 4.6 explains the two datasets (a Novice Abstract Corpus (NAC) and a Professional Abstract Corpus (PAC)) along with the criteria in data selection. Turning to the analysis techniques, Section 4.7 presents the data processing and procedures in move analysis, the inter-coder reliability, and corpus analyses. Section 4.8 describes corpus analysis procedures, namely the retrieval of the word frequency lists, keyword lists, and lexical bundle (LB) lists of the research abstracts (RABs) along with the statistical measures used for each function. Then, the rationale as to why certain statistical measures are adopted for each step of the corpus analysis is discussed. Section 4.9 takes into consideration the reliability and validity measures used in the study, before finally moving on to ethical considerations.

4.2 Research Paradigms

Guba and Lincoln (1994, p. 107) define *paradigm* as 'basic belief systems based on ontological, epistemological and methodological assumptions'. A paradigm represents a worldview that the researcher holds and the possible relationships among its parts. In conducting research, the researcher's position is conveyed through his/her ontology of how things really are, or in Grix's (2002, p. 175) words, 'what is out there to know about?' Epistemology is sometimes referred to as "theory of knowledge" (Crotty, 1998, as cited in Mack, 2010). Therefore, it should be noted that the choice of research paradigm basically depends on the aims of the research as guided by the research questions and the theory adopted to explain reality. The methodology employed in data collection is the reflection of the different ontological and epistemological views of different researchers.

The section that follows thus presents a brief overview of three research paradigms, namely positivism, postpositivism and social constructivism. A number of important characteristics of each paradigm are explored and discussed before turning to the ontological and epistemological stance of this study.

4.2.1 Three research paradigms

As regards the nature of reality, positivism posits that there is one absolute truth that can be identified, quantified and measured (empiricism). The job of the researcher is, in her search for seeking access to a particular truth, to impartially and objectively describe the phenomenon under investigation. This belief is commonly held by scholars researching in the natural sciences (Guba and Lincoln, 1994). Postpositivism, on the other hand, arose as a reaction to the tenets of positivism, proposing instead that although an external reality to our subjective understandings of it probably exists, it is elusive and constantly subject to refinements in our attempts to understand it ('falsifiability') (Ponterotto, 2005). The concept of falsification was introduced by Popper (1959, cited in Scotland, 2012, p. 10), who argued that 'every scientific statement must remain tentative forever'. In contrast to the two aforementioned paradigms, social constructivism postulates that there is no one single reality, that multiple realities exist, each differing in form and content depending on how individuals, groups, and cultures perceive their environment and the phenomena in it (Corbetta, 2003).

Ponteretto (2005, p. 132) noted that the rhetorical structure found in a research study can illuminate the ontological stance of the researcher. In the positivist and postpositivist stances, rhetoric is seen as 'precise and scientific, presented in an objective manner'. In contrast, the rhetoric of the constructivist is found to be subjective and interactive, and often presented in the first person singular. The relationship between the research and the nature of reality can also be differentiated among these three paradigms. While positivism purports to be value-free, maintaining that it is possible to keep a distance between the researcher and the researched, social constructivism maintains that researchers cannot be value-free and that the researcher's voice will always and inevitably be present in the data collection and interpretation (Corbetta, 2003). These ontological and epistemological differences lead to the deployment of different methods to research phenomena. While the methods used in positivism are based on quantitative approaches, both qualitative and quantitative methods can be found in postpositivism. In

the social constructivist paradigm, however, qualitative research methods tend to be used, such as ethnography or interviews.

4.2.2 Ontological and epistemological stance of the study

Having outlined how ontology and epistemology shape the research methodology and noting the main differences among these three research paradigms, this section explains the ontological and epistemological stance of this study. Epistemologically, the present study is situated in genre theory, where written texts are categorized into different types according to their communicative purposes. Written texts are thus closely related to their social context and can be regarded as 'strategies for responding to particular social situations' (Paltridge, 1997, pp. 17-18). The notion of discourse community and its membership also directs attention to how social groups have their norms and conventional practices on how things are done (Hyland, 2012b). Among the three different perspectives of genre theory (see Section 2.4), the English for Specific Purposes (ESP) approach was selected to be the way of seeking the nature of reality. Hyland's (2000) five-move model was adopted to analyze a number of Research Abstracts (RABs) written by novice and professional researchers because it covers a wider range of disciplines than Swales' (1990) IMRD model and Bhatia's (1993) four-move model (see Section 2.8 for discussion).

Hyland's framework postulates that an abstract consists of five moves as follows: (1) Introduction, (2) Purpose, (3) Methods, (4) Product, and (5) Conclusion. These moves have different communicative purposes (see Section 4.9.1). Arguably, it is difficult to determine whether the writers are consciously or unconsciously aware of the 'moves' while they compose their abstracts unless further interviews are carried out. Nevertheless, the researcher interprets and categorizes those written texts according to these five communicative purposes. As can be seen, multiple frameworks of move analysis exist, and understandings of how and why they are writing in the way that is shown in these abstracts are probably affected by social, and cultural influences.

Move analysis has long been acknowledged as a qualitative approach to discourse analysis (Biber *et al.*, 2007; Upton and Cohen, 2009), which allows a limited number of texts to be analyzed. However, recently the combination of corpus approaches and move analysis has made it possible to increase the number of texts under scrutiny. Researchers

are able to empirically analyze linguistic features based on a larger number of authentic texts. Although there was no consensus on whether corpus linguistics (CL) should be viewed as a theory or a methodology (Grix, 2011; McEnery *et al.*, 2006; Taylor, 2008), this study regards CL as a methodology suitable for seeking what realities are in terms of the linguistic features found in a NAC and a PAC. The debate is congruent with McCarthy (2001, p. 125), who claims that CL 'occupies an uncertain territory in applied linguistics'.

McCarthy (2001) suggests that CL attempts to search for objectivity about language use by disconnecting the researcher from the subjectivity of native speakers' intuition, or, as (Baker *et al.*, 2008, p. 277) put it, to make the researcher '(relatively) free from any preconceived or existing notions regarding their linguistic or semantic/pragmatic content'. Thus, it can be concluded that the empiricism of CL can be reflected through a positivist paradigm. However, it is observable that subjectivity still plays a central role in every stage of the analysis. Baker *et al.* (2008, p. 277) argue that:

> The analyst, informed by quantitative aspects, has to decide what texts should go in the corpus, and what is to be analysed, determine which corpus-based processes are to be applied to the data, and what the 'cut-off' points of statistical significance should be.

The retrieval of concordance lines should also be interpreted to identify wider themes. It should be noted that in fact subjectivity starts from the compilation of the corpus data, where the researcher needs to make a decision on which modes (spoken or written) and which genres (e.g., academic, business or medical) serve the aims of the research.

Despite the mistaken understanding that CL is purely a quantitative approach involving numbers and frequencies (Baker, 2013), the methodology in fact relies on a mixed or multi-methods approach. After quantitative results are generated, they need to be interpreted qualitatively, and the software becomes a quick aid for searching linguistic data through concordance lines (Baker and McEnery, 2015). As suggested by Creswell (2009, 2012), qualitative and quantitative research methods should not be viewed as two discrete dichotomies lying at the polar ends, but rather seen as forms of research along the continuum. This is in line with Baker (2004) who states that corpus-based research is not an entirely quantitative form of analysis because it relies on researchers' skills in order to interpret the findings. In a similar vein, Marchi and Taylor (2018) argue that identifying and classifying linguistic patterns through discourse analysis cannot be

achieved through purely qualitative methods. Patterns and their components must be quantifiable. The incorporation of CL thus offers greater rigor and objectivity to the social lens of discourse analysis. A mixed-methods approach is therefore viewed as complementary, countering the criticism that the quantitative nature of CL tends to disregard context, and the criticism over qualitative limitations of discourse analysis on the number of texts to be analyzed (ibid., pp. 4-5).

In sum, the ontological stance of this study is based on postpositivism, in which the written abstracts are organized and constructed in a way that serves the writers' communicative purposes and meets the expectations of discourse communities. The pre-defined theory or assumption regarding the moves found in abstracts can be falsified or refuted if the evidence in the two corpora shows the occurrences of some data which are against the theory or assumption. The empirical nature of CL provides the researcher with greater reliability and more generalizable results (McEnery and Wilson, 1996, p. 77) in terms of lexical choices and other linguistic patterns found in each move of the abstracts.

Before moving on to Section 4.3, the aims and research questions of the study are reiterated here. The first aim of the study is to analyze the rhetorical composition and move patterns of a number of English language RABs written by Thai novice and international expert writers from the fields of Education and Applied Linguistics. The second aim is to identify the lexical words, keywords and LBs used in those abstracts.

Overarching question: Are there any differences between the rhetorical moves and linguistic patterns of the NAC and those of the PAC? If yes, what are their pedagogical implications?

Sub questions:

1. What are some of the similarities and differences between the rhetorical moves found in abstracts written by novice researchers and those written by professional researchers in the fields of Education and Applied Linguistics?
2. To what extent are the most frequent lexical words in the NAC similar to those in the PAC?
3. What keywords are associated with each move in (a) the NAC and (b) the PAC?

4. What lexical bundles are associated with each move in (a) the NAC and (b) the PAC?

4.3 Theoretical Framework and Research Design

In this section, the theoretical framework underpinning the study is briefly described and then the research design is presented.

4.3.1 Theoretical framework

The theoretical framework employed in this study is illustrated in Figure 4.1. The figure was adapted from Paltridge and Phakiti (2015, loc 4500). As discussed in Chapter 2 and Chapter 3, the study is mainly based on two perspectives: the ESP approach to genre analysis and the corpus-based approach. The framework illustrates that novice and professional researchers are from different discourse communities, and therefore probably engage in different norms and community practices. These norms and practices are realized through the texts (abstracts) and can be perceived by investigating their rhetorical move compositional types and move patterns using the move analysis approach. Combining corpus-based techniques (wordlists, keyword lists and n-grams), the similarities and differences in lexical choices and linguistic patterns between the two corpora are investigated.

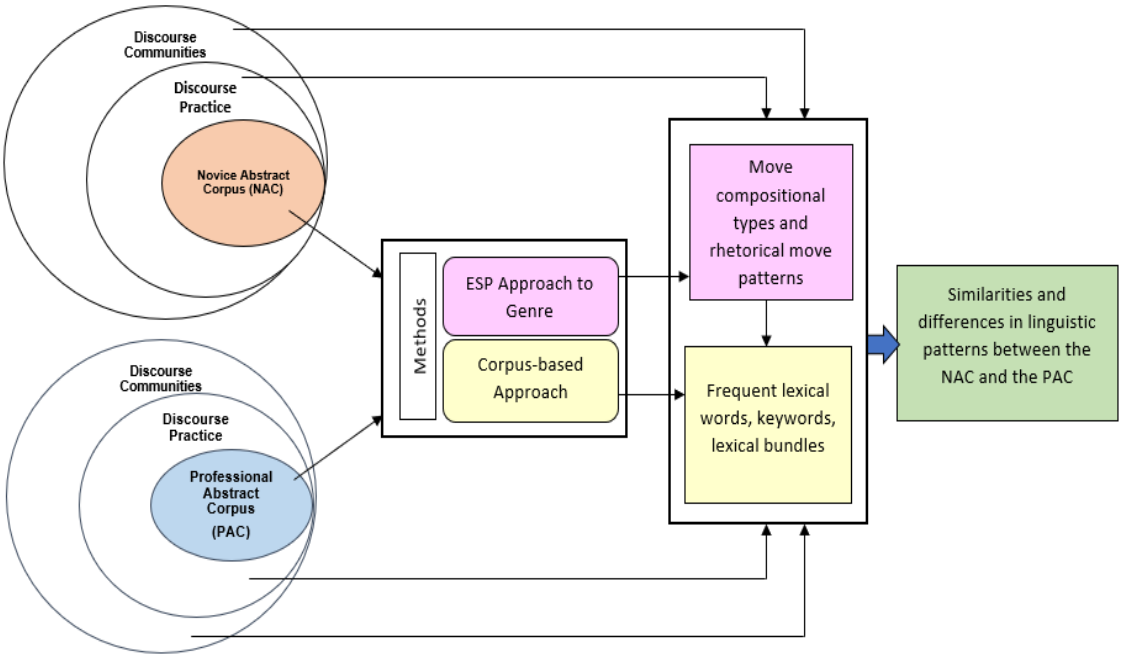

Figure 4.1 Theoretical framework
(*Adapted from Paltridge and Phakiti, 2015, loc 4500*)

4.3.2 Research design

In order to identify a number of distinctive linguistic patterns (i.e., rhetorical move patterns, frequent lexical words, keywords and LBs) in abstracts in the fields of Education and Applied linguistics, the study adopted a comparative research design with a sequential mixed-methods approach (see Figure 4.2).

- Two abstract data sets were compiled: one by novice writers and another one by expert writers;
- Hyland's (2000) framework was used to guide the analysis of move compositional types and move patterns of all the data in the two corpora;
- corpus-based techniques (wordlist, keywords and n-grams) were then used to investigate further the similarities and differences in lexical choices and linguistic patterns between the NAC and the PAC.

The data analysis was divided into two main stages, with the first stage involved with move analysis and the second stage was based on corpus analysis. After the data collection, all the selected abstracts were first analyzed in terms of move compositional types and patterns. The move analysis procedures can be regarded as qualitative because they rely on the researcher's judgement which sentence/clause/phrase belongs to m1, m2, m3, m4 or m5. However, these move types and patterns were quantified by descriptive statistics such as frequency counts and percentages, followed by inferential statistics, namely a chi-square to test whether there was a statistically significant difference among those frequently found patterns. The analysis in this stage was used to address RQ 1 regarding the similarities and differences of move rhetorical patterns and move structures of the NAC and the PAC (see details of how move patterns and structures are categorized in Section 5.2). All the abstracts were then manually tagged according to moves. Each move was stored in separate files for further analysis.

In the second stage, multiple techniques from AntConc software were utilized to generate word frequency lists, keyword lists, and LBs. Although these data were considered quantitative accompanied by numbers and statistics, they had to be further analyzed qualitatively by looking at sample concordance lines to see how they were employed in context. The word frequency lists of the two corpora were first generated to find the top twenty most frequently found lexical words. After that, the lists of high-frequency lexical words appearing in both corpora were compared. To address RQ2 regarding the degree

of similarities and differences in the usage of these words, the effect size (log-ratio) and the significance test (log-likelihood) were utilized. Each concordance line was then investigated to see whether the meanings and functions of these words were employed similarly in context. Finally, the information drawn from both corpora was compared to identify the similarities and differences in terms of lexical choices and linguistic patterns. The details and the analyses on word frequency lists, keyword lists, and LBs were used to address RQs 2, 3, and 4 respectively. More detailed descriptions of these analysis procedures can be found in Section 4.7 and Section 4.8.

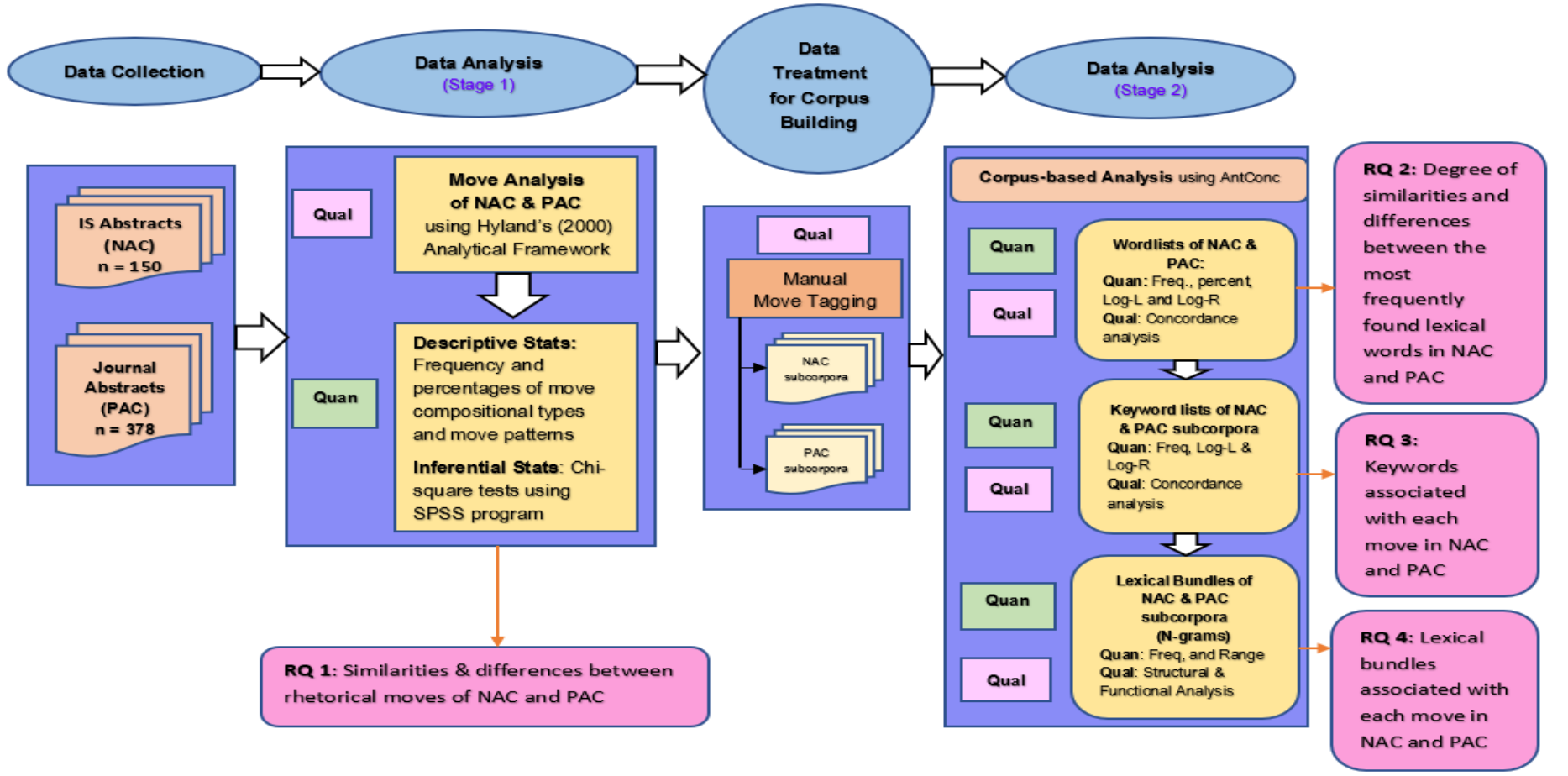

Figure 4.2 Research Design

4.4 Research Context of Novice Writers

This section presents background information on (1) the Language Institute, Thammasat University, where the data of the novice writers was collected, (2) some brief details regarding the two MA programs and also (3) the details of postgraduate students whose RABs were extracted to build up the NAC.

4.4.1 The Language Institute, Thammasat University (LITU)

LITU was officially founded in 1985 as an institute responsible for English language teaching for non-English major students from all faculties of Thammasat University (TU). Apart from offering foundation English courses and English for Specific Purposes (ESP) courses to undergraduates, LITU offers two international Master's Degree programs: the MA in Teaching English as a Foreign Language (MA in TEFL), and the MA in English for Careers (MA in CR). English is used as the medium of instruction, communication, and assessments on these international postgraduate programs.

4.4.2 Master of Arts (MA) degree programs

The source of the data selected for this study was students' Independent Studies (ISs) submitted for the two MA programs noted above from 2010 to 2013. The term 'IS' in this context is a piece of written work comparable to an MA dissertation in the UK.

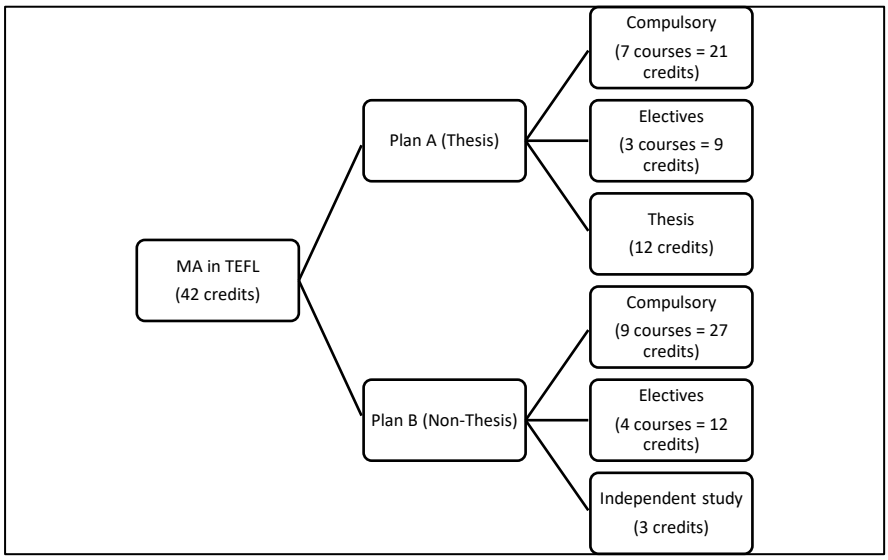

Figure 4.3 Curriculum for the MA in Teaching English as a Foreign Language

The curriculum for these two programs required students to undertake compulsory and elective courses totaling 42 credits. Despite the two optional paths that the students could select, the majority followed Plan B, meaning that they opted to undertake more taught courses and conduct an IS instead of writing a thesis to complete their degree. However, it is noticeable that the credits for an IS between the two programs were different (see Figures 4.3 and 4.4). While six credits were allotted for an IS of the MA in CR, only three credits were awarded to the IS on the MA in TEFL. Despite this discrepancy, the writing requirements, the length of the ISs, and the requirements for abstract writing in particular were similar. The abstracts written by these MA students comprise the NAC.

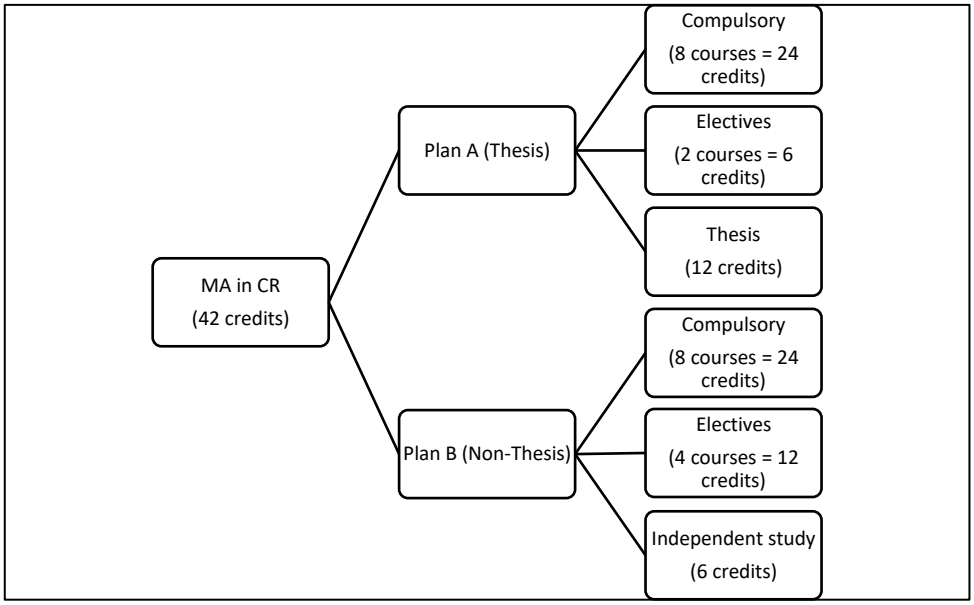

Figure 4.4 Curriculum for the MA in English for Careers

4.4.3 Background of the novice writers

Applicants who were admitted onto these international programs had to submit evidence of their English language ability, i.e. a TU-GET score (Thammasat University General English Test) of 500, a TOEFL Internet-based score of 61, a TOEFL computer-based score of 173, a TOEFL paper-based score of 500, or an IELTS band of 5.5, in addition to passing the entrance examination and a passing an interview held in English.

Upon completion of their programs, students are required to choose to write a thesis (Plan A) or write an IS *and* take a comprehensive examination (Plan B). Students who choose to write a thesis or Plan A are normally top students, and they represent only fewer than

1% of the total. It should be noted that only abstracts from ISs were used as thesis abstracts are expected to be more complicated with longer word counts and, as noted in Section 4.4.3, a few students chose to follow Plan A (thesis writing).

Despite undertaking courses in research methodology and writing, these students received no compulsory training modules on how to write a research abstract. The guidelines given for the abstract submission were an approximate length of no more than one-page in length. This might pose a question whether there might be a crucial difference of the average length between the NAC and the PAC. However, despite some differences, the approximate lengths of the two corpora shown in Table 4.1 were not as crucial as expected (233:175). Occasionally, a workshop on abstract writings was organized, but it was not an obligation for students to attend. It could be assumed that the students learned and followed the patterns of abstract writing from samples of the previous year graduates in order to ensure that their abstract conformed to the discourse community standards. The curriculums of the two programs were revised in 2015, leading to other additional requirements for program completion. One additional requirement is to disseminate part of their research study in either written or spoken form. This new policy could contribute to the increasing importance of abstract writing as it might help attract journal editors to consider the research for publication.

4.5 Data Collection Procedures

Not all the novice abstracts were retrievable online due to an unsystematic filing system. I managed to save the files directly from the CDs stored at LITU with the permission of the directors of the two postgraduate programs via email. Regarding the PAC, all the abstracts from the selected four journals were downloaded and saved in Word format and then converted into text files using AntFileConverter (Anthony, 2015).

Due to the variety of the published articles in each journal, only empirical research abstracts were included in the study. The researchers' nationality was not taken into account as these scholarly articles were believed to be native-like and met the rigorous standards of peer-review prior to publication. Review articles, conceptual papers, and book reviews were excluded as they were regarded as other genres. Articles in special issues were geared towards one particular subject, leading to an imbalance of information, and were therefore disregarded. However, the length of the abstracts was

not taken into account as it was expected that they would closely reflect the accepted length of this discourse community.

4.6 The Novice and Professional Corpora

Two corpora were created to compare the similarities and differences between the frequent words, keywords and LBs employed in abstracts written by novice and professional researchers.

4.6.1 Novice Abstract Corpus (NAC)

The NAC consists of 150 research abstracts, generated by stratified random sampling by strata in terms of year (Bryman, 2012) of abstracts from ISs written by Thai postgraduate students in MAs in CR and TEFL at TU, Thailand, during 2010-2013. As the total number of students taking the CR program during the specified period was 268, the sample represented approximately 28 percent of the total population. The remaining data from the TE program represented 79.8 percent of a total of 94 students. The imbalanced number of abstracts drawn from each year of the TE program was partly due to the different number of the yearly student intake. However, the four-year strata made it possible to have samples spread across 2010-2013.

4.6.2 Professional Abstract Corpus (PAC)

The PAC contains 378 research abstracts, selected from four international peer-reviewed journals, namely Language Learning (LL), TESOL Quarterly (TQ), English for Specific Purposes (ES) and ELT Journal (EL) over the same four-year period as the NAC. Given that one of the main objectives of this study is to find similarities and differences in the keywords found in the novice abstracts in comparison with the professional texts, the PAC was utilized as a reference corpus. These journals were chosen because of their high impact factors (ES = 1.659, LL = 1.612, TQ = 0.94, and EL = 0.720, as of 2014 in Journal Citation Reports, which is the most updated figure at the time of corpus compilation), in addition to being widely recognized by scholars in the relevant disciplines.

The size of the NAC corpus was established as 150 abstracts on the grounds that this was not too small, compared with previous studies (e.g., Ren and Li, 2011; Salager-Meyer, 1992; Samraj, 2005; Tankó, 2017), but was also a manageable size, allowing the

researcher to complete the qualitative and quantitative analysis within the timeframe set by the PhD program. The target novice groups were selected due to the convenience and accessibility of the abstract texts, since abstracts by Thai postgraduate students had not been systematically collected and made available online at the time this study was conducted.

Table 4.1 Details of research abstracts in the NAC and the PAC by year and source

Corpus	NAC		PAC			
Year	Careers (CR)	TEFL (TE)	English Language Teaching (EL)	English for Specific Purposes (ES)	Language Learning (LL)	TESOL Quarterly (TQ)
2010	19	22	25	17	20	23
2011	19	12	29	21	37	18
2012	18	19	22	22	37	23
2013	19	22	21	19	23	21
No. of files	**75**	**75**	**97**	**79**	**117**	**85**
Types	2,128	2,111	2,311	2,591	2,719	2,696
Tokens	16,438	16,692	12,922	14,395	18,569	15,231
Avg. length	231	235	146	197	173	193
Overall Types/ Tokens (files)	**3,081/ 33,130 (150)**		**5,536/ 61,117 (378)**			

In corpus design, balance and representativeness in accordance with the specific focus of the research questions should be taken into consideration (McEnery *et al.*, 2006, p. 73). The data selected by strata over the four-year-period led to a balanced corpus from the fields of Education and Applied Linguistics. In addition to being a focus of analysis in itself, the PAC was also used as a reference corpus against which the NAC could be compared in order to highlight and devote specific attention to the characteristics of the abstract writing of novices. The reason for using the PAC in this way, rather than another corpus of academic texts, is that unlike other available corpora of this kind, the PAC represents the specific genre under investigation, namely research article abstracts, rather than research articles or academic writing in general.

As illustrated in Table 4.1, a comparison of the total number of words in the two corpora shows that the PAC (61,117 words) is nearly twice the size of the NAC (33,130 words). It also contains more than twice the number of abstracts (378 versus 150). Nevertheless, a comparison of the length of the abstracts in each subcorpus shows that those produced

by novice writers tend to be longer than those written by professional writers: the length of the NAC texts ranges from 88 to 364 words, with an average of 233, while the length of the PAC texts ranges from 52 to 378, with an average of 175. This suggests that, on average, the professional writers tend to be more concise than the novices. It is worth noting that what Melander, Swales and Fredrickson (1997, p. 254) called 'house rules' might contribute to the longer average length of the NAC texts in relation to those of the PAC. As noted above, the general guidelines for the abstracts of students' ISs is that they should be one page in length.

Considering the differences in the corpus size, normalized frequencies or relative frequencies (see Section 6.2 for more detail) were, therefore, used instead of raw frequencies or absolute frequencies when comparing frequency counts between the corpora.

$$\textit{Normalized Frequencies} = \frac{\textit{raw frequencies}}{\textit{number of tokens in corpus}} \times \textit{basis of normalization (e.g. 10,000 or 100,000)}$$

As for the four journals, there is some variation in the abstract length suggested in the guidelines across the four journals. It should be noted that the conciseness of professional abstracts might be partly due to the guidelines provided by each journal, to which the writers are asked to conform. For example, LL and EL state that the length of an abstract should be approximately 150 words, while TQ and ES allow longer abstracts of 200 words. Of the four journals, only ES includes a minimum requirement (100 words).

The following sections focus on the data collection and data analysis procedures.

4.7 Data Processing and Procedures for Analyzing Moves

The approach to move analysis adopted in this study is based on Hyland's (2000) framework. This approach is detailed below, followed by an outline of the process of preparing and analyzing the NAC and PAC, including consideration of inter-coder reliability and the statistical measures adopted in the study.

4.7.1 Move analysis

Each abstract was first read through in order to familiarize the researcher with its overall content prior to categorizing its content into five moves used in Hyland's (2000) framework (see Table 4.2). Each move in the research abstracts might consist of one or more sentences, or could be composed of individual phrases or clauses. As stated in previous research (e.g. Dos Santos, 1996; Saeew and Tangkiengsirisin, 2014), embedded moves might also be found, that is, text relating to two or more moves could be combined in a single sentence. Apart from identifying moves, the researcher also analyzed the compositional types of the abstracts (see Section 5.2 for detail) so as to determine which moves are conventional, which optional, and which obligatory (see Section 5.5 for detail).

Table 4.2 Hyland's (2000, p. 67) five-move model of abstract analysis

Introduction	Establishes context of the paper and motivates the research or discussion.
Purpose	Indicates purpose, thesis or hypothesis, outlines the intention behind the paper.
Method	Provides information on design, procedures, assumptions, approach, data, etc.
Product	States main findings or results, the argument, or what was accomplished
Conclusion	Interprets or extends results beyond scope of paper, draws inferences, points to applications, or wider applications.

4.7.2 Inter-coder reliability

Gries (2011, p. 87) points out that subjective decision-making is required during the analysis (i.e. during the data coding); for example, categorization is likely to involve 'features that are not always clear-cut'. However, to avoid subjectivity as much as possible, and to seek as high a degree of reliability in the coding as possible, multiple coders were used and the steps described below were taken to evaluate and ensure inter-coder reliability.

Ten abstracts from novice writers of both disciplines (five each) and ten abstracts from professional writers randomly chosen from the four journals were piloted to check the reliability of the textual move analysis. Following brief instructions and training, two coders (PhD students in linguistics and in applied linguistics) were invited to ask any

questions to ensure that the procedures were clearly and mutually understood. All ten abstracts were then given to each of the coders, together with a table to record the moves they identified in each abstract. It should be noted that, prior to the inter-coder check, the abstracts had already been divided into units – i.e. sentences, clauses, and phrases – as seen in the modified text below. If embedded moves were found, the sentence in question was sectioned into clauses or phrases (see the modified text sample below). However, in cases where the coders did not agree with any preliminary division, this disagreement was resolved through discussion. The degree of inter-coder reliability was calculated, and the results yielded a high agreement of 95.7%.

An Example of Move Analysis

Original Text

> Motivation has long been identified as one of the main factors affecting English language learning (Gardner, 1985). This study examined the type and level of English language learning motivation (instrumental or integrative) of 30 first-year undergraduate students at an international institute of engineering and technology in Thailand. A modified motivational survey of 20 items adapted from Gardner's Attitude/ Motivation Test Battery (AMTB) was conducted. The data was analyzed by means of frequency, percentage, arithmetic mean and standard deviation. The main findings show that the students were relatively "highly" motivated and found to be slightly more "instrumentally" motivated to learn English. This reveals that instrumental motivation is a significant factor among this group of students learning English. Based on this study's findings, some relevant and useful motivational learning strategies are recommended for enhancement and improvement of the students' motivation. Relevant language improvement programs and activities are also discussed in accordance with the students' language difficulties, as communicated in the open-ended questions, to increase their motivation in English language learning.

TE_54_42 (Pilot 9)

Modified Text for Coding

(s1) Motivation has long been identified as one of the main factors affecting English language learning (Gardner, 1985).
(c2.1) This study examined the type and level of English language learning motivation (instrumental or integrative)
(p2.2) of 30 first-year undergraduate students at an international institute of engineering and technology in Thailand.

(s3) A modified motivational survey of 20 items adapted from Gardner's Attitude/ Motivation Test Battery (AMTB) was conducted.

(s4) The data was analyzed by means of frequency, percentage, arithmetic mean and standard deviation.

***Note**: 's' stands for sentence, 'c' stands for clause, and 'p' stands for phrase. In the sample above, the second sentence was divided into a clause and a phrase: c2.1 and p2.2. According to Hyland's Framework, c2.1 was categorized as move 2 (Purpose), while p2.2 was coded as move 3 (Method).*

Table 4.3 Inter-coder reliability for the NAC and the PAC

Moves	Coder 1				Coder 2			
	Units of Analysis	Agree	Disagree	%	Units of Analysis	Agree	Disagree	%
Introduction	7	7	0	100	4	4	0	100
Purpose	12	12	0	100	12	12	0	100
Method	32	31	1	96.9	28	21	7	75
Product	29	29	0	100	19	19	0	100
Conclusion	21	12	9	57.1	8	8	0	100
Total	101	91	10	90.1	71	64	3	95.7

After resolving disagreements in the coding, one issue that arose was that it would be problematic to classify the sentences according to each move as initially planned because numerous embedded moves were found, especially in the PAC texts. Moreover, breaking down the sentences into moves and then focusing on these in isolation in the analysis would reduce the discourse context being observed, potentially causing decontextualization, as mentioned by Baker (2006, p. 25). Accordingly, the moves were instead identified by annotation through tagging at the beginning and the end of sentences, so that the patterns of lexical use in each move could be observed and analyzed (see Table 4.4 for a tag set). However, as Baker (2006) points out, manual tagging can be error-prone and laborious. Consequently, after the tagging, all the annotations were double-checked by the researcher to try to limit possible errors.

It should be noted that the identification of moves in the texts of the novice writers was rather straightforward, with fewer embedded moves and less complexity of discourse usage. However, analyzing the professional abstracts was much more time-consuming, since the writers tended to combine two or more communicative functions in one sentence, leading to a higher degree of discourse complexity. It also became apparent that the descriptions provided in the move analysis framework were not comprehensive

enough to distinguish clearly between move 4 and move 5, as these two moves were closely linked and usually combined. As such, a clearer, more detailed explanation would help the analysis.

Table 4.4 Tag set, its definition and example

Tag	Definition	Example
<m1>	Beginning of move 1: Introduction	<m1> English language plays a crucial role in the study of TEP and TEPE students. It is used as a medium of instruction and communication throughout the courses. </m1> *N_TE_2010_080*
</m1>	End of move 1: Introduction	
<m2>	Beginning of move 2: Purpose	<m2> Therefore, the purpose of the present study is to investigate the errors in using the temporal prepositions "for" and "since" in different tense patterns. </m2> *N_CR_2012_057*
</m2>	End of move 2: Purpose	
<m3>	Beginning of move 3: Method	<m3> The task-based interactions of 42 college-level, native English-speaking learners of Spanish as a foreign language were examined. </m3> *P_LL_2010_186*
</m3>	End of move 3: Method	
<m4>	Beginning of move 4: Product	<m4> A relationship was found between learners' WM test scores and their tendency to modify output. Specifically, greater processing capacity was related to greater production of modified output during interaction. </m4> *P_LL_2010_186*
</m4>	End of move 4: Product	
<m5>	Beginning of move 5: Conclusion	<m5> Implications of these findings for theory, practice, and future research are discussed. </m5> *P_ES_2010_106*
</m5>	End of move 5: Conclusion	
<mx> <my>	Beginning of embedded moves, followed by moves (1, 2, 3, 4, 5)	<m3><m4> Informed by Vygotsky's (1978) sociocultural theory which puts talk at the core of successful teaching and learning, the analysis presented reveals how the teachers' thematic control leads to students' low interaction levels. </m4></m3> *P_EL_2010_002*
</my></mx>	End of embedded moves (which mirrors the beginning tag set)	

Once all the moves were annotated, a list of move patterns was compiled, along with frequency counts and percentages. To determine which moves are optional, conventional and obligatory, the move patterns were simplified by deleting any repeated moves and removing parentheses from embedded moves (see Tables 5.1a-c). Despite discrepancies

in categorizing the status of moves (i.e. whether they are optional, conventional or obligatory) in the texts that can be found among the studies reviewed in Section 5.5 (see Table 5.7), the present study followed the criteria established by Kanoksilapatham (2005) and Connor, Upton and Kanoksilapatham (2007), with 97-100% set as the frequency for a move to be classified as obligatory, 60% or more as conventional, and less than 60% as optional. As the number of abstracts under investigation has been increased from those previous studies which consist of less than 100 abstracts, the criteria to set 100% as obligatory move would be too harsh. This may result in the failure to classify any specific move as obligatory. Further explanation why 97-100% is classified as the obligatory move is provided in Section 5.5.

4.8 Corpus Analysis

After the completion of the move analysis and the annotation process, AntConc software was used to extract three kinds of information from both corpora, namely wordlists, keyword lists and cluster/n-grams, each indicating the frequencies of the features they identified. These data were then used to address the research questions (see Section 4.2.2) using the processes described below.

4.8.1 Frequency wordlist function

To address research question 2, which was concerned with the degree of similarity between the most frequent lexical words in the NAC and in the PAC, the wordlist tool was used to generate frequency wordlists of the two corpora for comparison. It is generally acknowledged that function words are the most frequent words in a corpus (Baker, 2006; McEnery and Wilson, 1996). In order to assess the similarities and differences among the 20 most prevalent lexical words found in each corpus, function words such as *the*, *of, and*, *into*, etc. were filtered out. As the corpora are of different sizes, comparisons were made using normalized frequencies.

4.8.2 Statistical measures for frequency wordlist

McEnery and Wilson (2011, p. 83) note that if two corpora are different sizes, frequency counts of words from the two cannot be compared directly without normalization. Because the sizes of the NAC and the PAC are 33,130 and 61,117 words respectively, a common base of 10,000 was used in calculating the normalized frequencies, as McEnery

et al. (2006, p. 53) suggest that the common base should be comparable to the corpus sizes. The findings were then compared using two statistical measures (log-likelihood or 'log-L' and log-ratio or 'log-R'). While log-L is a statistical significance statistic that indicates keyness (that is, whether the difference in the frequency of a specific word in the two corpora is due to chance), log-R is an effect size measure, and therefore indicates the degree of difference in the frequency of a particular word when comparing Corpus A with Corpus B. A positive log-R value is used to indicate that a particular word is more frequent in Corpus A than it is in Corpus B. A negative log-R indicates the reverse: that the word is more frequent in Corpus B than Corpus A. A larger value indicates a greater difference in frequency (i.e. a larger effect size) while a value closer to zero means that the two words occur with a very similar frequency in both corpora. Cases like this, where the effect size measure shows that two words have similar or the same frequencies in two corpora, have been called 'lockwords' by Baker (2011, p. 66). Taylor (2018) argues that lockwords can therefore be considered a counterpart of keywords, which are established on the basis of differences in frequency of occurrence in two corpora. However, it is worth noting that since 2011 the calculation of lockword, which is based on log-ratio statistic, has not been available in AntConc software. Therefore, the calculation was manually performed separately using an Excel template.

Hardie (2014) provides the following explanation of the meaning of the values of log-R, which is the binary log of the ratio of the relative frequencies:

- Log-R value = 0: the word in question occurs in the two corpora with the same relative frequency;
- Log-R value = 1: the word occurs two times (2 x 1) more frequently in the target corpus than in the reference corpus;
- Log-R value = 2: the word occurs 4 times (2 x 2) more frequently in the target corpus than in the reference corpus;
- Log-R value = 3; the word occurs 8 times (2 x 2 x 2) more frequently in the target corpus than in the reference corpus.

4.8.3 Keyword function

The keyword lists for the NAC were generated using the PAC wordlist as a reference corpus. Then, the NAC wordlist was used as a reference to find keywords in the professional abstracts. A threshold of 100 words was set to determine what the keywords are in each move of the abstracts and overall. Initially, the possibility of using the written

text section of the British National Corpus (BNC) as another reference corpus was explored, in order to compare the keywords of the NAC and the PAC with a source of general written English. However, keywords that were identified through this comparison with the BNC did not yield interesting results. Some of the words (e.g., *facebook* and *corpus*) occur mainly because the BNC is composed of texts from the 1980s-1990s (see Appendix J). Therefore, the decision was made to focus only on keywords identified using the PAC as a reference corpus. Grabowski (2015, p. 25) argues that a comparison against a general corpus is not as effective as a comparison of corpora of similar genres and text types. In Grabowski's study, four subcorpora, namely texts of patient information leaflets, summaries of product characteristics, clinical trial protocols and academic textbooks on pharmacology, were compared against the entire Corpus of English Pharmaceutical Texts in order to find keywords and LBs specific to each genre.

4.8.4 Statistical measures for keyword lists

Three statistical measures were used to identify the keywords: frequency, log-likelihood (log-L) and dispersion. In addition to frequencies, the inferential statistic, namely log-L, was used to determine whether differences in frequencies between the target corpus and reference corpus were statistically significant. Because log-L is calculated in relation to the corpus size, the raw rather than normalized frequencies were calculated to generate the keywords list, with the highest significance scores appearing at the top. Compared with the chi-square test, log-L is more suitable when the data is not normally distributed (McEnery and Wilson, 1996). In corpus linguistics, a widely accepted threshold of significance value is $p<0.01$ (Gabrielatos and Marchi, 2012). However, as the software showed only keyness, a critical value of at least 6.63 was therefore required to represent the same 99% level of significance.

The words at the top of the NAC keyword list are found to be words that have zero occurrence or very few occurrences in the PAC. This is because the keyword calculation method is based on a comparison of frequent words in one corpus (NAC) which occur with remarkably infrequent words in another corpus. Many of these words are found to have very little value as lexical words to be included in lists of words taught in their abstract writing. However, they are of great value in terms of raising novice writers' awareness that some of these words are terms that need not or should not necessarily be used in their abstract writing such as *(standard) deviation* and *SPSS*. In light of this

observation, formulae taken from the website associated with Rayson and Garside (2000) were employed to filter out words with zero occurrence (see Section 6.4 for keyword lists with zero occurrence in relation to the reference corpus). It should be noted that the analysis was done before AntConc introduced a new version equipped with various statistical measures in 2018. Only raw frequencies with keyness values were shown in the previous version.

Although the preliminary focus of the study is on the keyword lists rather than the frequency lists, the findings raise a number of issues that illustrate the importance of also paying attention to and interpreting the frequency lists. Focusing only on the keyword lists means that the differences between the NAC and the PAC are highlighted, which can encourage the assumption that the lexical choices in the two subcorpora are more likely to be different than similar, or that differences are more important than similarities. However, the frequency wordlists reveal that some of the most frequent lexical words are found with similar rates of occurrence in the two corpora.

The findings emerging from the search for keywords could certainly be beneficial in raising students' awareness of certain differences in the ways that novices and professionals approach abstract writing. Nevertheless, they also made clear that, in terms of supporting the development of pedagogical strategies and materials, other information about the two corpora and their characteristics also needed to be identified. As a result, the focus of the study shifted to the identification and analysis of LBs, which are recurring multi-word constructions that can be retrieved from corpora using the cluster/n-gram functions of corpus software.

4.8.5 Clusters/n-grams function for lexical bundle lists

The cluster/n-gram function of AntConc was used to extract multi-word units in each move of the two corpora. Broadly speaking, this analysis adopted the approach to LBs of Biber *et al.* (2004), in which recurrent bundles are extracted and listed by frequency. Then, the extracted LBs are categorized according to their grammatical structures and their functions. The LB analysis draws on the taxonomies of Chen and Baker (2010) for the grammatical structures, and Liu (2012) for the functions (see Section 7.3 for more detail).

4.8.6 Statistical measures for LBs

There are three criteria that need to be considered in the retrieval of LBs from a corpus: (1) the frequency of occurrence; (2) bundle length; and, (3) distribution. With respect to the last of these, three different terms - *distribution*, *range*, and *dispersion* - have been used in the literature to refer to the number of corpus texts or files that LBs are found in (Esfandiari and Barbary, 2017). In this study, the terms are used interchangeably. Some methodological issues related to overlapping bundles are also discussed in Section 7.2.4. The cluster size was set at 3 to 4 words. However, other parameters were adjusted according to the number of words found in each move (see Section 7.2 for further detail).

4.9 Reliability and Validity of the Study

The trustworthiness of a study can be judged by its reliability and validity. In what follows, these two concepts are discussed in relation to the present study.

4.9.1 Reliability

Reliability concerns the degree to which research can be replicated (Bryman, 2012) and produce the same results and similar conclusions. Bloor and Wood (2006, p. 147) argue that it is in practice impossible to fully achieve these due to a number of factors. However, reliability can be nevertheless be optimized in a number of ways. For example, meticulous records of the methodology and analysis procedures can be maintained in such a way that it allows other researchers to replicate the study (Bloor and Wood, 2006). Bryman (2012) affirms that replication is impossible if the procedures are not described in detail. With regard to data coding, reliability can be enhanced through cross-checking with other coders so that ambiguities can be resolved.

In this study, two types of reliability check were performed with the data coding and the classification of the lexical bundles. First, inter-coder reliability was performed in the pilot study before the move tagging process, as discussed in Section 4.8.3. MacKey and Gass (2005, p. 242) suggested that it is necessary to have more than one coder wherever possible, and these coders should be carefully selected and trained. MacKey and Gass (2005, p. 244) also pointed out that for simple percentages, above 75% may be regarded as good, despite the ideal percentage being over 90%. The intra-coder reliability was performed after one month to check the consistency in move tagging and the move

classifications. Of 528 files, 74 (14%) needed minor correction in terms of move tagging (e.g., the missing backslash) and move classifications. This resulted in 86% intra-coder reliability.

In terms of the lexical bundles, an intra-rater reliability was performed after two weeks to check the consistency in the classifications. Simple percentages were calculated for intra-coder reliability, resulting in 94% reliability. In addition to seeking reliability, the study provided exemple abstracts, along with the detailed move tagging procedures (see Section 4.8.1) and detailed analysis procedures (see Section 4.8.3) in order to allow replicability, which can in part enhance reliability. As mentioned in Dornyei (2007, p. 265), confidence in qualitative studies can be increased with a detailed procedure provided in the methodology if the researchers leave an 'audit trail'.

4.9.2 Validity

Validity checks of hypotheses can be performed using methodological triangulation (Layder, 1993, as cited in Baker and Egbert, 2016, p. 3). Triangulation involves using various methods, analysts, or datasets to examine the phenomena from more than one perspective. In this study, multiple methods and statistical measures are used. Some of the findings will be triangulated for validity checks and this can confirm the interpretations of the results.

4.10 Ethical Considerations

The present study aims to analyze abstracts of novice and professional researchers, which are commonly accessible freely as a preview of research articles. Therefore, there is no concern regarding copyright issues. Nevertheless, it is worth noting that text and data mining for non-profit research are exempted from copyright rules according to a recent document issued by the UK Intellectual Property Office (Baker and McEnery, 2015). This enables researchers to duplicate 'any copyright material for the purpose of computational analysis if they already have the lawful right to read the work' (ibid., p. 11). It should be acknowledged that this exception varies from country to country, and at present it applies to research within the UK.

In addition to considerations of copyright, the confidentiality and anonymity of the selected data must be maintained. This was achieved by storing all the files with

systematic codes for reference purposes. The codes P and N were used to differentiate between the professional and the novice groups, followed by group name, such as CR or TE, the year and the file number (see Appendixes C-D).

4.11 Summary

This chapter has provided a brief overview of the positivist, postpostivist and social constructivist research paradigms. Despite the paradigmatic differences between the genre-based approach and corpus-based approach, the incorporation of quantitative methods in the latter helps to strengthen the results of the former. Also, a corpus-based approach enables researchers to analyze a large number of texts in terms of lexical choices, and the results can be measured by statistical values. This study is situated in the social constructivist paradigm, as the theoretical framework showed that the NAC and the PAC came from different academic communities, in which a number of the discourse practices are more or less the same. The different practices of the two groups can be reflected in the written texts in the NAC and the PAC. A description of the two datasets, the criteria for their selection, and the data collection procedures have been provided. As for the analysis techniques, the data analysis procedures in relation to the move pattern classifications have been explained and a number of examples given in order to allow the replicability. The four methods from the corpus analysis involved in the retrieval of word frequency lists, collocates, keyword lists, and lexical bundle lists of the RABs along with the statistical measures used for each function have been discussed. The final section has highlighted the trustworthiness of the study in relation to reliability, validity and ethical considerations.

Chapter 5. Rhetorical Move Compositions and Move Patterns of Abstracts

5.1 Introduction

This chapter compares novice and professional abstracts by adopting Hyland's (2000) analytical framework of abstracts to analyze and explain particular communicative events or moves occurring in those abstracts. As mentioned in Section 2.8, the framework consists of five moves, as follows: (1) Introduction, (2) Purpose, (3) Method, (4) Product, and (5) Conclusion.

The aim of the analyses is to address the following question:

- **(RQ1)** What are the similarities and differences between the rhetorical move compositions and move patterns found in abstracts written by novice researchers and those written by professional researchers in the fields of Education and Applied Linguistics?

The discussion presents the move analysis from three angles, namely, the moves that the abstract is composed of, the sequence of move patterns, and the consideration of whether the move is optional, conventional or obligatory. Following the introduction, Section 5.2 begins with a discussion of how the moves have been classified, in terms of both move patterns and the moves in each abstract. Section 5.3 compares the most commonly found move compositional types of abstracts written by novice and professional researchers, in terms of the raw frequencies and the relative proportions. Sections 5.4 then compare the total frequencies and the proportion of move patterns occurring in the NAC and the PAC. In Section 5.5, the classifications of moves as optional, conventional, or obligatory are discussed before moving on to the summary in Section 5.6.

With regard to the rhetorical move composition, it is interesting to find out which move compositional types are the most frequently found in the novice and professional corpora. If those frequently used compositional types in the two corpora are found to be of the same type, the detailed information on further analysis of move patterns is important to see how similar or different they are. As mentioned in Section 1.4 that despite the recognition of the embedded moves in previous studies, the issue has not yet been studied

in depth, the findings would therefore contribute to the existing knowledge with regard to the embedded moves found in the abstracts written by novice and professional writers.

5.2 Move Classification

This section focuses on the move classification of the abstracts. The samples presented in Tables 5.1a-c show how the move patterns and types of compositional moves have been classified. In some abstracts, a move occurs more than once; where this is the case, the recurrence is not counted in the classification of move compositions.

The top three move compositional types {2, 3, 4, 5}, {2, 3, 4} and {1, 2, 3, 4, 5} found in either the NAC or the PAC are selected as exemplars. Despite some distinguishable variations of move sequences between the two corpora, the top three compositions are found to be the same as shown in Tables 5.2 and 5.3.

Table 5.1a A sample of abstracts with explanations of the classification of the move patterns and move composition type {2, 3, 4, 5}

m1 = Introduction, m2 = Purpose, m3 = Method, m4 = Product, m5 = Conclusion

***Note**: The tagging in angle brackets correspond to the moves with stands for embedding moves. The backslash inside the brackets suggest the ending of each tagging.*

Sample 1: P_EL_2010_011	
Abstract	**Explanation**
<m2><m3>*This pilot study explores* the apologies Japanese nursing college students thought they would use in their L1 (i.e. Japanese) and their L2 (i.e. English) when bumping into each other.</m3></m2>	Apart from telling the aim of the study by using the phrase, '***This pilot study explores***…', the first sentence also provides the readers with the detail of the population of the study (Japanese nursing college students). Therefore, the sentence is categorized as m3 embedded in m2.
<m3><m4>The students completed *a questionnaire*, the results of which indicated that they believed they should always apologize for bumping into someone. </m4></m3>	M3 is realized by stating that ***a questionnaire*** was completed. With the following clause mentioning the results found in the questionnaire, the sentence is thus classified as m4 embedded in m3.
<m4>The paper describes apologies students expected to use in both languages and their observations of real-world apologies in Japan and the US.</m4>	Continuing from the second sentence which some findings were presented, this sentence is categorized as m4. It seems that the writer tries to give more explanation of the apologies that the respondents indicated in the questionnaire. However, it should be noted that the sentence, if placed in the beginning of the abstract, could be interpreted as m2.
<m3><m5>*After reviewing* differences between their L1 and L2 apologies and their perceived and observed apologies, *the paper discusses* students' sociopragmatic, pragmalinguistic, and linguistic discoveries. </m5></m3>	The combination of m3 and m5 are shown in this sentence. The sentence begins with the method (m3) by using the phrase, '***After reviewing***…', followed by the phrase '*the paper discusses*…', which is interpreted as m5.
<m5>The findings have ***implications*** for instructors who are interested in investigating and teaching the speech act.</m5>	The word '***implications***' gives us a hint that the writer tries to draw some implications from this study in the conclusion part (m5).
Categorization	Move Pattern: 2(3)-3(4)-4-5(3)-5 Move Composition: 2, 3, 4, 5

Table 5.1b A sample of abstracts with explanations of the classification of the move patterns and move composition type {2, 3, 4}

m1 = Introduction, m2 = Purpose, m3 = Method, m4 = Product, m5 = Conclusion
***Note**: The tagging in angle brackets correspond to the moves with stands for embedding moves. The backslash inside the brackets suggest the ending of each tagging.*

Sample 2: N_TE_2014_143	
Abstract	**Explanation**
<m2><m3> ***This study focused on*** the attitudes of students who study in bilingual program in high school in Bangkok towards using English as a medium of instruction in content subjects. ***The aim of this study was to survey*** students' attitudes towards using English as a medium of instruction in three subjects which are math, science and social studies and to investigate the effects of using English as a medium of instruction on *the development of students' subject knowledge* and *English language skills*. </m3></m2>	These two sentences showed that m3 is embedded in m2. As a reference to what '***this study focused on***…', the first sentence is classified as m2. However, within the same sentence the information about the participants (see the underlined part) is described, which is therefore coded as m3. As for the second sentence, the phrase '***the aim of the study was to survey***…' is categorized as m2. Information about dependent variables (*using English as a medium of instruction*) and independent variables (*the development of students' subject knowledge and English skills*) is included in the second sentence, which is considered as m3.
<m3> ***The subjects*** of this research study were 60 students of M.3 who were studying in bilingual program in one private school in Bangkok, Thailand. They were ***asked to complete the provided questionnaires*** by rating their satisfaction and adding more comments about advantages and disadvantages of learning by using English in content subjects. </m3>	Further details concerning the participants are described in sentence 3, and details about tools of the study, and procedures in collecting data are in sentence 4. Therefore, these two sentences are categorized as m3.
<m4> ***The result of the study revealed that*** the students showed significantly more positive attitudes than negative attitudes. ***As to the students' comment***, most of them were satisfied with the benefit of using English as a medium of instruction in content subjects in terms of helping them in gaining more subject knowledge and world knowledge, encouraging them to raise self-development awareness and improving their English language skills. </m4>	The results of the study without any discussions, interpretation, or conclusions are shown in these two sentences. By saying that '***The result of the study revealed that*** …', it seems that the writer tries to guide the readers towards the findings. With reference to the phrase '*As to the students' comment*, …', the writer tries to give a hint of the next topic in the findings. Therefore, both sentences are categorized as m4.
Categorization	**Move pattern**: 2(3)-3-4, **Move composition**: 2, 3, 4

Table 5.1c A sample of abstracts with explanations of the classification of the move patterns and move composition type {1, 2, 3, 4, 5}

m1 = Introduction, m2 = Purpose, m3 = Method, m4 = Product, m5 = Conclusion

***Note**: The tagging in angle brackets correspond to the moves with stands for embedding moves. The backslash inside the brackets suggest the ending of each tagging.*

Sample 3: N_CR_2012_40	
Abstract	**Explanation**
<m1>***With the importance of*** English as the universal language, the number of learners of English as an international means of communication has been growing hand in hand with the number of non-native English teachers. As Thailand is entering the ASEAN Economic Community (AEC), which is expected to be established in 2015, ***increasing demand for*** *English acquisition can be anticipated*. This, together with an impact of the AEC in terms of the free flow of human resources, may ***result in an escalating number of*** non-native English teachers from other countries within Southeast Asia.</m1>	By stating the importance of the study with the phrases, '***With the importance of English*** …', '… ***increasing demand for*** ….', and '… ***result in escalating number of***…', the section is categorized as m1. According to Hyland (2000, p. 67), m1 provides readers with some contexts and motivates the research.
<m1><m2><m3>As learners' perceptions toward the language teachers ***is one of the most crucial components in*** acquiring a foreign language, this study of students' attitudes toward their English teachers ***were conducted aiming to*** measure <u>their degree of attitudes</u> toward non-native English teachers from the ASEAN community as well as to determine <u>the factors involved in</u> shaping such attitudes. </m3></m2></m1>	The continuation of m1 is shown in the dependent clause beginning with '*As learners' perceptions toward* …'. The phrase '***one of the most crucial components in*** …' is used to emphasize the importance of the research. The main communicative purpose of this sentence, however, is to tell readers the aim of the study by saying '…. ***were conducted aiming to*** …', which is considered as m2. Interestingly, m3 is also embedded in the sentence by providing readers with some hints regarding the variables of the study (<u>degree of attitudes</u>…, and <u>the factors involved in</u>…).
<m3>The data from this ***survey study*** was collected from 80 ***Thai undergraduate students*** from Assumption University ***in January 2013***.</m3>	Further details about the population (***Thai undergraduate students*** …, the type of study (***survey study***), and the period of data collection (***in January 2013***) are described in this sentence. Therefore, this sentence fits into m3.
<m4>***The key findings revealed that*** most students held positive attitudes toward non-native English teachers from the ASEAN community based on several factors: students' background and their views toward the teachers' performances in the aspect of communication effectiveness, teaching strategies and the relationship between teachers and students.</m4>	By saying '***the key findings revealed that***…' and no further discussions about the findings are provided, the sentence is categorized as m4.

Table 5.1c A sample of abstracts with explanations of the classification of the move patterns and move composition type {1, 2, 3, 4, 5} *(continued)*

<m5>The results from this study ***may serve as*** a basis for educators to develop English teaching strategies in Thailand.</m5>	By employing a hedge such as '***may serve as***…' in this sentence, it is likely that the writer tries to interpret the result and therefore needs to be cautious not to make a too strong claim.
Categorization	**Move pattern**: 1-2(1-3)-3-4-5, **Move composition**: 1, 2, 3, 4, 5

Having looked at how move patterns and move compositions in the abstracts are classified, the following sections will identify and examine the various types of move compositions of novice and professional writers. In addition, inferential statistics (namely chi-square tests) will be employed to test the differences in the two groups in using these types of move patterns.

5.3 Comparisons of Rhetorical Move Compositional Types between Novice and Professional Writers

5.3.1 Rhetorical move compositional types of novice writers

The analysis identifies the types of move compositions found in the NAC and its subcorpora. Table 5.2 shows the frequency of the seven types of move compositions (a - g) found in the NAC. Only four compositions (a - d) appear in both the CR and TE groups. M1 (the introduction move) was not found in the two most frequently found compositions in the NAC, which consisted of moves 2, 3, 4, 5 (38.66%), and moves 2, 3, 4 (26.66%). The third most frequent composition contained all five of the moves (24%). It is noticeable that three compositions (e, f and g) were specific to each subgroup of the NAC with only one example in each case: that is, the composition {2, 3} was found only in the CR group, while the compositions {1, 2, 3, 5} and {2, 3, 5} appeared only in the TE group.

Table 5.2 Frequencies of move compositional types in Novice Abstract Corpus (NAC)

Move compositions	CR*		TE*		NAC	
	Freq.	%	Freq.	%	Freq.	%
a) 2, 3, 4, 5	29	38.66	28	37.33	57	38.00
b) 2, 3, 4	23	30.66	16	21.33	39	26.00
c) 1, 2, 3, 4, 5	16	21.33	19	25.33	35	23.33
d) 1, 2, 3, 4	6	8.00	10	13.33	16	10.66
e) 1, 2, 3, 5	0	0.00	1	1.33	1	0.66
f) 2, 3	1	1.33	0	0.00	1	0.66
g) 2, 3, 5	0	0.00	1	1.33	1	0.66
Total	**75**	**100**	**75**	**100**	**150**	**100**

CR = English for Careers Group TE* = Teaching English as a Foreign Language Group*

5.3.2 Rhetorical move composition types of professional writers

Table 5.3 presents the rhetorical move compositional types found in the PAC, including the frequencies and percentages of each subgroup of the PAC and the overall picture of the compositions used by the professional writers as a whole. To facilitate the comparisons between the NAC and the PAC, the compositional types are arranged in the same sequence as the NAC data in Table 5.2 for the first seven types (a–g), instead of in order of the frequencies or the percentages.

Table 5.3 Frequencies of move compositional types in Professional Abstract Corpus (PAC)

Move compositions	EL*		ES*		LL*		TQ*		PAC	
	Freq.	%	Freq.	%	Freq.	%	Freq.	%	Freq.	%
a) 2, 3, 4, 5	23	23.71	26	32.91	56	47.86	26	30.59	131	34.66
b) 2, 3, 4	8	8.25	6	7.59	26	22.22	8	9.41	48	12.70
c) 1, 2, 3, 4, 5	21	21.65	28	35.44	19	16.24	29	34.12	97	25.66
d) 1, 2, 3, 4	14	14.43	8	10.13	11	9.40	9	10.59	42	11.11
e) 1, 2, 3, 5	8	8.25	5	6.33	2	1.71	8	9.41	23	6.08
f) 2, 3	3	3.09	1	1.27	0	0.00	1	1.18	5	1.32
g) 2, 3, 5	6	6.19	2	2.53	1	0.85	2	2.35	11	2.91
h) 1, 2, 3	12	12.37	1	1.27	0	0.00	0	0.00	13	3.44
j) 1, 3, 4, 5	1	1.03	0	0.00	2	1.71	2	2.35	5	1.32
j) 1, 3, 5	1	1.03	0	0.00	0	0.00	0	0.00	1	0.26
k) 1, 3, 4	0	0.00	1	1.27	0	0.00	0	0.00	1	0.26
l) 2, 3, 4	0	0.00	1	1.27	0	0.00	0	0.00	1	0.26
Total	**97**	**100**	**79**	**100**	**117**	**100**	**85**	**100**	**378**	**100**

EL = ELT Journal Group* *ES* = English for Specific Purposes Journal Group*
LL = Language Learning Journal Group* *TQ* = TESOL Quarterly Journal Group*

Comparisons of the abstracts (Tables 5.2 and 5.3) reflect seven different (basic) compositional structures in the NAC and 12 compositional types in the PAC. Although the patterns in the PAC exhibit greater variety than those in the NAC, the top four most common compositional types in the NAC are very similar to those in the PAC, with the exception that in the PAC types {1, 2, 3, 4, 5} and {2, 3, 4} are in the reverse order of the second and then the third place. The most common move compositional types in the PAC are as follows: {2, 3, 4, 5} (34.66%), {1, 2, 3, 4, 5} (25.66%), {2, 3, 4} (12.70%), and {1, 2, 3, 4} (11.11%). The remainders were found to occur with frequencies below 7%. It is worth noting that these compositional types (e-l), combined, account for 15.85%.

Although the compositional types used by the professional group were found to be of greater variety than the novice group (12:7), this might simply be attributable to the larger number of abstracts in the professional corpus. Because the seven compositional types (a – g) were found in both groups, these types were further analyzed statistically to determine whether the two groups employed them differently.

Therefore, a chi-square test was conducted using the SPSS v.23 program to test whether there was any statistically significant difference in the frequency of the seven compositional types employed by the two groups with the hypotheses below:

H_0: There is no statistically significant difference between the seven compositional types of abstracts written by the novice and professional groups at $p \leq .05$.

H_1: There is a statistically significant difference between the seven compositional types of abstracts written by the novice and professional groups at $p \leq .05$.

Table 5.4a A Cross-tabulation table with chi-square analysis showing the hypothetical variables (groups: NAC and PAC) and the compositional types of moves (a-g)

			Move Compositions							Total
			2, 3, 4, 5 (a)	2, 3, 4 (b)	1, 2, 3, 4, 5 (c)	1, 2, 3, 4 (d)	1, 2, 3, 5 (e)	2, 3 (f)	2, 3, 5 (g)	
Group	NAC*	Count	57	39	35	16	1	1	1	150
		Expected Count	55.4	25.6	38.9	17.1	7.7	1.8	3.5	150.0
	PAC*	Count	131	48	97	42	25	5	11	359
		Expected Count	132.6	61.4	93.1	40.9	18.3	4.2	8.5	359.0
Total		Count	188	87	132	58	26	6	12	509
		Expected Count	188.0	87.0	132.0	58.0	26.0	6.0	12.0	509.0

NAC = Novice Abstract Corpus* *PAC* = Professional Abstract Corpus*

Chi-Square Tests

	Value	df	Asymptotic Significance (2-sided)
Pearson Chi-Square	21.857[a]	6	.001
Likelihood Ratio	25.065	6	.000
Linear-by-Linear Association	10.395	1	.001
N of Valid Cases	509		

3 cells (21.4%) have expected count less than 5. The minimum expected count is 1.77.

Although the Cross-tabulation generated automatically by the SPSS program in Table 5.4a reveals that the Pearson chi-square is 21.857 ($p = .001$, $df = 6$), there are 3 cells (21.4%) with expected counts of 5 or less, meaning that the result is unreliable. As each of the last three types, {1, 2, 3, 5}, {2, 3} and {2, 3, 5}, appears only once in the NAC, leading some cells to have an expected count less than 5, these types were then recoded into one group and referred to as 'others', as shown in Table 5.4b.

Table 5.4b A Cross-tabulation table with chi-square analysis showing the hypothetical variables (groups: NAC and PAC) and the compositional types of moves (recoded a-g)

			Move Compositions					Total
			2, 3, 4, 5 (a)	2, 3, 4 (b)	1, 2, 3, 4, 5 (c)	1, 2, 3, 4 (d)	Others 1, 2, 3, 5/, 2, 3/, 2, 3, 5 (e, f, g)	
Group	NAC*	Count	57	39	35	16	3	148
		Expected Count	55.4	25.6	38.9	17.1	13.0	148.0
	PAC*	Count	131	48	97	42	41	343
		Expected Count	132.6	61.4	93.1	40.9	31.0	343.0
Total		Count	188	87	132	58	44	509
		Expected Count	188.0	87.0	132.0	58.0	44.0	509.0

NAC = Novice Abstract Corpus, PAC* = Professional Abstract Corpus*

Chi-Square Tests

	Value	df	Asymptotic Significance (2-sided)
Pearson Chi-Square	21.453[a]	4	.000
Likelihood Ratio	23.929	4	.000
Linear-by-Linear Association	7.746	1	.005
N of Valid Cases	509		

0 cells (0.0 %) have expected count less than 5. The minimum expected count is 12.97.

Table 5.4b reveals that the Pearson chi-square is 21.453 ($p \leq .001$, $df = 4$). There are no cells with expected counts of 5 or less, meaning that the result is reliable. Therefore, the alternative hypothesis was accepted with a risk of making a Type 1 error[7]. The results show that there is a statistically significant difference between the frequencies of the seven compositional types in the NAC and in the PAC, at $p \leq .05$.

As shown in the Cross-tabulation table, it is noticeable that the novice group tends to employ the compositional move {2, 3, 4} more frequently than the expected count (39:25.6). In contrast, the professional group tends to use this composition less often than the expected count (48:61.4). Regarding the last three compositions, which were recoded into one group, it is apparent that the novices seem to use these compositions (e, f, and g) much less frequently, as shown in the numbers of the actual count versus the expected count (3:13).

5.4 Comparisons of Rhetorical Move Patterns in the NAC and the PAC

5.4.1 Rhetorical move patterns in the NAC

This section examines the rhetorical move patterns found in the NAC and its subcorpora. For the classification of the move compositions in the previous section, their recurrence was ignored. In looking at the patterns of moves in this section, the sequence in which they occur is obviously important, and therefore instances of recurrence will be noted and counted. To begin with, for purposes of illustration two exemplar abstracts are organized into moves according to their communicative purposes. Given that some

[7] A type 1 error (false-positive) occurs if a null hypothesis is rejected according to the probability set prior to the testing. This is the error of accepting the alternative hypothesis because the results might occur due to chance.

sentences may have more than one communicative purpose, any identified secondary purposes are treated as signs of embedded moves and displayed in parentheses: for example '2(3)' indicates move 2, with move 3 embedded. While the move compositions in Sections 5.3 and 5.4 are separated by commas, it should be noted that hyphens here are used to separate each move, reflecting the fact that the notation indicates a sequential order. Below are two samples of move pattern classification.

Sample 4

<m2><m3> ***This study was conducted to*** investigate the attitudes of Math-Science Program students towards learning English as a core course. The area of the study intends to explore students' attitudes in various aspects, including their learning preferences, teachers, classroom activities and materials as well as their objectives in learning English. </m3></m2> <m3> The participants of this research study were M.4-M.6 Math Science Program students in Udonthani province. Questionnaires and open-ended questions were created and distributed to 94 participants to collect data and the SPSS/PC program was used to analyze the data which were obtained from the questionnaires. </m3> <m4> The results of the study reveal that most of the participants have positive attitudes towards learning English. However, the findings also reveal that students require some certain aspects including teaching content and supplementary materials to be improved; therefore, their learning will become more efficient. Interestingly, the data from open-ended questions indicate that students believe that self-learning and activities outside the classroom context can improve their English proficiency. </m4> <m5> The findings which reflect Math-Science Program students' attitudes towards learning English can be used to improve the English learning achievement of the students. </m5>

N_TE_2012_110

Sample 4 has been classified as pattern 2(3)-3-4-5. Although it fulfills the main objective associated with m2 of indicating the aim of the study ('*This study was conducted to investigate*….'), the first sentence also gives readers a sense of the research population in the reference to '*Math-Science Program students*'. M3 is therefore embedded in the first sentence. More details of the number of participants, the research tool, and the statistical program used are further explained in sentences 3 and 4, which together have therefore been classified as m3. The findings, or m4, are reported in the next three sentences, beginning with: '*The results of the study reveal that*….'. The abstract ends with: '*The findings… can be used to…*', reflecting some pedagogical implications; these sentences are thus classified as m5. It is worth noting that the abstract could be more concise if the detail regarding the program of the study (Math-Science Program) were presented only once to avoid redundancy. Some information regarding the statistical program can also be omitted. Interestingly, the writer probably assumes that all the

readers are Thai because an abbreviation (M.), commonly known among Thais to refer to 'Matthayom', or 'Mattayom', is used to represent the educational level/grade of the study.

Sample 5

<m2><m3> ***This study was conducted to*** analyze the grammatical errors types in the use of past simple and past continuous tenses by sixty first year students at Silpakorn University. The study aimed to find out the frequent errors and the causes of errors made by English and non-English major students. </m3></m2> <m4> The results of the study revealed that grammatical errors were based on the four types of errors: tense, omission, addition, and misformation. Errors in tense were problematic for both groups of students. The most frequent tense error was past continuous tense produced by the non-English major students, while past simple tense errors were produced less often. </m4> <m5> The findings of this study could be useful for the Thai EFL/ ESL students to enhance their English grammar proficiency in the past simple and past continuous tenses. Furthermore, it could be helpful to teachers for improving their teaching techniques and materials. Moreover, teachers can emphasize the problematic areas of the past simple and past continuous tenses to learn the causes of the students' errors. Knowledge about the occurrence of errors plays a significant role in students' improving their language learning in English and the development of teaching strategies to address these weaknesses. This could lead to more successful language learning in English grammar for Thai EFL/ESL students. </m5>

N_CR_2010_016

Sample 5 has been classified as an example of pattern 2(3)-4-5. The first two sentences reflect a combination of m2 and m3. As in the first sample, the writer of this abstract employed '*to + infinitive*' to explain the objective of the study, saying:

(sentence 1) This study was conducted to (*analyze*) ....

(sentence 2) The study aimed to (*find out*) ....

The writer then describes the findings (m4) and the conclusion in terms of the pedagogy (m5) in the following sentences. By addressing m3 and m2 together and not further explaining the method in another sentence, it seems that more space was available for m4 and m5. As can be seen, it is clear that approximately 50% of the space is devoted to m5 in relation to the implications of the study. It should be noted that the acronym 'EFL/ESL' was used without explaining what it stands for, implying that all the readers are insiders of the same discourse community. In fact, the abstracts could also be slightly

shorter if 'ESL' were omitted because English is regarded as a foreign language, not a second language, for Thai learners.

In addition to the two samples above, further exemplar abstracts can be found in Tables 5.1a-c, where the classification of both move composition and move patterns are illustrated. The following tables (5.5 and 5.6) show the move patterns in the NAC and then the PAC, arranged firstly according to the types of move composition involved, and then by the frequencies of the different move patterns that reflect a particular composition. It should be noted again that the different compositions are numbered (1, 2, 3, and 2, 3, 4, 5) and separated by commas. The patterns associated with each composition are separated by hyphens, as in 1-2-3 and 1-2(3)-3-4-5. The tables presented here record the three most frequent patterns associated with each composition; where there are additional patterns occurring only once or twice, these are grouped together as 'others' (for the full version of the table, recording all patterns separately, see Appendix F).

Table 5.5a Rhetorical variation of the NAC and its subcorpora organized by move compositional types

No.	Move Compositions	CR		TE		NAC	
	Move Patterns	**Freq.**	**%**	**Freq.**	**%**	**Freq.**	**%**
1	**2, 3, 4, 5**	**29**	**38.67**	**28**	**37.33**	**57**	**38.00**
1.1	2(3)-3-4-5	25	33.33	19	25.33	44	29.33
1.2	2(3)-4-5	1	1.33	2	2.67	3	2.00
1.3	2(3)-3-3(4)-4-5	0	0.00	2	2.67	2	1.33
1.4	Others	3	4.00	5	6.67	8	5.33
2	**2, 3, 4**	**23**	**30.67**	**16**	**21.33**	**39**	**26.00**
2.1	2(3)-3-4	18	24.00	13	17.33	31	20.67
2.2	2-3-4	2	2.67	1	1.33	3	2.00
2.3	2(3)-2-3-4	1	1.33	1	1.33	2	1.33
2.4	Others	2	2.67	1	1.33	3	2.00
3	**1, 2, 3, 4, 5**	**16**	**21.33**	**19**	**25.33**	**35**	**23.33**
3.1	1-2(3)-3-4-5	11	14.67	13	17.33	24	16.00
3.2	1-2(3)-2-3-4-5	1	1.33	1	1.33	2	1.33
3.3	Others	4	5.33	5	6.67	9	6.00
4	**1, 2, 3, 4**	**6**	**8.00**	**10**	**13.33**	**16**	**10.67**
4.1	1-2(3)-3-4	5	6.67	8	10.67	13	8.67
4.2	2(3)-1-3-4	1	1.33	1	1.33	2	1.33
4.3	1-3-3(4)-2(3)	0	0.00	1	1.33	1	0.67
5	**1, 2, 3, 5**	**0**	**0.00**	**1**	**1.33**	**1**	**0.67**
5.1	1-2-3-5	0	0.00	1	1.33	1	0.67
6	**2, 3**	**1**	**1.33**	**0**	**0.00**	**1**	**0.67**
6.1	2(3)-3	1	1.33	0	0.00	1	0.67
7	**2, 3, 5**	**0**	**0.00**	**1**	**1.33**	**1**	**0.67**
7.1	2(3)-3-5	0	0.00	1	1.33	1	0.67
Total		**75**	**100**	**75**	**100**	**150**	**100**

As evidenced in Table 5.5, there is not much variation in patterns of abstracts written by the novice writers. What stands out is that their patterns are likely to cluster around the first pattern (1.1, 2.1, 3.1, and 4.1) of certain move compositions. This means that certain prevalent patterns are likely to be written by these novices. It is also noticeable that the patterns grouped as 'others' under each move composition account for less than 6% of the novice abstracts. The three most prevailing move patterns are 2(3)-3-4-5, 2(3)-3-4 and 1-2(3)-3-4-5, with a proportion of 29.33%, 20.67% and 16.00% respectively, representing 66% of the novice abstracts. It is also apparent that words and phrases reflecting m3 are usually embedded in sentences that also reflect m2, with the exception of patterns 2.2 and 5.1. Excerpts from two abstracts where m2 occurs without m3 also represented in the sentence are displayed below.

Excerpt 1

(1) ***The purposes of this study are to*** explore the most common barriers contributing to communication breakdown between THAI flight attendants and foreign passengers, as well as to compare the intercultural barriers in regional routes and intercontinental routes in order to minimize the possible errors during communication. **(2)** One hundred and fifty questionnaires containing both close-ended questions and open-ended questions were used as a research instrument. **(3)** In order to collect the data related to personal information, listening problems, speaking problems and culture problems, questionnaires were randomly distributed to **flight attendants** at all levels at Thai Airways Operation Center. **(4)** Ninety six returned questionnaires were analyzed by SPSS program before presenting the result in the form of statistics.

N_CR_2011_029

In excerpt 1, the phrase '*The purposes of this study are to* …' clearly shows the communicative purpose of the first sentence. Although the subsequent phrase 'THAI flight attendants and foreign passengers' might be taken to refer to the participants of the study, sentence 3 (underlined) clearly reveals that only flight attendants were the respondents of the questionnaires. Given that foreign passengers were not involved, the first sentence should not be regarded as reflecting m2 with m3 embedded because it might mistakenly suggest that foreign passengers were also participants. It is noticeable again here that the statistical program is unnecessarily mentioned in the abstract.

Excerpt 2

(1) ***This study aimed to investigate*** the teacher's error correction methods and to find out the students' attitudes towards the teacher's error correction in written work. **(2)** The subjects of this study were thirty medical freshmen who took the EL317 (Reading and Writing for Academic Purposes) course at the Language Institute, Thammasat University during semester 2/2012. **(3)** The research instruments were a three-part questionnaire used to collect data on the students' attitudes and their last corrected written work in class.</m3>

N_CR_2012_054

In the first sentence of excerpt 2, there is no clear information regarding research design, participants, or research tools: no details are given about the project's data collection. Despite having mentioned *students' attitudes*, there are no specific details about the participants. Therefore, this sentence is coded as m2 without any embedded move. More information about the participants can be found in sentence 2, which mentions the number of participants, their field of study, the course they are undertaking, their affiliation, and the data collection period. The research instruments are mentioned in sentence 3. Sentences 2 and 3 are thus categorized as m3. It is noteworthy that some unnecessary information should be omitted, such as the course code (EL317) and the time and place of the data collection.

As seen in Table 5.5a, the move patterns in the NAC are grouped by move compositional types. The move patterns in the table that follows, however, are organized in terms of frequencies and percentages.

Table 5.5b Frequencies and percentages of move patterns in the NAC

No	Patterns	Freq.	%	No	Patterns	Freq.	%
1	2(3)-3-4-5	44	29.33	18	2(3)-3-4-3(2)-4	1	0.67
2	2(3)-3-4	31	20.67	19	2(3)-3-4-3(4-5)	1	0.67
3	1-2(3)-3-4-5	24	16.00	20	2(3)-3-3(4)-4	1	0.67
4	1-2(3)-3-4	13	8.67	21	2(3)-3-2-3-4-5	1	0.67
5	2(3)-4-5	3	2.00	22	2(3)-3-4(3)-4-5	1	0.67
6	2-3-4	3	2.00	23	2(3)-3	1	0.67
7	2(3)-3-3(4)-4-5	2	1.33	24	2(3)-2-3-4-5	1	0.67
8	2(3)-2-3-4	2	1.33	25	1-3-3(4)-2(3)	1	0.67
9	2(3)-1-3-4	2	1.33	26	1-2-3-5	1	0.67
10	1-2(3)-4-5	2	1.33	27	1-2(3)-5-3-4	1	0.67
11	1-2(3)-2-3-4-5	2	1.33	28	1-2(3)-4(3)-4-5	1	0.67
12	2-3-4-5	1	0.67	29	1-2(3)-3-4-5-4	1	0.67
13	2(3)-4	1	0.67	30	1-2(3)-3-4(5)	1	0.67
14	2(3)-3-5-4	1	0.67	31	1-2(3)-3-3(4)-5	1	0.67
15	2(3)-3-5	1	0.67	32	1-2(3)-3-3(4)-4-5	1	0.67
16	2(3)-3-4-4(5)-5	1	0.67	33	1-2(1-3)-3-4-5	1	0.67
17	2(3)-3-4-4(5)	1	0.67				

As indicated in Table 5.5b, a total of 33 different rhetorical patterns were found in the NAC. The move patterns are arranged in the table according to their frequency of occurrence. The top three frequently found move patterns, together accounting for 66% of the abstracts in the NAC, are 2(3)-3-4-5, 2(3)-3-4, and 1-2(3)-3-4-5. Closer analysis of the table demonstrates that m3 seems most often to occur as an embedded move. Only three exceptions were found: patterns 6, 12, and 26. Some patterns were found only once (patterns 12-33), accounting for 14.67%, or twice (patterns 7-11), representing 6.67%.

5.4.2 Rhetorical move patterns in the PAC

This section looks at the rhetorical move patterns in the PAC and its subcorpora arranged by move compositional types (12 types). A total of 163 move patterns were found in the PAC. Because some variations occurred only a few times in the PAC as a whole, they are grouped together here as 'others'. Only the two or three most frequent variations used by the group and subgroups are listed individually. The complete list of all one 163 variations can be found in Appendix G.

As seen in Table 5.6a, the top three most frequently found patterns are 2(3)-3-4-5, 1-2(3)-3-4-5, and 2(3)-3-4, with a proportion of 14.55%, 6.88% and 6.35% respectively, representing 27.78% of the PAC abstracts overall. It should be restated that the top three patterns in the NAC account for 66%, and thus there is undoubtedly more variation in the PAC than in the NAC. Interestingly, the top three move patterns also reflect the top three compositions (i.e. the most frequent pattern represents the most frequent composition). Despite the various patterns associated with each move composition, the most frequently found structural patterns tend to be the very basic patterns. These basic patterns[8] are found to be arranged in chronological order of m1 to m5, but in many cases m1 and m5 are not included. The basic patterns have only one embedded move, which is usually m3 embedded in m2. Complex structural patterns[9] are found with low

[8] Basic patterns are patterns arranged in chronological order from m1 to m5, though some moves may be omitted. The embedded moves (the combination of two moves) in these basic patterns occur only once. Samples of the basic patterns include 2(3)-3-4-5, 2(3)-3-4, and 1-2(3)-3-4-5.

[9] Complex patterns are patterns arranged in non-chronological order, with a repetition of some moves after an interruption of other moves. Samples of the complex patterns include 1-2(3)-5-3-4, 1-2(3)-4(3)-4-5, etc. There may be two occurrences of embedded moves or the occurrence of the three moves combined, such as 1-3-3(4)-2(3), 1-2(1-3)-3-4-5, etc.

frequencies.

In the case of those compositions that have patterns grouped together as ‘Others’, it is notable that this combined group has a frequency similar to or higher than the individually most frequent patterns; indeed, in some instances it is considerably higher. For example, the proportion of pattern 1.1 relative to those grouped as ‘others’ under the same composition was 14.55% to 15.87%. An even greater difference can be found when comparing pattern 2.1 with those grouped as ‘Others’ (category 2.4): 6.88% for pattern 2.1 versus 13.23% for 2.4.

As can be seen from Table 5.6a, there is greater variation in the PAC than the NAC with respect to the number of patterns that involve embedded moves. It is noticeable, for example, that m4 is occasionally embedded with m3. Such cases where the combination of move 3 and 4 occurs together is found in far fewer proportions in the NAC. Sample 1 below is an example of the non-embedded move structure of 2-3-4-5, which was grouped under the ‘Others’ category in item 1.4. The sample was chosen because in this study move 3 is usually found embedded in move 2, and a non-embedded pattern of moves 2 and 3 rarely occurs. In contrast, Sample 2 comes from an abstract where m4 is embedded by m3 in pattern 1-2(3)-4(3)-5.

Table 5.6a Rhetorical variation of PAC and its subcorpora organized by move composition types

No.	Move compositions	EL		ES		LL		TQ		PAC	
	Move Patterns	Freq	%	Freq	%	Freq	%	Freq	%	Freq	%
1	**2, 3, 4, 5**	**23**	**23.71**	**26**	**32.91**	**56**	**47.86**	**26**	**30.59**	**131**	**34.66**
1.1	2(3)-3-4-5	12	12.37	6	7.59	24	20.51	13	15.29	55	14.55
1.2	2(3)-3-3(4)-5	1	1.03	1	1.27	4	3.42	2	2.35	8	2.12
1.3	2(3)-4-5	0	0.00	4	5.06	3	2.56	1	1.18	8	2.12
1.4	Others	10	10.31	15	18.99	25	21.37	10	11.76	60	15.87
2	**1, 2, 3, 4, 5**	**21**	**21.65**	**28**	**35.44**	**19**	**16.24**	**29**	**34.12**	**97**	**25.66**
2.1	1-2(3)-3-4-5	5	5.15	9	11.39	5	4.27	7	8.24	26	6.88
2.2	1-2(3)-4-5	6	6.19	2	2.53	2	1.71	4	4.71	14	3.70
2.3	1-2(3)-3-3(4)-5	1	1.03	3	3.80	1	0.85	2	2.35	7	1.85
2.4	Others	9	9.28	14	17.72	11	9.40	16	18.82	50	13.23
3	**2, 3, 4**	**8**	**8.25**	**6**	**7.59**	**26**	**22.22**	**8**	**9.41**	**48**	**12.70**
3.1	2(3)-3-4	2	2.06	2	2.53	17	14.53	3	3.53	24	6.35
3.2	2(3)-3(4)-4	1	1.03	1	1.27	0	0.00	1	1.18	3	0.79
3.3	Others	5	5.15	3	3.80	9	7.69	4	4.71	21	5.56
4	**1, 2, 3, 4**	**14**	**14.43**	**8**	**10.13**	**11**	**9.40**	**9**	**10.59**	**42**	**11.11**
4.1	1-2(3)-4	4	4.12	1	1.27	3	2.56	1	1.18	9	2.38
4.2	1-2(3)-3-4	3	3.09	3	3.80	1	0.85	1	1.18	8	2.12
4.3	Others	7	7.22	4	5.06	7	5.98	7	8.24	25	6.61
5	**1, 2, 3, 5**	**8**	**8.25**	**5**	**6.33**	**2**	**1.71**	**8**	**9.41**	**23**	**6.08**
5.1	1-2(3)-5	4	4.12	1	1.27	0	0.00	2	2.35	7	1.85
5.2	1-2(3)-3-5	1	1.03	2	2.53	1	0.85	2	2.35	6	1.59
5.3	Others	3	3.09	2	2.53	1	0.85	4	4.71	10	2.65
6	**1, 2, 3**	**12**	**12.37**	**1**	**1.27**	**0**	**0.00**	**0**	**0.00**	**13**	**3.44**
6.1	1-2(3)	6	6.19	0	0.00	0	0.00	0	0.00	6	1.59
6.2	Others	6	6.19	1	1.27	0	0.00	0	0.00	7	1.85
7	**2, 3, 5**	**6**	**6.19**	**2**	**2.53**	**1**	**0.85**	**2**	**2.35**	**11**	**2.91**
7.1	2(3)-3-5	2	2.06	1	1.27	1	0.85	1	1.18	5	1.32
7.2	Others	4	4.12	1	1.27	0	0.00	1	1.18	6	1.59
8	**2, 3**	**3**	**3.09**	**1**	**1.27**	**0**	**0.00**	**1**	**1.18**	**5**	**1.32**
8.1	2(3)	1	1.03	0	0.00	0	0.00	0	0.00	1	0.26
8.2	2(3)-2	1	1.03	0	0.00	0	0.00	0	0.00	1	0.26
8.3	2(3)-3	0	0.00	1	1.27	0	0.00	0	0.00	1	0.26
8.4	3-3(2)-3	1	1.03	0	0.00	0	0.00	0	0.00	1	0.26
8.5	2(3)-3-2-3	0	0.00	0	0.00	0	0.00	1	1.18	1	0.26
9	**1, 3, 4, 5**	**1**	**1.03**	**0**	**0.00**	**2**	**1.71**	**2**	**2.35**	**5**	**1.32**
9.1	1-3-4-5	0	0.00	0	0.00	2	1.71	1	1.18	3	0.79
9.2	1-4-3-4(3)-5	1	1.03	0	0.00	0	0.00	0	0.00	1	0.26
9.3	3(1)-3-5-4	0	0.00	0	0.00	0	0.00	1	1.18	1	0.26
10	**1, 3, 5**	**1**	**1.03**	**0**	**0.00**	**0**	**0.00**	**0**	**0.00**	**1**	**0.26**
10.1	1-3-5	1	1.03	0	0.00	0	0.00	0	0.00	1	0.26
11	**1, 3, 4**	**0**	**0.00**	**1**	**1.27**	**0**	**0.00**	**0**	**0.00**	**1**	**0.26**
11.1	1-3(1)-3-4	0	0.00	1	1.27	0	0.00	0	0.00	1	0.26
12	**3, 4**	**0**	**0.00**	**1**	**1.27**	**0**	**0.00**	**0**	**0.00**	**1**	**0.26**
12.1	3-4	0	0.00	1	1.27	0	0.00	0	0.00	1	0.26
Total		97	**100**	79	**100**	117	**100**	85	**100**	378	**100**

Sample 6

> **(1)** ***This article uncovers*** the complexities and contradictions inherent in making decisions about L1 use in the English language classroom. **(2)** Through an analysis of data from classrooms in a Cypriot context and from interviews with Cypriot teachers, a number of functions for L1 use are identified, as are the teachers' rationales for using L1 for different functions. **(3)** Teachers' decision making, ***it emerges***, is often complex, based on either what they perceive as their students' affective needs or on their cognitive processes. **(4)** ***What is more***, teachers often under-report or differently report their use of L1 in the classroom, contradicting beliefs by their actions. **(5)** The construct of guilt is offered to explain these complexities and contradictions in the teachers' use of L1 in this study. **(6)** ***We conclude by suggesting that*** teachers should be supported in finding local solutions to local teaching problems, so that they better understand and exploit the resources available to them.
>
> *P_EL_2011_043*

Sample 6 has been categorized as pattern 2-3-4-5. The first sentence tells the readers the purpose of the study by saying, '***This article uncovers*** …'. Sentence 2 touches on the methodology by using the prepositional phrase '***through*** *an analysis of data from classroom ... and* ***from*** *interviews with* …'. Sentence 3 mentions the findings, signaled by the use of the phrase '***it emerges***'. The writers also used the expression, '***What is more, …***' to introduce some additional results in sentences 4 and 5. These three sentences relating to findings are thus classified together as m4. The last sentence, which is classified as m5, concludes by giving some pedagogical implications. This was accomplished by beginning the sentence with, '***We conclude by suggesting that***…'.

Sample 7

(1) Since 2001, the Education Bureau (EDB) in Hong Kong has been promoting a shift from traditional assessment of learning to assessment for learning, where classroom-based assessment is linked to teaching and learning, with students taking an active role in the assessment process. **(2)** In particular, secondary school students are encouraged to assess their own and peers' oral English through self- and peer assessment. **(3)** ***While peer assessment has been recognized as enhancing*** student learning if sensitively implemented, it is a new concept to many local students. **(4)** ***This paper reports on*** a case study that investigated students' perceptions of the implementation of peer assessment in their English speaking classes. **(5)** Through interview and classroom observation, ***the study reveals that*** despite the benefits that the participants perceived, they had serious concerns over the new assessment. **(6)** What these 'average/weak' students reported was sad but inspiring. **(7)** Implications are identified for those who plan to conduct peer assessment.

P_EL_2011_039

Sample 7 has been categorized as pattern 1-2(3)-4(3)-5. The first three sentences of the abstract describe the importance of peer assessment and why it was necessary to conduct the research by saying, '***While peer assessment has been recognized as enhancing*** …, it is a new concept to many local students'. Therefore, these three sentences are classified as m1. Sentence 4 is categorized as m3 embedded in m2 because the writer's intention is to inform the readers about the purpose of the study by saying, '***This paper reports on*** …'. However, the research design (a case study), which is regarded as m3, is also mentioned in the same sentence. Sentence 5 is classified as m4 embedded by m3. By saying '***the study reveals that*** ...', the writer presents the findings of the study (m4), which is the main communicative purpose of the sentence. However, the writer also uses the prepositional phrase '***through interview and classroom observation***' at the beginning to tell the readers something about the methodology (m3). Sentence 6 includes the interpretation of the writer, while sentence 7 states the implications. These two sentences are thus classified as m5. It is worth noting that this sample substantially differs from the previous three samples (Samples 4-5 of NAC and Sample 6 of PAC) in that more than 50% of the space is devoted to the introduction part. This reflects Hyland's (2000) argument that abstracts are regarded as a form of promotional discourse.

Table 5.6b Frequencies and percentages of move patterns in the PAC

No	Patterns	Freq.	%	No	Patterns	Freq.	%
1	2(3)-3-4-5	55	14.55	7	2(3)-4-5	8	2.12
2	1-2(3)-3-4-5	26	6.88	8	1-2(3)-3-4	8	2.12
3	2(3)-3-4	24	6.35	9	1-2(3)-5	7	1.85
4	1-2(3)-4-5	14	3.70	10	1-2(3)-3-3(4)-5	6	1.59
5	1-2(3)-4	9	2.38	11	1-2(3)-3-5	6	1.59
6	2(3)-3-3(4)-5	8	2.12	12	1-2(3)	6	1.59

***Note**: Only patterns occurring six or more times are included, i.e. 12 of the 163 move patterns found in PAC (see Appendix H for the complete list of all the move patterns)*

Table 5.6b shows the move patterns of the PAC arranged according to frequency. Overall, there are 163 move patterns in the PAC, but only the 12 patterns which occur at least six times are listed here. The three most frequently found patterns account for 27.78%, or just over one quarter, of the whole PAC. This contrasts noticeably with the NAC, where the top three patterns represent 66%, or nearly two-thirds, of the novice abstracts. To conclude, there is a greater variation found in the PAC than in the NAC. In addition, many of the patterns found in the PAC occur only once or twice, suggesting idiosyncratic patterns found based on the corpus. The findings may suggest that many professional researchers do not conform to conventional patterns, and as a group the variation is obvious, rather than denoting that each professional writer uses a variety of patterns in their abstract writing.

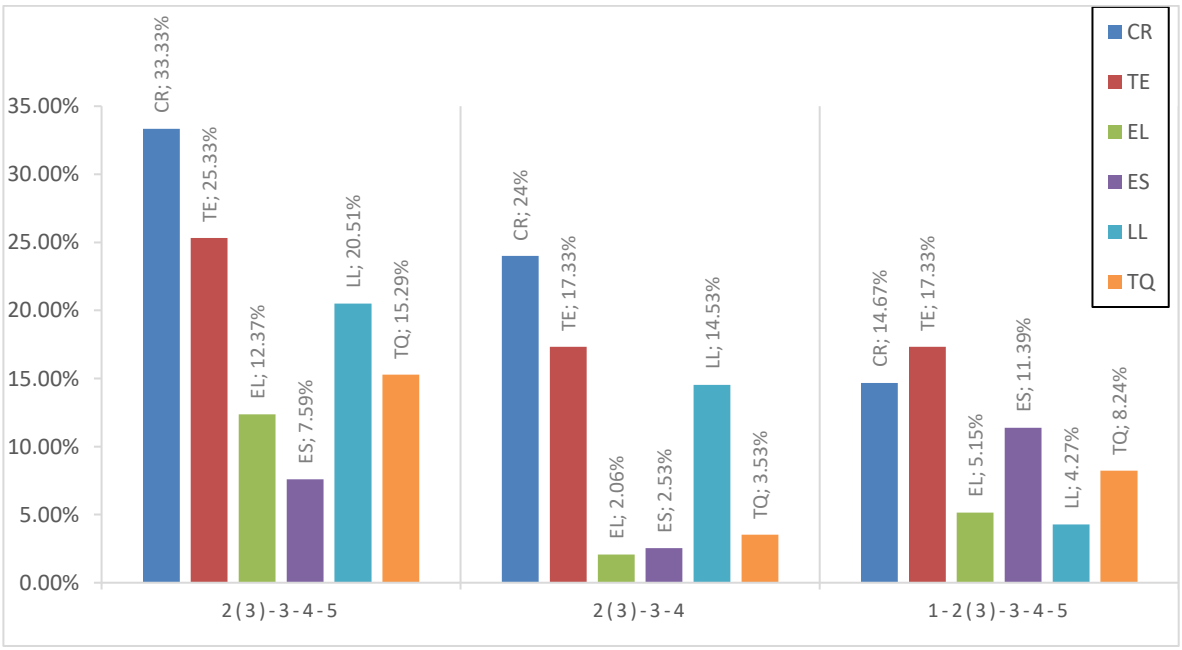

Figure 5.1 Bar chart of the most frequently found move patterns in each subcorpora

The top three frequently found move patterns of the NAC and the PAC are illustrated in Figure 5.1. Pattern 2(3)-3-4-5 are the most often found among CR, TE, EL, LL, TQ groups. Among these three patterns, the occurrence of pattern 1-2(3)-3-4-5 is the most commonly found in ES. It is noticeable that patterns 2(3)-3-4 and 1-2(3)-3-4-5 occur in TE group with the same proportion (17.33%). Among these top three patterns, patterns 2(3)-3-4 are the least frequently found in EL, ES and TQ abstracts.

5.5 Optional, Conventional, and Obligatory Moves

This section sets out an approach to categorizing moves in the NAC and the PAC abstracts as optional, conventional or obligatory, based on their frequency of occurrence. First, the criteria for distinguishing between the three categories will be established. After that, the frequencies and percentages of the moves used by the novice and the professional groups are presented in Table 5.8a, together with the corresponding categorization. Further details regarding the frequencies and percentages of the moves in the various subcorpora of the NAC and the PAC can be found in Table 5.8b, which provides fine-grained details of the similarities and differences between the subgroups of each corpus.

The criteria used to judge whether the moves are optional, conventional or obligatory were taken from Kanoksilapatham (2005, p. 272), Connor, Upton and Kanoksilapatham (2007, pp. 39-40) and Saeeaw and Tangkiengsirisin (2014, p. 83), as summarized in Table 5.7.

Table 5.7 Criteria for classifying moves as optional, conventional or obligatory

Author	Genre	Optional	Conventional	Obligatory
Kanoksilapatham (2005) and Connor, Upton and Kanoksilapatham (2007)	Research articles	< 60%	≥ 60%	97-100%
Saeeaw and Tangkiengsirisin (2014)	Abstracts	< 60%	≥ 60%	100%

Kanoksilapatham (2005) set her own criteria to analyze 60 biochemistry texts published in five leading peer-reviewed journals in the field. Her move analysis was based on Swales' (1990) analytical framework. In explaining how she established the criteria,

Kanoksilapatham suggests that a cut-off frequency of 60% for regarding moves as conventional, which is probably a little high, but that it is nevertheless beneficial to distinguish the conventional from those recognized as optional, which fall below 60%. Though not stated explicitly that at least 97% should be the cut-off point for obligatory move, her research team notes in a book chapter that 'the third move type can be considered obligatory, as it occurred in over 97% of the texts' (Connor, Upton and Kanoksilapatham 2007, p. 39).

The criteria used in Saeew and Tangkiengsirisin's (2014) study follow Kanoksilapatham (2005), with the difference that 100% is used as the cut-off point for moves to be classified as obligatory. Their study compares 200 abstracts in two different disciplines: Environmental Science and Applied Linguistics. One hundred research article abstracts from each field were selected from the top four journals with highest impact factors in their field. Hyland's (2000) framework was adopted for the analysis.

Considering the two criteria above, the difference in judging the obligatory move of Kanoksilapatham reflects the possibility that might occur when analyzing a large number of texts (NAC = 150, PAC = 358). It could be assumed that because of their inexperience in abstract writing, some novice writers might not be aware of the typical conventions, and therefore do not follow them. Kanoksilapatham's criteria will be adopted for the analyses in the present study.

Table 5.8a Frequency of moves 1-5 in the NAC and the PAC

Move	**NAC**			**PAC**		
	Freq.	**%**	**Status**	**Freq.**	**%**	**Status**
m1	52	34.67	Optional	183	48.41	Optional
m2	150	100.0	Obligatory	370	97.88	Obligatory
m3	150	100.0	Obligatory	378	100.00	Obligatory
m4	147	98.00	Obligatory	325	85.98	Conventional
m5	94	62.67	Conventional	268	70.90	Conventional
Total	**150**	**100.00**		**378**	**100.00**	

According to the criteria taken from Kanoksilapatham (2005, p. 272) and Connor *et al.* (2007, pp. 39-40), Table 5.8a demonstrates that in the corpus of abstracts produced by novice writers, m1 can be considered as optional (<60%), m2 and m3 as obligatory (100%), and m5 as conventional (60-97%). M4 was regarded as obligatory (98%). Similar results were found in relation to the corpus of abstracts produced by professional

writers, with the exception of m4. While m4 is classified as conventional in the abstracts written by the professionals (appearing in 85.98%), it is considered as obligatory in those written by the novices (appearing in 98%). In sum, three obligatory moves, one conventional move and one optional move are found in the NAC. In contrast, two obligatory moves, two conventional moves and one optional move are found in the PAC. This might contribute to the longer average length of the NAC than the PAC (NAC =233 and PAC =177.25). Considering the submission guidelines provided by the journals (see Section 4.4.3), it is worth noting that each publication has a different length criterion, ranging from 150 in EL and LL to 200 in TQ and ES. The minimum requirement is provided only in ES. The length suggested in the journal guidelines should not be taken lightly. This is different for novice researchers, who are given more flexible guidelines of one-page long.

Table 5.8b Frequency of moves 1- 5 in the subcorpora of the NAC and the PAC

Move	NAC				PAC							
	CR*		TE*		EL*		ES*		LL*		TQ*	
	Freq	%	Freq	%	Freq	%	Freq	%	Freq	%	Freq	%
m1	22	29.33	30	40.00	58	59.79	43	54.43	34	29.06	48	56.47
m2	75	100.00	75	100.00	95	97.94	77	97.47	115	98.29	83	97.65
m3	75	100.00	75	100.00	97	100.00	79	100.00	117	100.00	85	100.00
m4	74	98.67	73	97.33	67	69.07	70	88.61	114	97.44	74	87.06
m5	45	60.00	49	65.33	60	61.86	61	77.22	80	68.38	67	78.82
Total	**75**	**100**	**75**	**100**	**97**	**100**	**79**	**100**	**117**	**100**	**85**	**100**

NAC (CR = English for Careers Group, TE* = Teaching English as a Foreign Language Group)*
PAC (EL = ELT Journal, ES* = English for Specific Purposes Journal, LL* = Language Learning Journal, TQ* = TESOL Quarterly Journal)*

Table 5.8b sets out the frequencies of moves 1-5 in each of the subcorpora of the NAC and the PAC. With regard to the NAC, the most noticeable difference between the CR and TE abstracts relates to the frequency of m1. Approximately 10% difference between the CR and TE abstracts is found in the occurrence of m1 (CR = 29.33%, TE = 40%). When it comes to the PAC, m1 is less frequent in LL abstracts than in those of the other three journals (EL, ES and TQ). While the professional writers tend to incorporate m2 and m3 with a similar degree of frequency across the four different sources, there is clearly variability in the representation of m4 and m5 among subgroups. Figure 5.2 clearly depicts the similarities and differences among these subgroups.

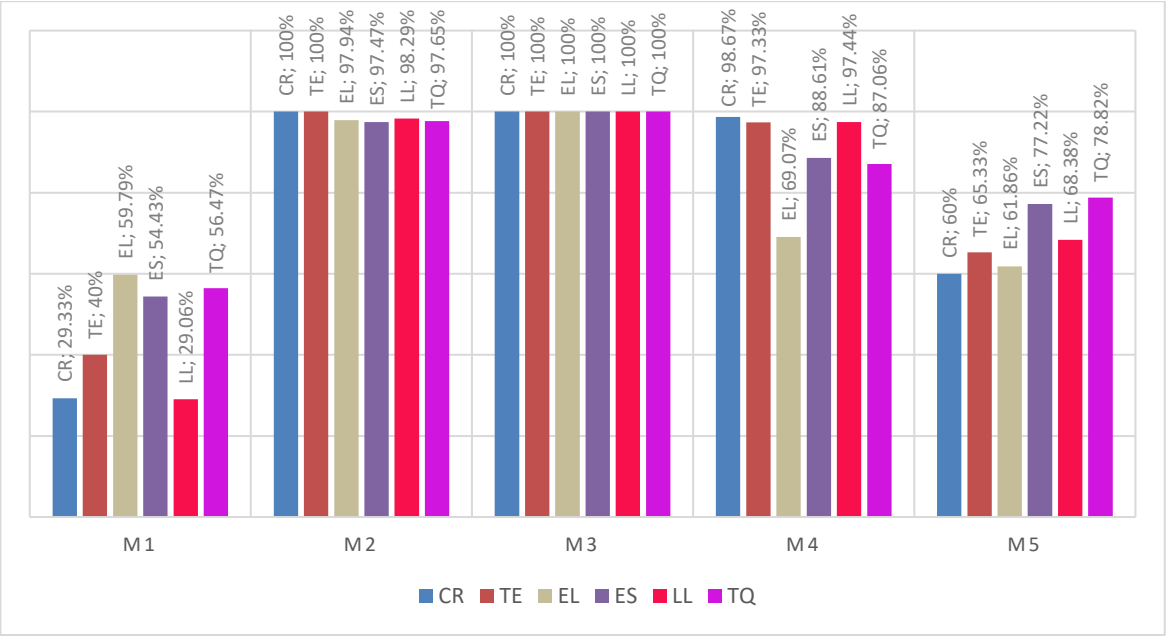

Figure 5.2 Bar chart of moves in the subcorpora of the NAC and the PAC

A closer examination of the frequencies related to the NAC subcorpora shows two noticeable differences between them. The first difference is related to the use of m1 in the subgroups of both the NAC and the PAC. As illustrated in Figure 5.2, m1 is more frequent in TE abstracts than in CR abstracts, with a difference of approximately 10% between the two groups (40%:29.33%). Differences in the frequency of m1 are also found in the subgroups of the PAC. While similar proportions of m1 (54.43 - 59.79%) are found in EL, ES, and TQ abstracts, a considerably lower proportion of m1 is found in LL abstracts (29.06%).

Another difference can be seen when comparing the frequencies of m4 and m5 across the NAC and the PAC and their various subcorpora. Although both the CR and TE abstracts of the NAC contain a higher percentage of m4 when compared with m5 (CR = 38.67% and TE = 32% differences), only LL abstracts in the PAC show a similar tendency (29.06% difference). However, it should be noted that when comparing the usage of m4 and m5 of the other PAC subcorpora (EL, ES and TQ), approximately 10% difference are found.

5.6 Summary

This chapter has addressed research question one. By using Hyland's (2000) move analysis framework, the similarities and differences found in abstracts written by novice

and professional writers have been analyzed from three perspectives: move compositions, move structural patterns and the consideration of whether the different kinds of moves are optional, conventional or obligatory. With regard to the move compositions, the findings show that there are seven compositional types of abstracts found in the NAC, namely {2, 3, 4, 5}, {2, 3, ,4}, {1, 2, 3, 4, 5}, {1, 2, 3, 4}, {1, 2, 3, 5} {2, 3, 5}and {2, 3}, while twelve types are found in the PAC with the seven types similar to NAC. The other five compositional types include {1, 2, 3}, {1, 3, 4, 5}, {1,3, 5}, {1, 3, 4}, and {3, 4}.

With reference to the move patterns, it has been found that a greater number of patterns associated with the various move compositions occur in the PAC than in the NAC. Overall, a total of 33 patterns has been found in the NAC, compared to 163 patterns in the PAC. Despite a large number of patterns found in both the NAC and the PAC, it was found that the basic patterns, rather than the complex patterns, are much more frequent in both groups. The tendency to employ the basic patterns in both groups is demonstrated in the high percentages clustering around the top three or four most frequently found patterns, while the remaining patterns have much lower frequencies, constituting fewer than 10% of the NAC abstracts and fewer than 5% of the PAC abstracts. The greater variety of patterns shown in the PAC reflects how creative and skillful professional writers are. Many of the patterns in the PAC have been found to be idiosyncratic, having only one occurrence.

Having adopted Kanoksilapatham's (2005) and Connor *et al.*'s (2007) criteria for categorizing the moves into three types (optional, conventional and obligatory), the most noticeable difference between the two groups has been shown to relate to m4. While m4 can be classified as an obligatory feature of the NAC abstracts, it is only conventional in the PAC. The findings show that the occurrences of m2 and m3 (97.88% and 100%) are frequently found in the PAC, and these moves are categorized as obligatory. Compared with professional writers, who should strictly follow the publication guidelines, novice writers, having more lenient submission guidelines, tend to produce longer abstracts with a greater variety of move inclusion as obligatory (m2, m3 and m4). As mentioned in Chapter 4, novice writers are informed only about the approximate length of their abstract (1 page), whereas journals suggest to professional writers that they follow the submission guidelines. However, although not all writers follow the guidelines, their articles are

nevertheless accepted for publication. Some variations in move categories are found among the subgroups of the corpora. One obvious difference relates to the frequency of m1. In comparison with other subgroups, m1 is less likely to appear in CR and LL, with frequencies of around 30%.

Chapter 6. Word Frequency Lists and Keyword Lists

6.1 Introduction

Adopting Hyland's (2000) analytical framework, the analyses in the previous chapter have shown how the move composition types and move patterns of the Novice Abstract Corpus (NAC) differ from those of the Professional Abstract Corpus (PAC).

A number of previous studies (e.g. Samraj, 2005; Kanoksilapatham, 2007), which have placed more emphasis on the analyses of move structures and organization of the text, tend to limit the number of texts under investigation to allow more time for performing the labor-intensive qualitative process. This chapter demonstrates that a corpus-based approach, when combined with the top-down qualitative approach, offers more insights into the lexical words and phrases used in each move. Compared with using solely the move analysis approach, the corpus-based method makes it possible to analyze a large amount of data at once, allowing for empirical evidence to be compared and displayed within the context of real language usage.

In the sections that follow, frequency lists and keyword lists extracted from the NAC and the PAC corpora are analyzed using a corpus-based approach in order to answer the following questions:

- **(RQ2)** To what extent are the most frequent lexical words in the NAC similar to those in the PAC?
- **(RQ3)** What keywords are associated with each move in (a) the NAC and (b) the PAC?

Section 6.2 presents lists of the twenty most frequent words in the NAC and the PAC, along with their subcorpora. A comparison of the NAC and the PAC word frequency lists highlights similarities and differences between the two corpora. Section 6.3 offers in-depth analyses of some lexical words by looking at samples from concordance lines. Section 6.4 focuses on the keyword lists, generated and arranged according to moves (m1 = Introduction, m2 = Purpose, m3 = Method, m4 = Product, and m5 = Conclusion) so as to specify the keywords that are characteristics of each move. These keyword lists are divided into: (1) The NAC keyword lists in relation to zero occurrence in the PAC and

(2) The PAC keyword lists in relation to zero occurrence in the NAC. Section 6.5 summarizes the major findings presented in this chapter.

6.2 Wordlists of the NAC and the PAC

Before keyword retrieval, word frequency lists were generated for the NAC and the PAC corpora. Tables 6.1 and 6.2 show that, when all words are included, function words unsurprisingly dominate the frequency lists. The function words include articles, pronouns, conjunctions and auxiliary verbs such as *the, of, and, in, to, a,* and *were*. Since the focus in this study is on lexical words, these function words were subsequently excluded from the frequency lists in order more easily to identify which content words were prevalent in the two datasets (see Tables 6.3 and 6.4).

In order to compare two or more corpora of markedly different sizes, normalized frequencies are needed. The selection of a common base should be taken with care to avoid a numberical distortion which might affect the subsequent calculations of statistical significance. In terms of confidence level determination, the more samples that are tested, the greater the precision of the confidence level. Kilgarriff (2005, p. 268) thus asserts in his experiment that "[g]iven enough data, H_0 is almost always rejected". Given the size of the NAC and the PAC subcorpora (33,130 words and 61,117 words respectively), a common base of 10,000 was therefore used in calculating the normalized frequencies.

6.2.1 Wordlists of the NAC and the PAC: All words (function and lexical)

Tables 6.1 and 6.2 show the raw and normalized frequencies of the twenty most frequent words (including function words) in the NAC and the PAC, broken down also by their subcorpora. Only five lexical words (highlighted in gray) appear in the list for the NAC as a whole: *English*, *students*, *study*, *reading* and *language*. Two additional lexical words appear in the separate subcorpora lists (*Thai* in CR and *learning* in TE), but these are not among the twenty most frequent for the NAC as a whole.

Table 6.1 Twenty most frequent words in the NAC and its subcorpora

NAC = Novice Abstract Corpus, consisting of two subcorpora

(CR = English for Careers Program, TE = Teaching English as a Foreign Language Program)

Rank	CR			TE			NAC		
	Wordlist	Raw freq.	Norm freq. (per 10,000)	Wordlist	Raw freq.	Norm freq. (per 10,000)	Wordlist	Raw freq.	Norm freq. (per 10,000)
1	the	1367	831.56	the	1291	773.29	the	2658	802.20
2	of	620	377.15	of	605	362.38	of	1225	369.71
3	and	551	335.18	and	575	344.41	and	1126	339.83
4	in	500	304.15	in	486	291.11	in	986	297.58
5	to	470	285.91	to	466	279.13	to	936	282.49
6	english	322	195.88	english	312	186.88	english	634	191.34
7	study	216	131.39	students	311	186.28	students	498	150.30
8	a	209	127.14	a	226	135.37	a	435	131.29
9	were	191	116.19	were	205	122.79	study	412	124.34
10	students	187	113.75	was	203	121.59	were	396	119.51
11	was	186	113.15	that	201	120.40	was	389	117.40
12	that	179	108.89	study	196	117.40	that	380	114.69
13	this	159	96.72	for	153	91.64	for	305	92.05
14	for	152	92.46	this	145	86.85	this	304	91.75
15	their	120	73.00	their	143	85.65	their	263	79.37
16	as	118	71.78	by	135	80.86	as	243	73.34
17	on	115	69.96	reading	132	79.07	by	230	69.42
18	with	113	68.74	language	127	76.07	reading	218	65.79
19	from	100	60.83	as	125	74.87	on	213	64.28
20	thai	99	60.22	learning	105	62.89	language	206	62.17

Table 6.2 Twenty most frequent words in the PAC and its subcorpora

PAC = Professional Abstract Corpus, consisting of four subcorpora
(EL = English Language Teaching Journal, ES = English for Specific Purposes Journal, LL = Language Learning Journal, and TQ = TESOL Quarterly Journal)

Rank	EL			ES			LL			TQ			PAC		
	Wordlist	Raw freq.	Norm freq. (per 10,000)	Wordlist	Raw freq.	Norm freq. (per 10,000)	Wordlist	Raw freq.	Norm freq. (per 10,000)	Wordlist	Raw freq.	Norm freq. (per 10,000)	Wordlist	Raw freq.	Norm freq. (per 10,000)
1	the	719	556.42	the	987	685.65	the	948	510.28	the	879	579.36	the	3533	578.54
2	of	536	414.80	of	662	459.88	of	786	423.08	of	652	429.74	of	2636	431.66
3	and	465	359.85	and	543	377.21	and	664	357.41	and	552	363.17	and	2224	364.03
4	in	400	309.55	in	448	311.22	in	535	287.98	in	438	286.71	in	1821	297.71
5	to	350	270.86	to	348	241.75	to	381	205.08	to	354	232.01	to	1433	234.33
6	a	280	215.91	a	253	173.67	a	354	190.55	a	310	203.66	a	1197	195.20
7	that	181	140.07	that	161	111.84	language	238	128.11	language	190	125.23	that	749	122.49
8	this	151	116.85	this	150	104.20	I	229	123.26	that	189	124.57	language	618	101.20
9	for	140	109.12	english	148	102.81	that	217	116.80	as	181	118.64	for	573	93.67
10	their	130	100.60	for	137	95.17	learners	155	83.43	english	166	109.41	this	550	89.90
11	on	124	95.96	on	132	91.70	on	154	82.89	for	143	93.59	on	538	88.10
12	language	115	89.00	is	106	73.64	for	152	81.82	this	131	85.68	english	522	85.48
13	students	112	86.67	students	104	72.25	were	139	74.82	on	128	84.37	as	498	81.39
14	teachers	106	82.03	are	99	68.77	with	138	74.28	students	113	73.16	study	409	66.81
15	as	98	75.84	as	97	67.38	english	128	68.90	study	109	71.18	their	389	63.21
16	study	92	71.20	their	92	63.91	as	122	65.67	their	104	66.57	students	386	62.88
17	is	90	69.65	research	91	63.22	study	121	65.13	were	99	64.59	with	385	62.72
18	are	86	66.55	by	90	62.52	this	118	63.52	with	95	61.30	is	367	60.10
19	learners	81	62.68	study	87	60.44	learning	114	61.36	teachers	91	59.98	were	362	59.12
20	english	80	61.91	with	81	56.27	was	109	58.67	is	87	57.34	I	352	57.64

Among the four subcorpora of the PAC in Table 6.2, only two lexical words (*English* and *study*) appear in the wordlists for each subcorpus. Overall, only five lexical words (*language*, *English*, *study*, *students*, and *l*) appear in the top twenty for the combined PAC wordlist. Comparing the NAC wordlist with the PAC wordlist, it was found that four lexical words (*English*, *students*, *study* and *language*) appear in both, with only one different lexical word in each of the lists—the word *reading* in the NAC and the letter *l* in the PAC. It should be noted that these two words were in fact used by both novice and professional writers, but not often enough to appear among the twenty most frequent words. However, this raises further questions as to whether the exclusion of the function words would reveal more similarities (or perhaps more differences) of lexical words from the two datasets; if some specific words are found in both lists, to what extent the novices used the word differently from the professionals is shown by an effect size measure (see Section 6.2.2 for more detail).

Although the normalized frequencies allow us to make comparisons between the two corpora to judge whether a given word is more frequent in the NAC or the PAC, they do not on their own indicate clearly how large any differences are or whether they are statistically significant. Two statistical measures (log-likelihood, hereafter '*log-L*', and log-ratio, hereafter '*log-R'*) were therefore adopted to allow for more in-depth analyses, presented in Section 6.2.2. These two measures served different purposes. Log-L - a keyness statistic - was employed to determine whether the difference in frequency of a given word in the NAC and the PAC is due to chance. The cut-off point for this measure was set at 6.63, corresponding to p = .01, which is commonly used as the minimum confidence level acceptable in corpus studies (Paquot and Bestgen, 2009). The rationale for regarding log-L as a more accurate and preferable significance measure than chi-squared was discussed in Chapter 4 (Section 4.8.4).

In recent decades, the reliability of research findings based merely on null-hypothesis significance testing has become a controversial issue. Many researchers (e.g., Gabrielatos and Marchi, 2012; Gries, 2005; Hardie, 2014) have echoed Kilgarriff's (2005) argument that calculations based merely on *p-value* or statistical probability are debatable, especially when dealing with a large amount of data, where results tend to be statistically significant, although differences may be very subtle. Conversely, with a small amount of data, findings might not be statistically significant despite a high degree of difference. In

order to determine the degree of difference in the frequency with which a specific word is used in corpus A in relation to corpus B, an effect size measure should be used. The second statistical measure mentioned above, log-R, is one such effect size measure. The closer the log-R is to zero, the more similar word usage is between the two corpora (see Section 4.7.3 for more details regarding log-R interpretation).

Despite AntConc's (3.5.2) integration of log-likelihood (log-L) and log-ratio (log-R) functions in the latest version, these statistical results occur only with keyword lists, not frequency lists. With the frequency list function, the words, along with their ranking and raw frequency, are shown in Figure 6.1. However, log-R (the effect size statistics) offers a fine-grained description of the degree of similarities and differences among the words in the two lists, while log-L (the keyness value) provides information as to whether the difference occurs by chance. Arguably, the two statistics can be applied to examine whether there are any similarities in word usage in the frequency lists of any two corpora, rather than placing an emphasis on the degree of differences in the keyword list.

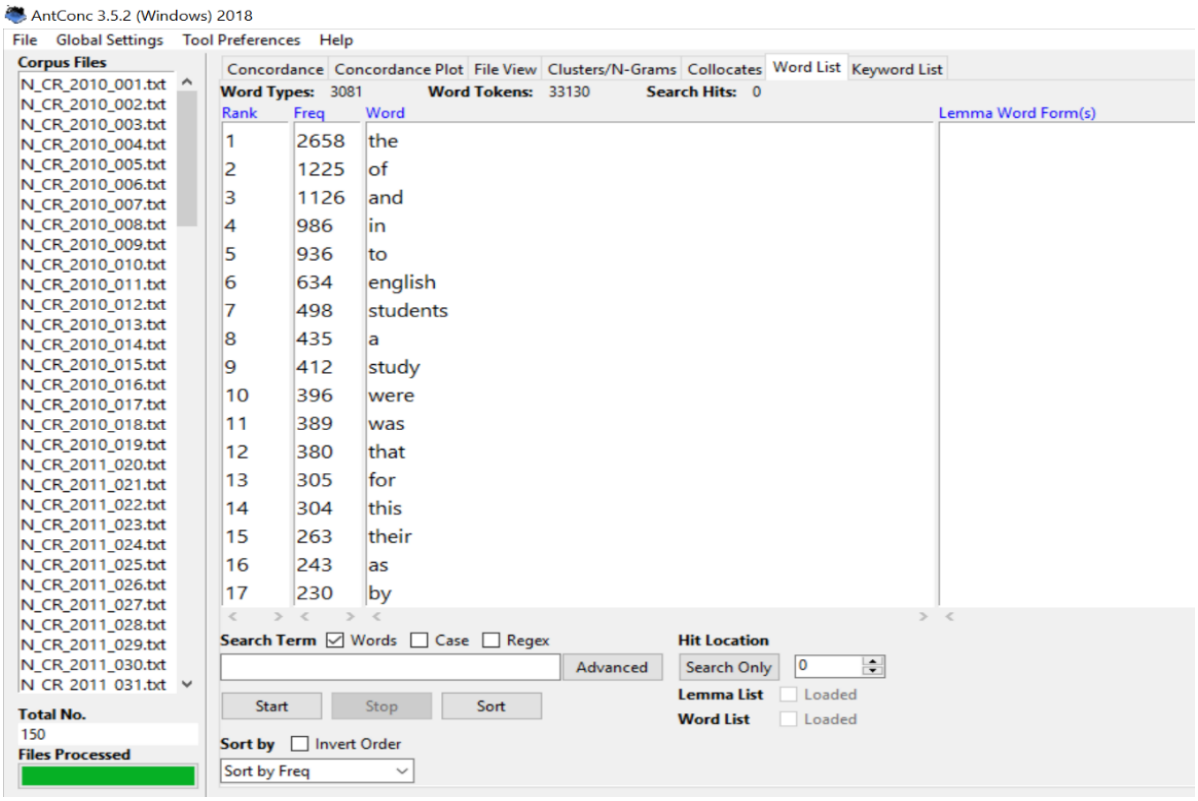

Figure 6.1 Wordlist function display

6.2.2 Wordlists of the NAC and the PAC: Exclusion of function words

This section addresses the question:

- **(RQ2)** To what extent are the most frequent lexical words in the NAC similar to those in the PAC?

Table 6.3 shows the ranking, raw frequency, keyness value (log-L), and effect size value (log-R) of the top twenty lexical words in the NAC wordlist relative to the PAC wordlist. The words are arranged according to their ranking and frequency in the NAC. Those highlighted in gray are the twenty most frequent words found in both datasets.

Table 6.3 Twenty most frequent lexical words in the NAC in relation to the PAC with their rank, raw frequency, log-likelihood (log-L) and log-ratio (log-R)

Wordlist	NAC		PAC		Log-L	Log-R
	Rank	Raw freq.	Rank	Raw freq.	(Keyness)	(effect size for NAC)
english	6	634	12	522	186.17	1.16
students	7	498	16	386	164.42	1.25
study	9	412	14	409	77.64	0.89
reading	18	218	51	132	204.67	3.05
language	20	206	8	618	39.36	-0.70
thai	22	197	-	-	411.92	9.51
used	23	175	68	100	92.03	1.69
learning	27	152	25	265	0.31*	0.08
strategies	28	151	203	42	149.90	2.73
data	30	140	50	133	29.67	0.96
research	31	137	28	226	1.06*	0.16
skills	33	128	142	56	90.62	2.08
results	35	124	32	191	2.41*	0.26
using	36	124	87	86	49.57	1.41
use	37	121	34	182	2.98	0.29
teachers	39	118	29	223	0.05*	-0.03
questionnaire	40	114	273	32	112.54	2.72
communication	41	112	134	58	234.19	8.69
program	42	104	542	17	217.46	8.58
findings	43	99	42	160	1.06*	0.19

Asterisk (*) = keyness value below the cut-off point which was set as ≥ 6.63 or p ≤ 0.01
Minus (-) in Log-R = underuse of the word in NAC in relation to PAC

As shown in Table 6.3, log-R values of the following words - *learning* = 0.08, *research* = 0.16, *results* = 0.26, *teachers* = -0.03 and *findings* = 0.19 - reveal small differences in frequency between the NAC and the PAC. The subtle differences are also backed up by log-L values, which prove to be statistically insignificant (see log-L values with asterisks). In other words, the novice and the professional writers employed these words in their abstract writing with a somewhat similar degree of frequency.

The three words with the greatest degree of difference in their frequencies in the two corpora are *Thai*, *communication*, and *program* (log-R = 9.51, 8.69, and 8.58 respectively). The word *Thai* in particular is not found in any abstract written by the professionals, indicating the focus of the novice corpus, which relates specifically to *Thai* (of which the majority are used as an adjective to indicate nationality or language), and at the same time showing the lack of research studies associated with *Thai* in the PAC. Other words that are found to be used with some degree of difference (but having log-R values less than 8) in the two corpora are *reading*, *strategies*, and *questionnaire*. This suggests that many of the research studies undertaken by novice researchers focus on *communication* and *strategies*. It also implies that these novices tend to use *questionnaires* as a tool for data collection. Among these six words, *reading* also appears in the list of twenty most frequent PAC words, suggesting that both novice and professional writers' research involved *reading* to a certain extent, though with a different degree of indicating the word in their abstract writing.

Table 6.4 Twenty most frequent lexical words in the PAC in relation to the NAC with raw frequency, log-likelihood (log-L) and log-ratio (log-R)

Wordlist	PAC		NAC		Log-L	Log-R
	Rank	Raw freq.	Rank	Raw freq.	(Keyness)	(effect size for PAC)
language	8	618	20	206	39.36	0.70
english	12	522	6	634	186.17	-1.16
study	14	409	9	412	77.64	-0.98
students	16	386	7	498	164.42	-1.25
I	20	352	216	22	183.58	3.12
learners	23	313	68	69	54.54	1.30
learning	25	265	27	152	0.31*	-0.08
research	28	226	31	137	1.06*	-0.16
teachers	29	223	39	118	0.05*	0.03
results	32	191	35	124	2.41*	-0.26
article	22	182	1905	1	147.34	6.62
use	34	182	37	121	2.98*	-0.29
writing	36	171	57	79	1.41*	0.23
analysis	40	162	106	44	18.64	1.00
findings	42	160	43	99	1.06*	-0.19
group	44	158	131	37	24.75	1.21
based	47	139	123	39	14.78	0.95
data	50	133	30	140	29.67	-0.96
reading	51	132	18	218	106.32	-1.61
teaching	53	128	45	90	3.51*	-0.38

Asterisk (*) = keyness value below the cut-off point which was set as ≥ 6.63 or p≤ 0.01
Minus (-) in Log-R = underuse of the word in PAC in relation to NAC

As indicated in Tables 6.3 and 6.4, the lexical wordlists of the NAC and the PAC show that there are a number of similarities and differences in the two corpora. Of the top twenty

content words, the 12 words highlighted in gray boxes are found in both lists, though with different ranks and frequencies. The words include *English*, *students*, *study*, *reading*, *language*, *learning*, *data*, *research*, *results*, *use*, *teachers*, and *findings*. While the sameness of these words reflects the language-related research in both corpora, the differences of the words in the lists represent the specific focus and interests of that study.

Despite having these 12 words in both lists, the log-L values show that in terms of frequency there is no indication of whether novice and professional writers use the following eight words (*learning*, *research*, *teachers*, *results*, *use*, *writing*, *findings*, and *teaching*) differently in their abstract writing. Only small differences in the frequency of these terms are shown in the log-R values, ranging from -0.38 to 0.23. The striking difference in the lexical usage in the PAC relative to the NAC is found with the word *article* and *l*, with log-R values of 6.62 and 3.12 respectively. This demonstrates that the professional writers used the word *article* more than 64 times in comparison with the novices, and used the letter *l* roughly eight times more than the novices.

In summary, the most frequently found lexical words in (a) the NAC and (b) the PAC are the words displayed in Tables 6.3 and 6.4. Overall, 12 words were found in the top twenty lists for both corpora, but only eight words appeared to be used with a very similar degree of frequency. The similarities were established firstly by looking at log-R values which are very close to zero, and then confirming by log-L values (log-L $\leq$ 6.63). The greatest degrees of difference in lexical word frequency are found in the following words: *Thai*, *communication*, and *program* from the NAC, and *article* and *l* from the PAC. The word *Thai* is unique to the novice corpus, with no occurrences at all in the PAC. Given that the thesis is mainly motivated by pedagogic concerns, these findings are rather trivial and uninteresting, and they did not provide much pedagogically useful information about the data, reflecting the fault of the method.

In the next section, concordance lines will be examined to gain more insights into how these words were used in context.

6.3 Evidence from Concordance Lines: Similarities and Differences

With regard to similarities, the following words were used in the NAC and the PAC with a high degree of frequency. Despite the similar frequencies, of course, the ways in which these words are used in context can differ. The concordance analysis will provide a clearer

picture of the function of each word along with its senses of meaning. Because some words/word forms belong to more than one word class, a part of speech (POS) classification is required. At present, various software can help with POS tagging, such as TagAnt and the CLAWS tagger. However, the decision to employ a manual analysis was made. The reason why automatic tagging is not used in this study is that the software was still unable to distinguish some borderline cases such as between *-ing* form acting as a participle and *-ing* form acting as a modifier. Although automatic taggers could facilitate the tagging process and reduce workload in POS tagging, especially with large corpora, Hunston (2002, p. 83) proposed that, "the tagger may be wrong, and the human user's judgment is more reliable in some cases. Thus, a tagger with an accuracy rate of 96%, say, may be 100% accurate for many words, but only 70% accurate for some words."

Given that automatic POS-tagging cannot be 100% reliable in identifying the correct POS for some words/word forms, each occurrence of the words under discussion in Sections 6.3.1 and 6.3.2 was checked manually. The benefits of manual exploration are that the POS of the target lexical words can be identified more accurately by looking at its context from the concordance lines. An exploration will not be placed on every word in Tables 6.3 and 6.4 due to word limits, but rather on some lexical words that are used with a similar degree displayed in those tables.

6.3.1 Words used in the NAC and the PAC with a similar degree

In this section, four of the eight words which were found to be used with a similar degree of frequency evidenced by log-R values will be further explored by analyzing their parts of speech. These four words were selected on the basis that they are more related to abstract writing in general than the remaining four.

Similarities: ***learning***, ***research***, *teachers*, ***results***, *use*, *writing*, ***findings*** and *teaching*

(1) LEARNING

As shown in Table 6.5, the concordance analysis showed that the majority of *learning* (41.45%) in the NAC functions as a pre-noun modifier such as *vocabulary learning strategies* (19 occurrences), *learning environment* (4 occurrences), and *learning styles* (4 occurrences). In contrast, it was found that the majority of the occurrences of *learning* (54.34%) in the PAC are used as a noun, as in *learning* (15 occurrences), *learning of* (10 occurrences), *implicit learning* (7 occurrences), and *L2 learning* (7 occurrences). The

second most prevalent word class of *learning* in the NAC was found to be nouns (53 occurrences, or 34.87%) such as *language learning* (14 occurrences), *English learning* (8 occurrences) and *self-directed learning* (7 occurrences), while the second most prevalent word class was found to be pre-noun modifiers in the PAC (89 occurrences or 33.58%) such as *learning process* (8 occurrences), *learning context* (4 occurrences) and *learning environment* (4 occurrences).

Table 6.5 POS of LEARNING

Learning	Pre-noun modifier	Noun	Gerund	Participle	Total
NAC	63 (41.45%)	53 (34.87)	34 (22.37%)	2 (1.32%)	152 (100%)
PAC	89 (33.58%)	144 (54.34%)	28 (10.57%)	4 (1.13%)	265 (100%)

When used as a gerund, 17 out of 34 tokens in the NAC are found after either *towards* (11 tokens) or *toward* (6 tokens). With respect to the preposition *toward(s)*, it raises the question of whether the novice writers are more likely to use *British English* than *American English*. But, returning to Table 6.3, Section 6.2.2, this assumption might not be accurate as the word *program* is found to be among the top lexical words (instead of *programme*). Therefore, it should be noted here that some Thai novice writers are likely to employ a mixed spelling of both *British English* (BE) and *American English* (AE), even in one text. For example, the words *toward* and *analysed* were found in the abstract coded as N_TE_2013_32, the words *analyse* and *organization* appeared in N_TE_2012_121, and the words *analyse* and *emphasized* occurred in N_TE_2013_132. Functioning as a gerund, the most prevalent pattern of *learning* in the PAC (10 out of 28 tokens) occurs after *of*. Moreover, the word is rarely used as a participle in both datasets, with only two occurrences (1.32%) appearing in the NAC and four occurrences (1.13%) in the PAC.

(2) RESEARCH

Table 6.6a shows that the word *research* in the NAC is used as a pre-noun modifier (70 occurrences, or 51.09%) slightly more than as a noun (67 occurrences, or 48.91%). Functioning as a pre-noun modifier, it tends to occur with *instrument*, as in *research instrument* (27 occurrences) and with *this … study* as in *this* ***research*** *study* (17 occurrences). When used as a noun, the word tends to occur with the determiner *this*, as in *this research* (31 occurrences), followed by *the research* (14 occurrences) and *further research* (7 occurrences). In contrast, the word *research* in the PAC is used as a noun (141 occurrences, or 62.39%) more than as a pre-noun modifier (85 occurrences, or 37.61%).

The majority of *research* functioning as a noun is used with the determiner *the* as in *the research* (31 occurrences), followed by *previous research* and *research on* (9 occurrences each), *future research* and *little research* (7 occurrences each). When functioning as a pre-noun modifier, the word tends to occur with *articles*, as in *research articles* (15 occurrences), followed by *engagement* (5 occurrences), and *article* (4 occurrences).

Table 6.6 POS of RESEARCH

Research	Pre-noun Modifier	Noun	Total
NAC	70 (51.09%)	67 (48.91)	137 (100%)
PAC	85 (37.61%)	141 (62.39%)	

Functioning as a noun, two verbs that usually occur with the word *research* in the NAC are *was* (15 occurrences, or 22.39%), with five out of 15 occurrences used in the *passive voice*, and *aimed* (6 occurrences, or 8.96%). In the PAC, verbs that are likely to occur with research include *has* (13 occurrences, or 9.22%), which is used in the *present perfect*, as shown in Figure 6.2, and *are* (6 occurrences, or 4.26%) in the *passive voice*, as in *research are discussed* and *research are given*.

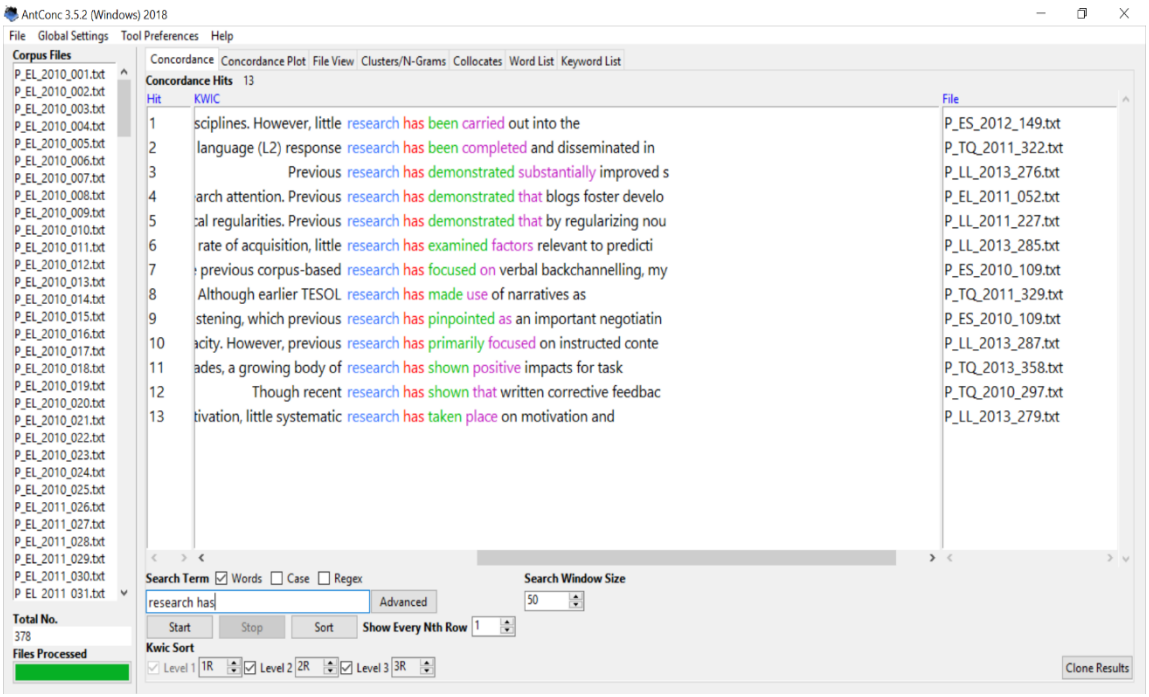

Figure 6.2 The word *research* with the *present perfect tense*

(3) RESULTS

Table 6.7 reveals that although the word *results* can function as either a noun or a verb, it was found that nearly 100% of the occurrences in the NAC and 100% in the PAC are its use as a noun. Only one case in the NAC (0.81%) is used as a verb as shown in the concordance line below.

(1) English speaking skill have played the important role in today world which ***results in*** high competition …

(P_LL-2011_212)

Table 6.7 POS of RESULTS

Results	Noun	Verb	Total
NAC	123 (99.19%)	1 (0.81%)	124 (100%)
PAC	191 (100.00%)	0 (0.00%)	191 (100%)

As a noun, the word *results* in the NAC is usually followed by the following verbs: *revealed* (17 occurrences), *showed* (12 occurrences), *show* (17 occurrences) and *indicated* (7 occurrences). In contrast, the verbs accompanying *results* in the PAC include: *show* (24 occurrences), *suggest* (22 occurrences), *showed* (19 occurrences), *indicate* (10 occurrences), and *revealed* as well as *indicated* (9 occurrences each).

Table 6.8 POS of FINDINGS

Findings	Noun
NAC	99 (100%)
PAC	160 (100.00%)

The concordance analysis of the word *findings* shows that the majority in the NAC tends to occur after the determiner *the* (78 occurrences, or 78.79%) and *these* (5 occurrences, or 5.05%). The word is usually followed by prepositions *of* + N (20 occurrences) and *from* + N (7 occurrences). Verbs that are used with this word include *revealed* (25 occurrences), *indicated* (9 occurrences), and *showed* (6 occurrences). In contrast, the word *findings* in the PAC appears with the determiner *the* (70 occurrences) and *these* (30 occurrences). It is also followed by the prepositions *of* (14 occurrences) and *from* (6 occurrences). Verbs that appear with *findings* include *show* (24 occurrences), *suggest* (22 occurrences), *showed* (19 occurrences), *indicate* (10 occurrences), and *revealed* (9 occurrences).

The subsequent section focuses on the lexical words in the top 20 wordlists which are found to be used with a striking degree of difference between the NAC and the PAC.

6.3.2 Words used in the NAC and the PAC with a striking degree of difference

The top three lexical words that are found to be most different in terms of percentage and ratio in the NAC when compared to the PAC are *Thai* (log-R = 9.51), *communication* (log-R = 8.69), and *program* (log-R = 8.58), while those found in the PAC, when compared to the NAC, are *article* (log-R = 6.62) and *l*, the majority of which stands for *language* (log-R = 3.12). The word '*Thai*' obviously appears only in the NAC. Due to word limits, only two words - '*program*' and '*article*' - will be further explored here.

Differences: *Thai*, *communication,* and ***program*** (NAC), ***article*** and *l* (PAC)

(1) PROGRAM

As shown in Table 6.9, the concordance analysis of the NAC revealed that the majority of the uses of the word *program* (78 instances, or 75%) were used in relation to the participants' study program, such as *Math-Science Program students, English and Japanese Program*, and *TEFL program*. However, the remainder (25 instances, or 24.04%) were used in relation to computer programs, specifically '*SPSS program*' or a statistical software program (23 instances), and '*Excel program*' (2 instances). The only instance of the word '*program*' which did not fit into these categories was the word '*TV program*' (1 instance, or 0.96%).

Table 6.9 POS of PROGRAM

Noun	Relating to computer program	Relating to study	Relating to TV
NAC	25 (24.04%)	78 (75.00%)	1 (0.96%)
PAC	1 (5.88%)	15 (88.24%)	1 (5.88%)

The concordance analysis of the PAC showed that out of 17 occurrences, the majority (15 occurrences, or 88.24%) was also used with regard to the participants' study program, such as *TESOL program* and *reading program*. One instance each (5.88%) was employed in relation to a computer program and a television program.

(2) ARTICLE

In relation to the word *article* in the PAC, 177 occurrences (97.25%) were employed in the sense of a piece of writing (see example 2). Only five of the 182 instances (2.75%) were found not to be used in this sense (see example 3). The analysis, therefore, reveals that some of the words, when employed in the academic context, convey a specific meaning or a specific function.

(2) The ***article*** concludes with some practical implications for ESP practitioners.

(P_ES_2013_174)

(3) In each group, both noun-***article*** and noun-adjective gender agreement processing were examined behaviorally …

(P_LL_2010_180)

Table 6.10 POS of ARTICLE

Word Sense	A piece of writing	Part of speech
NAC	-	1 (100%)
PAC	177 (97.25%)	5 (2.75%)

6.4 Keyword Lists: Zero Occurrence

The term 'keyword' in corpus linguistics refers to a word whose frequency is statistically high (positive keywords) or low (negative keywords) compared to a reference corpus, which is set as a benchmark (McEnery *et al.*, 2006, p. 347). By comparing a word frequency list of the target corpus with that of the reference corpus, results generated by AntConc software will show the keyword list, along with its keyness and effect size statistics. In this study, a comparison was made between the word frequency lists of the NAC and the PAC. When the NAC is used as the target corpus, the PAC will be used as the reference corpus, and vice versa.

In this section, the keyword lists of the NAC and the PAC that are found to have zero occurrence in relation to the reference corpus will be discussed. Due to word limits, only these keywords with zero occurrence in comparison with the reference corpus are discussed, but not the keywords occurring at the minimum frequency of five because these words undoubtedly show how the NAC and the PAC differ in their lexical choice usage. The lists, which were divided according to moves, were arranged by log-L and filtered by log-R and range (see Appendix I for the statistical detail of each word). The range was

used here to minimize the tendency of some words classified as keywords but found only in one or two specific files. The cut-off threshold for the range was set at five for the overall NAC and 10 for the overall PAC. The thresholds of range in each move were set at three for the NAC and six for the PAC. The arbitrary cut-off setting relies on the corpus sizes, so inconsistencies among different research are often found. The threshold of the PAC was double the NAC because the corpus size of the PAC was roughly double that of the NAC.

6.4.1 The NAC keyword list in relation to zero occurrence in the PAC

Table 6.11 presents the NAC keyword list in relation to zero occurrence in the PAC. The keywords in m1 are found to be related to the background of the studies, such as place and nationality (*Thai, Thailand* and *foreigners*) and the acronym '*AEC*', indicating the current situation regarding the inauguration of the ASEAN Economic Community. Proper nouns related to places, such as *Bangkok, Thammasat* and *Silpakorn*, and those related to nationality (i.e., Thai) are still found as keywords in m2. It is notable that the word

Table 6.11 The NAC keyword list in relation to zero occurrence in the PAC, arranged by log-L and filtered by range
(Criteria: Log-L ≥ 6.63, Range ≥ 5)

Move	Keywords (arranged by log-L)
m1 (Introduction)	**thai**, **thailand**, employees, **aec**, due, effectively, economic, **foreigners**
m2 (Purpose)	**thai**, **bangkok**, satisfaction, employees, objectives, **thammasat**, analyze, **silpakorn**
m3 (Method)	thai, **spss**, **percentage**, bangkok, **deviation**, thammasat, employees, **likert**, ltd, officers, **convenience**, majoring, objectives, silpakorn, **checklist**, December, technology
m4 (Product)	thai, problem, needed, **satisfied**, **rated**, having, **satisfaction**, mail, **ranked**, wanted, physical, communicate, intrinsic, media, suggestions
m5 (Conclusion)	thai, employees, able, company, **encourage**, program, **recommended**, techniques, thailand, hotel, reliable, **respondents**.

objectives, signifying the communicative purpose of this move, has zero occurrence in the PAC. Although the proper nouns found in m1 and m2 are also present in m3, what stand out as the keywords in this move are those related to the data analysis process, such as *spss, percentage, deviation, likert, convenience* and *checklist*. These words clearly show that it is unnecessary to indicate the statistical program used for analysis. Stating the percentage of the results and the standard deviation in an abstract is also uncommon. Four keywords (*satisfied, rated, satisfaction* and *ranked*) in m4 are found to be interrelated.

These words suggest that a number of novice writers provided a description regarding their questionnaire. The four words are also related to the keyword *likert* in m3. Two keywords that signify the main communicative purpose in m5 are *encourage* and *recommended.* As regards the word *encourage*, the novice researchers used this word to describe pedagogical implications. With regard to *recommended*, the concordance lines reveal two main types of usage. The first type is used as recommendations for further research or for teaching and learning, and is used in the passive voice. The second type of usage is related to the open-ended questionnaire that the respondents recommended regarding what should be done.

6.4.2 The PAC keyword list in relation to zero occurrence in the NAC

As shown in Table 6.12, a greater number of PAC keywords are not found in the NAC. With regard to m1, the distinctive keywords can be grouped into those related to the significance of the study (i.e., *development, attention, literature* and *critical*), and conjunctions signifying contrast and an adverb of frequency (i.e., *while, but, yet, despite* and *rarely*). It is worth noting that on some occasions the word *but (also)* was used as a coordinating conjunction to denote the significance of the study, as in the examples below.

(1) Vocabulary knowledge …., not only as a means to facilitate first language **acquisition but also** as an integral element in the learning of English as a second or foreign language.
(2) The unit not only exemplified the theory of project-based learning (PBL) for the student-teachers **but also** involved them in a hands-on experience.

It is noticeable that a variety of verbs, namely *investigates, describes, discusses* and *presents*, are PAC keywords with zero occurrence in relation to m2 of the NAC. However, after checking the concordance lines of the NAC, they reveal that only *investigate* appears in m2 of the NAC, but it was used as an infinitive verb (as in the example below). Three other words do not occur in m2 of the NAC at all.

(3) The purpose of this study was to investigate …

The keywords in m3 of the PAC can be grouped as those proper nouns related to places, languages and nationality, such as *Spanish, Chinese, Hong Kong, Dutch, Mandarin, Swedish,* and *(United) States*. This reflects that in the Method move, the researchers are likely to indicate the nationality of the participants. This means that the occurrences of

Thai or *Thailand* in the NAC should not be regarded as inappropriate. However, what seems to be different is that these words appear repeatedly in every move of the NAC, but these kinds of keywords occur only in m2 and m3 of the PAC.

Table 6.12 The PAC keyword list in relation to zero occurrence in the NAC, arranged by log-L and filtered by range

Move	Keywords (arranged by log-L)
m1 (Introduction)	**development**, into, based, between, **attention**, education, linguistic, **while**, **but**, assessment, feedback, input, where, esp, **literature**, focused, g, genre, **yet**, learner, new, paper, task, tesol, themselves, article, competence, **critical**, **despite**, develop, **rarely**
m2 (Purpose)	article, **investigates**, education, context, genre, chinese, **describes**, spanish, word, languages, current, hong, individual, kong, may, over, drawing, forms, investigation, role, **discusses**, practice, **presents**, processes, related, attention, children, interaction, experiences, results, small, time
m3 (Method)	n, **spanish**, **chinese**, **investigates**, **articles**, interaction, model, **analyses**, **hong**, input, **kong**, children, **describes**, languages, may, drawing, **measured**, processes, small, across, conditions, experiences, new, system, talk, tesol, authors, feedback, modeling, noun, planning, complexity, peer, fluency, our, **presents**, sessions, critical, **discusses**, esp, g, identity, large, longitudinal, pragmatic, structural, beliefs, collaborative, comparison, conversation, novice, value, author, **discussion**, **dutch**, efficacy, interactions, lectures, processing, recognition, **states**, times, translation, video, whereas, account, audio, capacity, complex, disciplinary, **discussed**, engagement, **gains**, gap, his, making, **observed**, professionals, relevant, **shows**, treatment, vs, wp, adults, **associated**, direct, **evaluated**, experiments, initial, insights, irregular, long, middle, month, pronunciation, reflection, although, backgrounds, **draws**, episodes, **followed**, **gathered**, home, **illustrate**, **influenced**, journal, **mandarin**, nature, norms, old, per, perspective, **swedish**, via
m4 (Product)	**analyses**, **gains**, spanish, over, forms, discourse, across, **feedback**, children, disciplinary, **our**, genre, writers, accurate, **evidence**, **interview**, languages, article, longer, months, range, regular, **we**, acquisition, **emerged**, **influenced**, institutional, **led**, particular, posttest
m5 (Conclusion)	article, **we**, effects, **I**, model, frequency, processing, professional, first, genre, **indicate**, memory, **our**, **conclude**, discourse, meaning, nature, changes, **evidence**, interaction, **show**, written

As m2 and m3 are commonly found embedded together, some keywords in m2 (e.g., *investigates, describes* and *presents*) are also found in m3. However, a greater variety of verbs are found in this move, such as *analyses, measure, gains, shows, observed, evaluated, followed, gathered* and *draws*. What stand out as keywords in m4 of the PAC includes verbs, such as *gains, emerged, influenced* and *led*, and pronouns and possessive adjectives (such as *our* and *we*). Although the word *analyses* can function as a verb, it was found to function mainly as a noun in this move (for example, *our analyses revealed that*

…, *subsequent analyses showed that* …, and *quantitative analyses suggest that*…). With regard to the use of pronouns and possessive adjectives, this suggests that some professional researchers employed these words to signify engagement with the readers, which is in line with Hyland's (2005) claim that self-mention is important in showing the standpoint of writers when making an argument. Self-mention can also be found as keywords in m5 of the PAC, where the words *we, our* and *I* are found. Among the three verbs (*indicate*, *conclude* and *show*) that are found as keywords in m5 of the PAC, the word *conclude* clearly signifies the communicative purpose of this move, and it is noteworthy that this word is not found in m5 of the NAC.

6.5 Summary

This chapter has, firstly, presented the wordlists of the NAC and the PAC with regard to the twenty most frequently found words. In subsequent sections, when function words were filtered out, the top twenty lexical words of the NAC and the PAC have been shown. Although twelve lexical words were found in both corpora, the effect size measure called log-R revealed that only eight (***learning***, ***research***, *teachers*, ***results***, *use*, *writing*, ***findings*** and *teaching*) were found to be used with a similar degree of frequency. Log-R also showed that the lexical words used with a striking degree of difference were *Thai*, *communication,* and ***program*** from the NAC, and ***article*** and *l* from the PAC. The words highlighted in bold were further analyzed by looking at their contexts in the concordance lines. Apart from focusing on the wordlists, this chapter has also explored the keywords that have zero occurrences in relation to the reference corpus. The results suggest a number of distinctive characteristics in terms of the lexical choices used in each move of the corpora.

Chapter 7. Lexical Bundles in Abstract Writing

7.1 Introduction

This chapter has two main aims: (1) to identify the multi-word constructions, or lexical bundles (hereafter LBs) which characterize each move of the research abstracts (RABs) in the target corpora (NAC and PAC); and (2) to classify these bundles according to their grammatical structures and functions. The findings will therefore address the following question:

- **(RQ4)** What lexical bundles are associated with in (a) the NAC and (b) the PAC?

The chapter is divided into five sections. After the introduction, Section 7.2 discusses the criteria used in extracting the LBs. Then Section 7.3 provides the description of the categories of the LB structural and functional analyses. In Section 7.4, the comparisons of structural, functional and concordance analyses according to each move are presented and discussed. Section 7.5 summarizes the key findings of the chapter.

Before proceeding to the criteria for the LB retrieval, it is important to review some important characteristics of LBs that have been pointed out in the literature. For example, LBs are 'usually not idiomatic and are not complete structures, but are important building blocks of discourse' (Biber *et al.*, 2012, p. 7) which are 'semantically transparent and formally regular' (Hyland, 2008a, p. 6). Moreover, the distribution patterns of LBs in academic prose tend to be 'phrasal', while those in conversation are likely to be 'clausal' (Chen and Baker, 2010, pp. 30-31). 'Phrasal' in this sense means that the majority of LBs can be categorized as either noun phrases, prepositional phrases, or verb phrases, while 'clausal' means that those LBs consist of a subject and a predicate, though they tend to be incomplete, appearing in a fragment form. Nevertheless, Biber and Barbieri (2007, p. 273) claim that 'the extent to which a speaker or writer relies on lexical bundles is strongly influenced by their communicative purposes, in addition to general spoken/written differences', which means that the distribution of LBs is more distinguishable by genre, rather than by mode (Hyland, 2012a). LBs recur or can be found in multiple texts within a certain register (Esfandiari and Barbary, 2017, p. 22). Having recognized some distinctive characteristics of LBs, it is interesting to examine how certain LBs occur in

academic prose, specifically in each move of the research abstracts, and whether there will be more phrasal than clausal LBs, as suggested by Chen and Baker (2010).

The next section discusses three important criteria and some methodological issues in the retrieval of LBs from the move corpora of NAC and PAC.

7.2 Criteria for the LB Retrieval

Three criteria were used in extracting the LBs from the corpora: frequency of occurrence, length, and distribution. The exact nature of these criteria was established in order to generate a manageable number of LBs, appropriate to serve the purpose of the study, time constraints, and the word limits.

7.2.1 Frequency cut-off points of lexical bundles

With regard to the frequency criterion, studies have used a variety of arbitrary cut-off points, such as 25 times per million words (pmw hereafter) (Ädel and Erman, 2012; Chen and Baker, 2010), 20 times pmw (Hyland, 2008a; 2008b), 30 times pmw (Shahriari, 2017) and 40 times pmw (Biber and Barbieri, 2007; Grabowski, 2015). In some instances, more than one threshold frequency has been used in a single study (Cortes, 2013; Esfandiari and Barbary, 2017). For example, Cortes (2013) set thresholds of 20, 10, 8 and 6 times pmw in order to identify 4-, 5-, 6- and 7-word bundles respectively in a one-million-word corpus of research article introductions, while Esfandiari and Barbary (2017) used values of 40, 20 and 10 times pmw to examine 4-, 5- and 6-word LBs respectively in a pair of two-million word corpora as part of a cross-linguistic study of psychology research articles (English and Persian). All in all, the minimum cut-off point of at least 10 times pmw for 4-word bundles is most commonly used, with lower cut-off points for longer LBs (Salazar, 2014, p. 13). These thresholds are used in part as a factor to determine which LBs are worth examining further. It is also important to note that one possible factor contributing to the use of different threshold frequencies is the corpus size.

Although the cut-off point of 40 times pmw has been used in some studies (e.g. Biber and Barbieri, 2007; Grabowski, 2015), this threshold is not suitable in the present study for two main reasons. Firstly, considering that the corpus size of NAC is 33,130 words, with PAC consisting of 61,117 words, the raw frequencies that correspond to 40 occurrences pmw would be only one token (rounding down from 1.33) in NAC and two tokens

(rounding down from 2.44) in PAC. Raising the cut-offs from 40 to 50 or 60 times pmw would have little effect: the corresponding raw frequencies in NAC would only increase to 2 (rounding up from 1.66 and 1.99), while those for PAC would be 3 or 4 (rounding down from 3.06 or up from 3.67). The relationships between these normalized frequencies and corresponding raw frequencies are shown in Table 7.1. When using these raw frequencies for LB extractions, too many LBs will be extracted, and it is likely that some of them may occur due to chance. It is therefore questionable that a cut-off frequency set at 40 pmw is valid and applicable to all corpus sizes.

Table 7.1 Normalized frequency cut-off points and corresponding raw frequency

Corpus	Norm freq. (pmw)	Corresponding raw freq.	Norm freq. (pmw)	Corresponding raw freq.	Norm freq. (pmw)	Corresponding raw freq.
NAC	40	1.33	50	1.66	60	1.99
PAC	40	2.44	50	3.06	60	3.67

The second reason why a threshold of 40 times pmw should be reconsidered is related to the research question (RQ4), which aims to identify lexical bundles in relation to the different moves (the communicative purposes). In this respect, rather than extracting LBs from the NAC/PAC corpus as a whole, each of the annotated move corpora are examined in turn. Considering the number of word tokens in each rhetorical move subsection of the corpora, as illustrated in Table 7.2, it is apparent that this involves focusing on relatively smaller amounts of words.

Table 7.2 Type/token of rhetorical moves and the cut-off frequency thresholds

Moves	NAC		Cut-off	PAC		Cut-off
	Types	Tokens	raw freq.	Types	Tokens	raw freq.
m1	798	2633	3	2046	8345	8
m2	1058	5865	6	2711	15018	15
m3	1811	14526	15	3999	30498	10
m4	1875	12127	12	3108	17732	18
m5	952	3993	4	2158	9857	10

A recent study by Bestgen (2018) suggests that the corpus size and the length of LBs should be taken into consideration when determining an appropriate cut-off threshold. The four corpora Bestgen analyses are carefully selected to reflect different sizes (one corpus of 4,200,000 words, two of 150,000 words each, and one of 50,000 words), different modes (one spoken corpus and three written corpora) and different types of

speakers/writers (two corpora of native English speakers, and two of non-native speakers), so that the findings can be compared with previous similar studies. Using a statistical significance test (Fisher's exact test) to evaluate different frequency thresholds for different sizes of lexical bundles, Bestgen (2018, p. 215) proposes that the following normalized frequency thresholds should be applied in the case of three-word bundles: 25 occurrences pmw for the 4,200,000-word corpus, 80 occurrences pmw for the 150,000-word corpora, and 140 occurrences pmw for the 50,000-word corpus. In making these recommendations, he convincingly argues that the smaller the corpus size, the higher the frequency thresholds should be, in order to confirm that the results are not due to chance. When it comes to the four-word bundles, his study also demonstrates that cut-off thresholds ranging from 10 to 40 occurrences pmw prove effective for the 4,200,000-word corpus, while a raw frequency of four is required for the 150,000-word corpora and a raw frequency of three for the 50,000-word corpus.

In light of this, the present study considered setting a higher normalized frequency for the bundle retrieval. Taking Bestgen's (2018) proposition regarding the higher cut-offs into consideration, it seems that the rigorous frequency threshold might become problematic with small specialized subcorpora because it will result in a very small number of bundles or perhaps even zero bundles for the analysis, thereby potentially disregarding some interesting results. Consequently, the decision to employ the size of the sub-corpora of the NAC as a basis for setting a cut-off frequency point was made, with a proportional raw frequency of one occurrence per 1000 tokens. This means that in m1 of the NAC, which has the total of 2,633 words, the raw frequency cut-off point will be 3 (rounding up from 2.63) (see Table 7.2 above for further details). However, problems still exist because, when applied to the PAC, this approach results in a very small number of LBs, as seen in Table 7.3a. For example, with a frequency threshold of 18 used as a cut-off, only one LB type is identified in m4 of the PAC (see the gray highlight).

As shown in Table 7.3a, the results revealed that the total number of LBs in the NAC is 168 types and 2,189 tokens, while the total number of LBs in the PAC is 34 types and 648 tokens. This suggests that LBs are more common in the NAC than in the PAC. The highest number of LBs was found in m2 of both corpora, with 59 bundles in the NAC and 16 bundles in the PAC. LBs were least frequent in m4 of both corpora, with 19 bundles in the NAC and only one bundle in the PAC.

Table 7.3a Types and tokens of the LBs in each move of the NAC and the PAC, together with the initial threshold criteria for LB extraction

Moves	LBs in NAC		Cut-off raw freq.	Range	LBs in PAC		Cut-off raw freq.	Range
	Types	Tokens			Types	Tokens		
m1	21	96	3	3	5	45	8	8
m2	59	767	6	6	16	332	15	15
m3	36	790	15	15	4	151	30	30
m4	19	349	12	12	1	19	18	18
m5	33	187	4	4	8	101	10	10
Total	**226**	**2189**			**34**	**648**		

The decision was made to adjust the cut-off frequencies until the number of LBs in PAC is nearly equivalent to the number of LBs found in the NAC, in order that more LBs in PAC could be identified and explored (see Table 7.3b). For example, if the number of LBs in m1 of the NAC was found to be 21, the frequency cut-off points were adjusted until the results in PAC nearly reached that number. In this case, the frequency cut-off was adjusted to 4, resulting in 28 LBs with 148 tokens. In practice, exact equal numbers could not be achieved because some of the LBs at the bottom of the lists occur at the same low frequency, meaning that there would be no principled way of retaining some of this set of LBs, while excluding others. To limit the LB number to more manageable sizes, only the top 20-30 LBs are examined, depending on the preliminary results. Supposing the preliminary results of LBs in the NAC show 25 LBs in m1 and 50 LBs in m2, the number of LBs for further scrutiny will be roughly 20 in m1 and 30 in m2. In terms of the comparison between the two corpora, the results will be indicative in that which types of bundles with regards to their structure and function each group tend to rely on. However, it is important to note that direct comparisons between the number and proportion found are not appropriate because of the need to apply different frequency cut-off criteria.

As shown in Table 7.3b, the final threshold criteria for LB extraction from the PAC were adjusted according to the number of LBs extracted in each move. After the adjustment, the number of LBs identified in each move of both corpora are broadly similar. The number of LBs in the NAC versus the PAC are as follows: m1 (21:28), m2 (59:45), m3 (36:34), m4 (19:32) and m5 (33:38). Because only the top twenty to thirty LBs are investigated, the cut-off points in PAC have been adjusted until approximately 30 LBs or more are shown. The final number of LBs in PAC is 177, which is closer to the total of LBs in the NAC (168). After the retrieval, the number of LBs identified for each corpus

was further filtered in cases where overlap between LBs occurs. An explanation of the ways in which bundles could overlap and how they were filtered is provided in Section 7.2.4.

Table 7.3b Types and tokens of the LBs in each move of the NAC and the PAC, together with the final threshold criteria for LB extraction

Moves	LBs in NAC		Cut-off raw freq.	Range	LBs in PAC		Cut-off raw freq.	Range
	Types	Tokens			Types	Tokens		
m1	21	96	3	3	28	148	4	4
m2	59	767	6	6	45	661	10	10
m3	36	790	15	15	34	726	14	14
m4	19	349	12	12	32	303	7	7
m5	33	187	4	4	38	283	5	5
Total	**168**	**2189**			**177**	**2121**		

7.2.2 Length of lexical bundles

With regard to the cluster length, only three-word and four-word bundles were investigated because three-word bundles are claimed to be the most prevalent (Hyland, 2012a), while four-word bundles are the most frequently studied (Bestgen, 2018). Longer bundles like five-word bundles are not the focus of the study. This is due to the fact that with the thresholds set in Table 7.3a above, only a handful of five-word clusters were found after the initial exploration (see Table 7.3c), and many of these were simply extensions of three- or four-word bundles.

Table 7.3c Types and tokens of the LBs in the NAC and the PAC according to length (three-, four- and five-word LBs)

Moves	LBs in NAC						LBs in PAC					
	three-word		four-word		five-word		three-word		four-word		five-word	
	types	tokens	types	tokens	types	tokens	types	tokens	types	tokens	types	tokens
m1	17	76	4	20	2	8	4	37	1	8	0	0
m2	41	542	18	225	11	125	13	279	3	53	0	0
m3	29	652	7	138	2	33	4	151	0	0	0	0
m4	14	265	5	84	2	30	1	19	0	0	0	0
m5	25	149	8	38	3	14	8	101	0	0	0	0
Total	**126**	**1684**	**42**	**505**	**20**	**210**	**30**	**587**	**4**	**61**	**0**	**0**

7.2.3 Distribution of lexical bundles

To avoid the extraction of LBs being influenced by the idiosyncratic, repeated usage of a single writer, it is important to apply an additional criterion: each LB must be found in

multiple texts in the corpus (Biber and Barbieri, 2007). As with the arbitrary frequency thresholds, there is no consensus with respect to what constitutes an appropriate distribution threshold. Some studies (Hyland, 2008a, 2008b) set the threshold as a minimum percentage, such as 10%, while other studies either set a fixed minimum range (Ädel and Erman, 2012; Chen and Baker, 2010; Cortes, 2013; Qin, 2014) or different minimum ranges for different sizes of corpus (Biber and Barbieri, 2007).

In the present study, the range was set according to the frequency cut-off thresholds outlined in section 7.2.1. Given the concise nature of abstracts, using 10% distribution would mean that any particular LB must appear in at least 15 NAC texts and 38 PAC texts, which is not practical because the frequency thresholds which were set according to the proportion of the tokens in each move would definitely fail to meet this criterion. Therefore, the decision to use the same cut-off thresholds with the range criteria was made. This is to guard against the recurrence of the same LBs occurring in only a few texts as concise as abstracts (the average length of the abstracts is only 233 words in the NAC and 177 words in the PAC).

7.2.4 Exclusion or inclusion of overlapping LBs

This section discusses some other methodological issues in relation to questions about overlapping LBs, and provides a rationale explaining why these LBs should be excluded or retained.

Chen and Baker (2010, p. 33) point out that some short bundles are components of longer bundles, and argue that these overlapping bundles should be counted only once to avoid inflating results. In their study, two kinds of overlapping LBs are identified. The first is referred to as 'complete overlap', in which two four-word bundles are from the same five-word bundle. These four-word bundles also occur with the same frequency. For example, both *it has been suggested* and *has been suggested that* are components of *it has been suggested that*, a five-word bundle. Another type of overlapping is called 'complete subsumption', in which a five-word bundle is derived from two or more four-word bundles, but the four-word bundles do not have the same occurences. For example, *as a result of* (17 occurrences) and *a result of the* (5 occurrences) are a subset of *as a result of the*. Therefore, Chen and Baker combined each case of the overlapping bundles into one longer unit (12 cases in total). In this study, those three-word LBs with complete overlap

to four-word bundles and the same frequency of occurrence will be combined as one. For example, in the case of having both *is one of* (7 occurrences) and *is one of the* (7 occurrences) in the extraction, only the latter bundle is retained.

However, in the case of complete subsumption, the decision to retain was made. For example, the three-word bundle in m1 of the NAC *one of the* can be a subset of a longer five-word bundle such as *one of the most important*, or a four-word bundle such as *one of the most* or *is one of the*, but not vice versa. For example, *one of the* occurs 16 times in m1 of the NAC, and *of the most important* occurs 4 times, but there are only 6 occurrences of *one of the most*. If combining the longer bundles with the shorter ones, only 6 occurrences of *one of the most* can be counted because not all instances of *one of the* occur in conjunction with the word *most*. In addition, despite the fact that *of the most important* is a part of *one of the most important*, the criteria limit the analysis to four-word bundles only. This was because, after some concordance exploration, it was found that one of the shorter bundles might meet the frequency and the dispersion criteria to be included for the analysis, but the longer one might not. In addition, attempting to combine LBs that have different frequencies of occurrence would complicate the process and cause confusion. Therefore, in the case of the overlapping bundles that occur with different frequencies, the decision to retain those bundles without exclusion was made.

Table 7.4 Number of bundles before and after the exclusion of the complete overlaps

Move	**Before exclusion (NAC)**		**After exclusion (NAC)**		**Before exclusion (PAC)**		**After exclusion (PAC)**	
	No. of LBs		No. of LBs		No. of LBs		No. of LBs	
	Types	Tokens	Types	Tokens	Types	Tokens	Types	Tokens
m1	21	80	16	73	28	148	23	128
m2	36	610	32	530	45	661	34	513
m3	36	723	33	790	34	726	30	633
m4	22	372	17	299	32	303	19	206
m5	33	187	27	154	38	283	32	249
Total	**148**	**1972**	**125**	**1846**	**177**	**2121**	**138**	**1729**

As shown in Table 7.4 below, the total number of LBs identified in the NAC was previously 148 types, with 1972 tokens. After excluding the complete overlap bundles, the total number is 125 types, with 1846 tokens. In contrast, the total number of LBs identified in the PAC was originally 177 types, with 2121 tokens. After excluding overlaps, the number is 138 types with 1729 tokens.

After the refinement of these bundles, comparisons of LBs in terms of their structures and functions in each move of the two corpora were made. The findings are presented in the following section.

7.3 Categories of Structural and Functional Analyses

LBs are commonly classified according to their structures and functions. In this study, the structural analyses are based on the clusters automatically generated using the thresholds indicated in Table 7.3b. It is worth noting once again that these bundles are among the top twenty to thirty bundles identified in each move in the two corpora, representing salient linguistic patterns in the moves. The taxonomies used to classify the LBs in this study were adapted from the studies of Chen and Baker (2010) for structural analysis and Liu (2012) for functional analysis. In structural terms, the bundles are classified into four broad categories: (1) NP-based, (2) PP-based, (3) VP-based, and (4) Others. Two main differences between this study and Chen and Baker's study lie in discrepancies in the 'VP-based' and 'Others' classification. Firstly, while Chen and Baker (2010) placed clusters such as *little is known about,* and *in order to make* in the VP-based categories on the basis that they contain verbs, this study takes other aspects of their form into consideration when categorizing these two clusters. For example, *little is known about* is classified as 'NP + V' under the 'others' category. This is because *little* in this clause functions as a pronoun and is followed by a verb. The cluster *in order to make* was categorized under PP-based or prepositional phrases because it begins with a preposition. Secondly, in Chen and Baker's study, there was a small number of LBs belonging to 'Others', which means that the category is not prominent. In this study, one salient type of cluster that emerges from the 'NP + V' subcategory of others involves the combination of a noun phrase with a verb phrase, in other words a cluster that is clausal. There are two reasons why Liu's (2012) taxonomy was adopted along with Chen and Baker's (2010). Firstly, the fine-grained subcategories and comprehensive examples it provides make the process of classification easier, in comparison with other studies that use a small number of subcategories, illustrated by relatively few examples. Secondly, some aspects of Liu's classification were found to correspond with the findings emerging from the present study, for example in terms of the very high frequency of 'NP + action verb' clusters.

The following subsections, 7.3.1 and 7.3.2, give full details of the structural and functional characteristics of the four types of bundles and their subcategories.

7.3.1 Structural classification

In this section, descriptions of the structural classification are provided. As noted above, the present study adopts and adapts the taxonomies of Chen and Baker (2010) to categorize the LBs found in each type of move in the two corpora. The LBs are grouped into four broad categories: (1) NP-based; (2) PP-based; (3) VP-based; and, (4) Others.

NP-based refers to (a) any noun phrases with pre-modifiers, (b) those with post-modifiers, and (c) those with both pre- and post-modifiers.

PP-based refers to constructions consisting of a preposition plus a noun-phrase fragment. The PP-based category is divided into two subcategories: (1) those starting with a preposition *of* and (2) those starting with a preposition other than *of*. While Chen and Baker (2010) did not classify PP-based into subcategories, this study divides it into these two subcategories. It is observable that among all the prepositional phrases, those consisting of the preposition *of* plus an NP fragment occur so frequently that their proportional distribution in some moves is found to be higher than or similar to that of all other types of PPs combined, such as in m2 (11.32%: 2.45%), m3 (13.14%: 6.64%), m4 (10.03%: 11.71%) of the NAC.

As noted in the previous section, Chen and Baker (2010, p. 34) grouped any clusters containing a verb component into the VP-based group. In contrast, the present study uses the **VP-based** category for any bundles beginning with a verb component, as well as those beginning with anticipatory subject *it* and *there*, while LBs such as *in order to make* are assigned to PP-based on the basis that they begin with a preposition. The VP-based category is composed of four subcategories: (1) VP-active for those beginning with an active voice verb; (2) VP-passive for those beginning with a passive voice verb; (3) (V) + to infinitive for those containing a to infinitive, regardless of whether or not there is another verb in front of the infinitive; and, (4) those beginning with anticipatory ***it/there***.

The fourth main category is **Others**, which covers any other phrases, clauses, or fragments that do not belong to the other three categories. The Others category consists of two subcategories: (1) NP + V for those beginning with a noun phrase plus a verb phrase; and (2) Miscellaneous for those that count as 'others' but do not belong to the NP + V subcategory.

7.3.2 Functional classification

Lexical bundles normally perform one of three basic functions: (1) Referential; (2) Discourse/Textual Organizer; or (3) Stance/Interpersonal/Personal Expressions (Biber *et al.*, 2004; Chen and Baker, 2010; Liu, 2012). This functional taxonomy originated from the pioneering work of Biber *et al.* (1999), which considered a wide range of genres and modes (written and spoken) and has been used as a starting point for many subsequent studies (e.g., Chen and Baker, 2010; Cortes, 2013; Esfandiari and Barbary, 2017; Liu *et al.*, 1999). This study specifically follows Liu's (2012) adaptation of Biber *et al.*'s original taxonomy, which adds a fourth category, referred to as 'research-topic related', to the original three. Therefore, the LBs identified in NAC and PAC are grouped according to the following four functional categories.

Although the LBs functional classification was also made in Hyland's (2008b) study (see Section 3.4 for details), his classification was not easy to follow. For example, an LB *in the present study* was identified as both research-oriented (location subcategory) and text-oriented (structural signals subcategory). This made it questionable why an LB *in the next section* was classified as only text-oriented (structural signals subcategory), but not as research-oriented (location subcategory). Without providing any samples from concordance lines and explanation, it was unclear why the classification was made as such.

Referential LBs provide attributive features. Liu (2012) divides the referential functions into 9 subcategories as follows: (1.1) intangible framing such as *in terms of*, *based on the*; (1.2) tangible framing such as *in the form of, in the sizes of* (not found in the list of NAC/PAC LBs); (1.3) identification/identity specification such as *one of the, one of the most*; (1.4) qualifying; (1.5) quantity specification such as *many of the*, *some of the* (not found in the list); (1.6) referential text/group such as *of this study, in this article*; (1.7) referential time/sequence such as *in the academic year, in recent years*; (1.8) reporting/description/interpretation such as *the findings suggest, results show that*; and, (1.9) multifunctional such as *the development of, be used to* (see Appendix M for more examples). Of these 9 subcategories, no examples of category 1.4, qualifying referential LBs, were identified at the top of the NAC or PAC lists.

Discourse/Textual organizers, as the name suggests, are used to organize texts and develop their coherence. They are divided into two subcategories, namely linking LBs

such as *in addition the, as well as,* and topic introduction LBs such as *with respect to, with regard to*. No examples of the second subcategory were found in this study.

Stance/Interpersonal/Impersonal Expressions are used to convey the writer's evaluation or attitudes with regard to his/her proposition. They are broadly divided into two subcategories: (1) epistemic stance and (2) attitudinal/modality stance (e.g., *important role in, will be beneficial, need to be,* etc.). No examples of epistemic stance clusters were found among the top twenty to thirty LBs from each move in the two corpora.

Research-related topic LBs are those LBs that signal the topic or the focus of the research study such as *English language learning, the acquisition of, learners of English* and so forth.

Complete lists of all the LBs identified in NAC and PAC, arranged according to their functions, can be found in Appendix L.

7.4 Comparisons of Structural and Functional Analysis of LBs in the NAC and the PAC

7.4.1 Structural and functional analysis of m1: Introduction

A. Structural Analysis of m1

Table 7.5 shows that structurally the most prevalent LBs in m1 of the NAC are NP-based, accounting for more than half of the total (57.53%), followed by VP-based (17.81%), PP-based (16.44%), and Others (8.22%). Similarly, NP-based LBs are also the most common type of cluster in m1 of the PAC (59.38%). However, in contrast with the NAC, where VP-based and PP-based LBs appear at a similar proportion (17.81%: 16.44%), the proportion of PP-based LBs (22.66%) in m1 of the PAC is much higher than that of VP-based LBs (7.81%). Among three types of NP-based LBs in m1, those with a post-modifier are found to be the most common in both corpora (NAC = 39.73% and PAC = 42.19%). It is notable that *of*-phrase LBs are not found in the PAC, with all of the PP-based clusters in that corpus belonging to the other PPs category (22.66%). Despite the proportion of *of*-phrases in the NAC (5.48%), there are fewer than the combined number of other PPs (10.96%). Two types of VP-based LBs are found in each corpus. In the NAC, the VP-active type accounts for 13.70% of clusters, while VP-inf. accounts for 4.11%. VP-active

clusters are also found in the PAC (3.13%), but at a slightly lower proportion than anticipatory *it/there* (4.69%). The remaining clusters, categorized as 'Others', are all clausal (NP + V) in PAC (10.16%), while in NAC they are split between clausal clusters (4.11%) and one type of miscellaneous LB: a phrase acting as a modifier (*very important for*, 4.11%).

Table 7.5 Proportional distribution of lexical bundles in m1 of the NAC and the PAC, categorized by structure

Category	Structural patterns	NAC m1	PAC m1	Examples
1. **NP-based**	NP with pre-modifier	10 (13.70%)	8 (6.25%)	*an important role* (NAC) *a second language* (PAC)
	NP with post-modifier (fragment)	29 (39.73%)	54 (42.19%)	*one of the most* (NAC) *a number of* (PAC)
	NP pre- and post-modifiers	3 (4.11%)	14 (10.94%)	*important role in* (NAC) *second language l* (PAC)
	Subtotal	**42 (57.53%)**	**76 (59.38%)**	
2. PP-based	*of* + NP fragment	4 (5.48%)	-	*of the most important* (NAC) -
	Other PPs	8 (10.96%)	29 (22.66%)	*in terms of* (NAC) *as a second language* (PAC)
	Subtotal	**12 (16.44%)**	**29 (22.66%)**	
3. VP-based	VP-active	10 (13.70%)	4 (3.13%)	*is one of the* (NAC) *have investigated the* (PAC)
	Anticipatory ***it/there*** + verb	-	6 (4.69%)	- *there has been* (PAC)
	(V) + to infinitive	3 (4.11%)	-	*to communicate with* (NAC)
	Subtotal	**13 (17.81%)**	**10 (7.81%)**	
4. Others	NP + V	3 (4.11%)	13 (10.16%)	*who work in* (NAC) *little is known about* (PAC)
	Miscellaneous (modifier)	3 (4.11%)	-	*very important for* (NAC) -
	Subtotal	**6 (8.22%)**	**13 (10.16%)**	
	Total	**73 (100%)**	**128 (100%)**	

B. Functional Analysis of m1

As shown in Table 7.6, three functional categories of LBs are found in m1 of the two corpora, namely *referential*, *discourse/textual organizers* and *research-related topics*. In terms of the distribution, of all the LBs identified in the NAC, 67.65% of them can be classified as referential, while in the PAC it is 48.51%. In the NAC, LBs functioning as discourse/text organizers are found with a similar proportion to those related to research

topic (15.69%:16.67% respectively). The notable thing is that, in contrast with the NAC, a much higher proportion of PAC LBs can be categorized as research topic-related: they account for 47.01% of PAC LBs, much higher than the 16.67% in the NAC, and only slightly lower than the proportion of PAC LBs that are referential (48.51%).

Table 7.6 Proportional distribution of lexical bundles in m1 of the NAC and the PAC, categorized by function

Category	Functions	NAC	PAC	Examples
		m1	m1	
1. Referential	Intangible framing	5 (4.90%)	5 (3.73%)	*in terms of* (NAC) *the context of* (PAC)
	Identification/ identify specification	29 (28.43%)	12 (8.96%)	*is one of the* (NAC) *there has been* (PAC)
	Quantity/specification	29 (28.43%)-	10 (7.46%)	*one of the most** (NAC) *a number of* (PAC)
	Referential text/group	-	5 (3.73%)	- *previous research has* (PAC)
	Referential time/sequence	-	4 (2.70%)	- *in recent years* (PAC)
	Reporting/ description/ interpretation	6 (5.88%)	12 (8.96%)	*who work in* (NAC) *have investigated the* (PAC)
	Multifunctional	-	17 (12.69%)	- *the effects of* (PAC)
	Subtotal	**69 (67.65%)**	**65 (48.51%)**	
2. Discourse/ Textual organizers	Linking	3 (2.94%)	6 (4.48%)	*due to the* (NAC) *as well as* (PAC)
	Attitudinal/Modality Stance	13 (12.75%)	-	*important role in* (NAC)
	Subtotal	**16 (15.69%)**	**6 (4.48%)**	
3. Research topic-related		17 (16.67%)	63 (47.01%)	*the english language* (NAC) *second language esl* (PAC)
TOTAL		**102* (100%)**	**134* (100%)**	

Note: one of the most (raw freq. = 6) is classified as both '*Identification/ identify specification*' and '*Qualifying*'; thereby making the total number increases from 59 to 65.

Some concordance lines taken from the NAC and the PAC are shown below, to illustrate specific examples of the different types of LBs and indicate how they are used in the context of the introduction move. The examples are randomly selected to show at least two different LBs in each sub-category, except where there is only one relevant LB, in which case two instances of the same bundle are presented (see Appendix M for the identified list of LBs).

C. Concordance Analysis of m1

NAC: NP-based

NP with pre-modifier

(1) *… is vital in order for EFL learners to master the English **language**….*
(2) *… had discovered ten strategies to motivate English language **learning** in Hungary.*

NP with post-modifier

(3) *Reading is **one** of the most important language skills for ensuring success…*
(4) *Effective **use** of reading strategies will certainly enhance the reader's comprehension.*

NP with pre- and post-modifiers

(5) *Listening, the most frequently used skill, plays an important **role** in daily communication.*
(6) *At present, reading plays an important **role** in all EFL/ESL students' academic lives.*

NAC: PP-based

PP-of phrase

(7) *Motivation is one **of** the most important factors for …,*
(8) *English is considered one of the most important international languages ….*

Other PPs

(9) ***Due** to the need for implementation of alternative ways of teaching….*
(10) *This, together with an impact of the AEC **in** terms of the free flow of human resources…*

NAC: VP-based

VP-active

(11) *News writing **is** one of the genre that has its own conventions.*
(12) *E-mail **is** the most popular channel that ….*

(V)+ to inf.

(13) *…that employees use **to** communicate with their Swiss suppliers and customers.*
(14) *…especially people who have **to** communicate with foreigners on a regular basis.*

NAC: Others

NP + V

(15) *English communication skills are very important for employees **who** work in multinational companies, …*

Miscellaneous

(16) *It is very **important** for teaching assistants to be able to …*

PAC: NP-based

NP with pre- modifier

*(17) …languaging plays a crucial role in learning a second **language** (L2).*

*(18) Immigrant success stories found in English as a second **language** (ESL)*

NP with post-modifier

*(19) The practicum is **one** of the most important aspects of …*

*(20) Although these rich descriptions have illustrated **a number** of important functions …,*

NP with pre- and post-modifiers

*(21) Few second-**language** (L2) reading studies have examined the relationship…*

*(22) Although several studies have been conducted that investigated the attitudes of English as a second **language** (ESL)…*

PAC: PP-based

PP-others

*(23) …school, however, rarely have room for extensive English **as** a second language instruction*

*(24) …on different aspects of their classroom teaching **as** well as having the opportunity to…*

*(25) Feedback in writing has **in** recent years attracted the attention of ….*

PAC: VP-based

VP-active

*(26) Several studies **have** investigated the effect of dictionary use on …; however, …*

*(27) However, few studies to date **have** investigated the use –and effects of –English…*

Anticipatory *it/there* + verb

*(28) To date, **there** has been little empirical research exploring …*

*(29) While **there** has been extensive L1 research into the benefits of….*

PAC: Others

NP + V

*(30) …, which **previous research has** pinpointed as an important negotiating skill.*

*(31) Several previous **studies have investigated** the use of questions to …*

*(32) …, very **little is known** about the marking of relevance in this genre.*

The NAC examples above indicate that novice writers provide the motivations of their research by highlighting how important the topic is. This is shown in LBs such as *one of the, is one of the, one of the most, of the most important, important role in, is the most, very important for*. Apart from highlighting the research significance of the topic, the occurrence of some LBs such as *English as a, English language learning, second language l, use of reading strategies, as a second, as a second language, a second*

language ESL, English for Specific Purposes and *a foreign language* gives the readers an indication of what the topic is, though research topic-related LBs are relatively infrequent in the NAC, compared with the PAC. Many of these LBs are among the top twenty most frequent (see Appendix M for details). Then, the specific focus can be narrowed down with LBs such as *in terms of, who work in.* The contexts of LBs in some of the examples (10, 15 and 16) reveal issues current at the time of the data collection and the area of the program of the study, reflecting the particular interests of a small culture within the wider academic community. For example, the LB *in terms of* in example 10 shows how the founding of the ASEAN Economic Community (AEC) plays a role in human resources recruitment. The LBs *very important for* and *who work in* in examples 15 and 16 show the relationship between a study's topic and its participants or target readership. This is a specific characteristic of the abstracts associated with the Master's Degree in English for Careers program, in which the students usually conducted a study related to their careers.

Although texts in both NAC and PAC contain LBs that direct their readers to the significance of the studies, it is worth noting that the PAC concordance lines clearly reveal how professional writers identify not only the importance of the research but also the gap that it fills in relation to previous research. Three linguistic features can be identified as techniques that help achieve these two aims: (1) the use of present perfect tense in many contexts (examples 20, 21, 22, and 25-31); (2) the use of *contrastive markers*, such as *although, however, while* appearing in the context near the LBs, as in examples 20, 22, 23, 27, and 29; and, (3) the use of *adverbs, pronouns* and/or *determiners* that denote a small amount, such as *few, little, rarely*, as in examples 21, 23, 28 and 32. This contrasts with what is seen in the NAC texts, where attention is certainly placed upon the importance of the study, but the gap in research is rarely identified. Concordance analysis shows that the novice writers also use present perfect construction in m1, but with much lower frequencies compared with the professional writers (NAC = 32 occurrences, PAC =133 occurrences). With regard to the use of contrastive markers, the words *while*, *although*, and *however* was found to have zero occurrence, four occurrences and seven occurrences respectively in m1 of the NAC. However, these words are found with higher frequencies in the PAC (*while* = 15 occurrences, *although* = 13 occurrences and *however* = 26 occurrences). Moreover, when the words denoting a small quantity are examined, only two occurrences of *few*, one occurrence of *little* and zero occurrence of *rarely* are found in m1 of the NAC. In contrast with the novices, professional researchers are demonstrating

their expertise by citing relevant literature and pointing out the contribution that their own research makes relative to previous studies.

7.4.2 Structural and functional analysis of m2: Purpose

A. Structural analysis of m2

As shown in Table 7.7, the most frequent LBs in m2 in the NAC belong to the category of 'Others', specifically in the subcategory of 'NP + V'. It is notable that this is not in line with the observation made by Chen and Baker (2010) that LBs in academic prose are more likely to be phrasal than clausal, though this was found to be applicable to m1 (see Table 7.5). NP-based and VP-based clusters occur at a similar proportion (21.70% and 23.96%), followed by PP-based (13.77%). Two types of VP-based LBs are found in this move in the NAC. However, it is worth noting that (V) + *to* inf. clusters are found with a much higher proportion, compared with VP-active (21.89%: 2.80%). *Of*-phrases are found with a higher proportion than other PPs (11.32%: 2.45%), which is parallel to what was seen in m1 of the NAC.

Table 7.7 Proportional distribution of lexical bundles in m2 of the NAC and the PAC, categorized by structure

Category	Structural patterns	NAC	PAC	Examples
		m2	m2	
1. NP-based	NP with pre-modifier	10 (1.89%)	49 (9.55%)	*this research study* (NAC) *a second language* (PAC)
	NP with post-modifier (fragment)	93 (17.55%)	182 (35.48%)	*the purpose of this* (NAC) *the extent to which* (PAC)
	NP with pre- and post-modifiers	12 (2.26%)	10 (1.95%)	*students attitudes towards* (NAC)* *first language l* (PAC)
	Subtotal	**115 (21.70%)**	**231 (45.03%)**	
2. PP-based	*of* + NP fragment	60 (11.32%)	10 (1.95%)	*of this research* (NAC) *of english as* (PAC)
	Other PPs	13 (2.45%)	84 (12.71%)	*in order to* (NAC) *as a second language* (PAC)
	Subtotal	**73 (13.77%)**	**94 (18.32%)**	
3. VP-based	VP-active	11 (2.80%)	35 (6.82%)	*find out the* (NAC) *reports on a study* (PAC)
	(V) + to infinitive	116 (21.89%)	15 (2.92%)	*was to investigate* (NAC) *to investigate the* (PAC)
	Subtotal	**127 (23.96%)**	**50 (9.75%)**	
4. Others	NP + V	215 (40.57%)	170 (33.14%)	*this study aimed to* (NAC) *this article explores* (PAC)
	Total	**530 (100%)**	**513 (100%)**	

*Note: *Apostrophe is omitted in the clusters 'students attitudes towards', which makes the phrase a three-word cluster. This is because AntConc does not treat (') as part of a word (Anthony, 2013).*

In contrast with the NAC, the most prevalent type of LB in m2 of the PAC is NP-based (45.03%), which was also the case for m1 of the PAC. The 'NP +V' subcategory of 'Others' is also found with a high proportion (33.14%), followed by PP-based (18.32%) and VP-based (9.75%). Among the PP-based, *of*-phrases are found with a much lower proportion than other PPs (1.95%: 12.71%), which is in sharp contrast with the NAC.

B. Functional analysis of m2

Table 7.8 reveals that while LBs related to three functional categories (referential, discourse/textual organizers and research-related topics) are found in the NAC, only two (referential and research-related topics) are represented in the PAC. Of all the LBs identified in the NAC, 95.28% can be classified as referential, with just over 2% in each of the remaining two categories. Looking more closely at the referential subcategories, the highest proportion in the NAC is found to be reporting/description/interpretation (74.15%). In the PAC, the ratio of referential versus research-related topics is 61.99%: 38.01%. Similar to the NAC, the reporting/description/interpretation function dominates the referential category (47.76%).

Table 7.8 Proportional distribution of lexical bundles in m2 of the NAC and the PAC, categorized by function

Category	Functions	NAC m2	PAC m2	Examples
1. Referential	Intangible framing	-	12 (2.34%)	- *the extent to which* (PAC)
	Referential text/group	100 (18.87%)	17 (3.31%)	*of the study* (NAC), *the present study* (PAC)
	Reporting/ description/ interpretation	393 (74.15%)	245 (47.76%)	*to find out* (NAC), *this paper reports on* (PAC)
	Multifunctional	12 (2.26%)	44 (8.58%)	*the use of* (NAC) *the development of* (PAC),
	Subtotal	**505 (95.28%)**	**318 (61.99%)**	
2. Discourse/ Textual organizers	Linking	13 (2.45%)	-	*in order to* (NAC) -
3. Research topic-related		12 (2.26%)	195 (38.01%)	*students attitudes towards* (NAC) *as a foreign language* (PAC)
TOTAL		**530 (100%)**	**513 (100%)**	

C. Concordance analysis of m2

NAC: NP-based

NP with pre-modifier

*(33) This research **study** aims to examine the language needs and problems ….*
(34) The purpose of this research **study** is to investigate …

NP with post-modifier

*(35) The **purposes** of this study are to access the needs for English…*
*(36) The **purposes** of the study were to investigate…*
*(37) The **purposes** of this research study was to …*

NP with pre- and post-modifiers

*(38) The purposes of this study were to investigate students' **attitudes** towards learning …*
*(39) The primary purposes of this study were to identify students' **attitudes** towards using ….*

NAC: PP-based

Of + PP fragment

*(40) The purposes **of** this study are to explore the most common barriers…*
*(41) The main objective **of** this study was to find out whether …*
*(42) The purpose **of** this research is to investigate …*

Other PPs

*(43) The study was conducted **in** order to explore techniques and preparation in …*
*(44)… .analyze the degree of the problems faced by the Thai staff …**in** order to find possible solutions.*

NAC: VP-based

VP-active

*(45) The study aimed to **find** out the frequent errors and the causes of errors ….*
*(46) Moreover, the study sought to **find** out the listening comprehension strategies used by …*

VP-inf.

*(47) The purposes of this study were **to** investigate the ability of …*
*(48) This study aimed **to** examine the reading motivation in English of …*
*(49) The objectives of this study were **to** explore the attitudes of Japanese workers*

NAC: Others

NP + V

*(50) **This study aimed** at examining the types and frequency of vocabulary learning strategies….*
*(51) **This study was** conducted with 40 students at a language institute …*
*(52) This **study aimed to** investigate the problems using the four English communication skills…*

PAC: NP-based

NP with pre- modifier

*(53) …. this study investigates the use of … by learners of Chinese as a second **language**.*

*(54) The present **study** examined the effects of ….*

*(55) This study investigated second language (L2) **learners**' acquisition of Enlgish….*

NP with post-modifier

*(56) …the interrelationships between learners' attributions and self-efficacy and their achievements in learning **English** as a foreign language.*

*(57) …to determine the **effects** of explicit phonics instruction and sight word instruction on…*

*(58) It then investigates the **extent** to which teachers rely on textbooks to …*

NP with pre- and post-modifiers

*(59) … this study specifically traced the effects of learners' first **language** (L1), and their…*

*(60) The study examined whether individual differences in high school first **language** (L1) …*

PAC: PP-based

Of + PP fragment

*(61) This study extends research into the use **of** English as a lingua franca in…*

*(62) … sentences containing filler-gap dependencies by late-learning speakers **of** English as a second language (L2) ….*

Other PPs

*(63) …the use of morphosyntactic particle DE by learners of Chinese **as** a second language.*

*(64) … approach was deployed in this study to account for variation **in** second language (L2)*

*(65) This article reports **on** a study of the actual curriculum-making practices…*

PAC: VP-based

VP-active

*(66) This paper **reports** on a study to assess the reaction of Japanese university students…*

*(67) This paper **reports** on a collaborative study that tracks the development…*

*(68) This paper **reports** on a qualitative exploration of student perceptions of…*

VP-inf.

*(69) This paper reports on a study that was carried out **to investigate** the effect of pair work on ….*

(70) To investigate the role of word recognition in reading, ….

PAC: Others

NP + V

(71) Motivated by both these gaps, ***this study examines*** *how researchers in two paradigms evaluate…*

(72) ***This study investigates*** *the use of this and these as pronouns versus determiners in …*

(73) ***This study explores*** *the influence of the learning environment on the second language acquisition of Swedish.*

The NAC examples illustrate that, in indicating the aims of the study, novice researchers tend to begin with NPs – such as *this research study, the purpose(s) of this (research) study, the purpose of this research, the main objective of the study* – followed by V.to be + to inf. or aim(s) to inf. These LBs clearly signal the function of indicating purposes, facilitating the move identification. The use of NP + (V) *to* inf. is very common in the NAC as shown in the concordance lines and in Table 7.7, accounting for 21.89%.

The pervasiveness of this kind of 'NP + V' construction is also found in m2 of the PAC, but with some differences. Rather than using (V) + *to* inf, various action verbs (i.e., *reports, investigates, examines, describes* and *explores*) are found to immediately follow the NPs (i.e., the paper and the/this (present) study), making the sentences more concise. In other words, verb phrases such as *aim(ed) to/at* are generally absent. Although the present simple tense and past simple tense are found in both corpora, it seems that the past simple tense is more frequently found than the present simple tense in the NAC. In contrast, the present simple tense is more prevalent in the PAC.

7.4.3 Structural and functional analysis of m3: Method

As seen in Table 7.9, the most frequently found LBs in m3 of the NAC belong to the 'NP + V' subcategory of 'Others', accounting for 41.91 %, followed by NP-based (20.32%), PP-based (18.99%), and VP-based (16.47%). In contrast, NP-based is found to be the most prevalent type of LB in the PAC, with a proportion of 55.65%, followed by PP-based (17.65%), 'NP +V' (16.90%) and VP-based (6.03%). Among three types of NP-based clusters, those with post-modifiers are found with the highest proportion in both corpora. While *of*-phrases are more frequently found than other prepositional phrases in the NAC (13.14%: 6.64%), other prepositional phrases occur more often than *of*-phrases in the PAC (15.96%: 2.53%). As for the VP-based LBs, the most prevalent subcategory in the NAC

is (V) + *to* inf (12.59%), while the most prevalent in the PAC is VP-active (3.95%), followed by a slightly lower proportion of (V) + *to* inf (2.37%).

A. Structural analysis of m3

Table 7.9 Proportional distribution of lexical bundles in m3 of the NAC and the PAC, categorized by structure

Category	Structural patterns	NAC	PAC	Examples
		m3	**m3**	
1. NP-based	NP with pre-modifier	49 (6.78%)	78 (12.32%)	*the academic year* (NAC) *a foreign language* (PAC)
	NP with post-modifier (fragment)	86 (11.89%)	226 (35.70%)	*the subjects of* (NAC) *the effects of* (PAC)
	NP-with pre- and post-modifiers	-	65 (10.27%)	- *native speakers of* (PAC)
	Others	18 (2.49%)	-	*mean and standard* (NAC) -
	Subtotal	**153 (20.32%)**	**369 (55.65%)**	
2. PP-based	*of* + NP fragment	95 (13.14%)	16 (2.53%)	*of this research* (NAC) *of english as a* (PAC)
	Other PPs	48 (6.64%)	101 (15.96%)	*in order to* (NAC, PAC) *as a second language* (PAC)
	Subtotal	**143 (18.99%)**	**117 (17.65%)**	
3. VP-based	VP-active	15 (2.07%)	25 (3.95%)	*was a questionnaire* (NAC) *reports on a* (PAC)
	VP-passive	18 (2.49%)	-	*was used to* (NAC)
	(V) + to infinitive	91 (12.59%)	15 (2.37%)	*to find out* (NAC) *to investigate the* (NAC, PAC)
	Subtotal	**124 (16.47%)**	**40 (6.03%)**	
4. Others	NP + V	303 (41.91%)	107 (16.90%)	*this study aimed to* (NAC) *this study examines* (PAC)
	Total	**723 (100%)**	**663 (100%)**	

B. Functional analysis of m3

Table 7.10 demonstrates that, in terms of function, three of the four types of LBs are present in m3 of both the NAC and the PAC: referential, discourse/textual organizers and research topic-related. Referential LBs in the NAC are found with a strikingly high proportion of 94.74%, while discourse/textual organizers and research topic-related LBs are both found in similarly small proportions (2.77% and 2.49% respectively). Of all the referential LBs, 65.01% belong to the reporting/description/interpretation subcategory. In contrast, the most prevalent functional LBs in the PAC are research-topic related

(51.18%), followed by referential LBs (43.76%) and discourse/textual organizers (2.77%). As in the NAC, reporting/description/interpretation LBs are the most frequently found referential subcategory (29.86%).

Table 7.10 Proportional distribution of lexical bundles in m3 of the NAC and the PAC, categorized by function

Category	Functions	NAC m3	PAC m3	Examples
1. Referential	Referential text/group	181 (25.03%)	35 (5.53%)	*of this research (NAC)* *in this study (PAC)*
	Referential time/ sequence	16 (2.21%)	-	*the academic year (NAC)* -
	Reporting/ description/ interpretation	470 (65.01%)	189 (29.86%)	*to analyze the (NAC)* *this study examines (PAC)*
	Multifunctional	18 (2.49%)	53 (8.37%)	*was used to (NAC)* *the use of (PAC)*
	Subtotal	**685 (94.74%)**	**277 (43.76%)**	
2. Discourse/ Textual organizers	Linking	20 (2.77%)	32 (5.06%)	*in order to (NAC)* *as well as (PAC)*
3. Research topic-related		18 (2.49%)	324 (51.18%)	*mean and standard deviation NPAC)* *native speakers of (PAC)*
	TOTAL	**723 (100%)**	**633 (100%)**	

C. Concordance analysis of m3

NAC: NP-based

NP with pre-modifier

*(74) The research **instrument** employed for collecting data was a questionnaire consisting of…*

*(75) This research **study** adopted a mixed-method approach for data collection and data analysis …*

*(76) The subjects used in this study were 240 students studying in the academic **year** 2013 at…*

NP with post-modifier

*(77) The **subject** of this study consisted of 43 graduate students of the Master of Arts …*

*(78) The **purpose** of this study is to survey the attitudes of employees towards English courses at…*

*(79) The **purpose** of this research is to investigate the Thai employees' opinion on the factors affecting motivation in studying English at Siam Denso Manufacturing Co.,Ltd.*

NP-others

*(80) …the results are presented in the form of frequency, percentage, **mean and standard deviation**.*

*(81) … by using the SPSS program and the findings obtained were assessed by using Percentage, **Mean and Standard Deviation** with descriptive statistics (frequency).*

NAC: PP-based

Of + PP fragment

*(82) A questionnaire adapted from …was used as the research instrument **of** this study.*

*(83) …60 foreign travelers who consented to be participants **of** the study in Bangkok and Krabi…*

*(84) The subjects **of** this research study were 60 students of M.3 who were studying in …*

Other PPs

*(85) **In** order to collect the data related to…, questionnaires were randomly distributed to …*

*(86) Thirty-five (35) students were engaged **in** this study, of whom …*

NAC: VP-based

VP-active

*(87) The research instrument **was** a questionnaire asking about understanding of …*

*(88) The instrument utilized in this present study **was** a questionnaire.*

VP-passive

*(89) The SPSS 17 program **was used** to analyze the data in terms of ….*

*(90) A 20-item questionnaire with a 5-point Likert scale **was used** to collect data.*

NAC: Others

NP + V

*(91) **The study was** a cross-sectional survey conducted with 120 Thai high school students…*

*(92) **This study aimed to** evaluate the teenage audience' s satisfaction with English language…*

*(93) **The data was** analyzed by means of frequency, percentage, arithmetic mean and standard deviation.*

PAC: NP-based

NP with pre- modifier

*(94) …. lexicogrammatical problems found in Cantonese English as a second **language** (ESL) learners' written English output.*

*(95) ….use of motivational strategies and students' motivated behavior in the English as a foreign **language** (EFL) context of Iran.*

*(96) The present **study** investigated the effects of …. on 37 Japanese students' English writing ability and motivation over 3.5 years.*

NP with post-modifier

*(97) In our course, through a task-based approach including the **use** of digital recorders for …*

*(98) One hundred and twenty 12-year-old Taiwanese **learners** of English as a foreign language were assigned …*

*(99) ….investigates the development of writing proficiency in **English** as a foreign language (EFL), …*

NP with pre- and post-modifiers

*(100) …participated in direct and semidirect interviews in both Japanese (first language [L1]) and English (second **language** [L2]).*

*(101) …, we selected LOCNESS, a corpus of young adult native **speakers** of English…*

PAC: PP-based

Of + PP fragment

*(102) The participants were 90 learners **of** English as a foreign language, who were randomly …*

*(103) Thirty learners **of** English as a second language were randomly assigned to …*

Other PPs

*(104) Participants (N =106) came from six intact sections of French **as** a second language (FSL) courses.*

*(105) … a study on learning new and relearning forgotten words of French **as** a foreign language …*

*(106) Learners' grammar and vocabulary knowledge, **as** well as listening, writing, and speaking skills were examined…*

PAC: VP-based

VP-active

*(107) This paper **reports** on a case study that investigated…*

*(108) This article **reports** on a qualitative study that explored…*

(V) + to infinitive

*(109) **To** investigate the existence and validity of these beliefs, a series of test…*

(110) The study was designed to investigate the development of third-person singular (3SG) ….

PAC: Others

NP + V

*(111) **This article reports** on a study of the actual curriculum-making practices…*

*(112) …, this **study investigates** the extent to which self-consciousness and assertiveness are...*

*(113) **This study examines** the relationship between EFL teachers' emotional quotient (EQ) and their pedagogical success …*

Three distinctive features are noticeable from the concordance lines of the NAC: the use of 'NP + V', the use of *of*-phrases in reference to text/group, and the details of methodology. The use of 'NP +V' such as *the study was, the study aimed to* and *the data was* can be seen in examples 91-93. It is worth noting that 'V. to be' constructions are more commonly found than action verbs for describing details. Various LBs that refer to text/group are identified in the top thirty clusters in the NAC, such as *of this study, of the study, of this research, in this study*, as shown in examples 76-79, 82-84 and 86. The details of the methodology identified by m3 LBs of the NAC include *was a questionnaire, the subjects of, the research instrument, mean* and *standard deviation* (see Appendix L for more details). It is noticeable that, although questionnaires were frequently used as a research tool in the studies being described, many of the novice writers still refer to their participants or respondents as *subjects*, which reflects the positivist norms in scientific experiments for describing samples.

In contrast, the clusters that are commonly found in the PAC include 'NP-based' structures and the LBs that fit into the functional category of research topic-related. These two features are in fact closely related because instances of NP-based LBs can be identified as research topic-related, such as *English as a, a second language, second language l, English as a second* and *a foreign language*, as seen in examples 94, 95, 100, 102 and 103.

7.4.4 Structural and functional analysis of m4: Product

A. Structural analysis

As shown in Table 7.11, the two prominent structural patterns in m4 of the NAC are 'Others' (33.44%) and NP-based (31.44%), followed by PP-based (21.74%) and VP-based (13.38%). 'NP + V' LBs, a subcategory of 'Others', account for 28.76% of clusters, which is slightly lower than noun phrases with a post-modifier (31.44%). Similarly, 'NP + V' is identified as the most prevalent type of LB in m4 of the PAC, though with a higher proportion of 49.51% in this subcorpus. The second most prevalent type of cluster identified in the PAC m4 is NP-based (24.76%), followed by VP-based (20.70%), and PP-based (8.80%).

Table 7.11 Proportional distribution of lexical bundles in m4 of the NAC and the PAC, categorized by structure

Category	Structural patterns	NAC	PAC	Examples
		m4	m4	
1. NP-based	NP with pre-modifier	-	8 (3.88%)	- *the experimental group* (PAC)
	NP with post-modifier (fragment)	94 (31.44%)	43 (24.87%)	*the results of the* (NAC, PAC) *the use of* (NAC, PAC)
	Subtotal	**94 (31.44%)**	**51 (24.76%)**	
2. PP-based	of + NP fragment	30 (10.03%)	-	*of the study* (NAC) -
	Other PPs	35 (11.71%)	25 (8.80%)	*in addition the* (NAC) *in terms of* (NAC, PAC)
	Subtotal	**65 (21.74%)**	**25 (8.80%)**	
3. VP-based	VP-active	40 (13.38%)	20 (9.71%)	*revealed that the* (NAC) *showed that the* (NAC, PAC)
	Anticipatory **it/there** + verb	-	8 (3.88%)	- *there was a* (PAC)
	Subtotal	**40 (13.38%)**	**57 (20.07%)**	
4. Others	NP + V	86 (28.76%)	102 (49.51%)	*the findings revealed that* (NAC) *analyses revealed that* (PAC)
	Miscellaneous	14 (4.68%)	-	*that the students* (NAC) -
	Subtotal	**100 (33.44%)**	**102 (49.51%)**	
	Total	**299 (100%)**	**206 (100%)**	

B. Functional analysis

Table 7.12 demonstrates that three of the four functional types of LBs can be identified in m4 of the NAC, namely referential, discourse/textual organizers and research topic-related. Of all the LBs identified, by far the most prevalent ones are referential, accounting for 90.11%, followed by equal proportions of discourse/textual organizers and research topic-related. Similarly, three of the four types of functions are found for LBs in m4 of the PAC, but in this case these are referential, discourse/textual organizers and stance/interpersonal/personal expressions. In m4 of the PAC, 91.90% of LBs can be classified as referential, with the highest proportion in the subcategory of reporting/description/interpretation (70.42%). It is notable that some attitudinal/modality stance LBs are found in m4 of the PAC, but not in the NAC.

Table 7.12 Proportional distribution of lexical bundles in m4 of the NAC and the PAC, categorized by function

Category	Functions	NAC m4	PAC m4	Examples
1. Referential	Intangible framing	21 (7.42%)	10 (3.52%)	*in terms of* (NAC, PAC)
	Referential text/group	30 (10.60%)	22 (7.75%)	*of the study* (NAC) *the control group* (PAC)
	Quantity specification	26 (9.19%)	-	*the majority of* (NAC) -
	Reporting/ description/ interpretation	178 (62.90%)	200 (70.42%)	*study revealed that* (NAC) *it was found that* (PAC)
	Multifunctional	-	29 (10.21%)	- *differences in the* (PAC)
	Subtotal	**255 (90.11%)**	**261 (91.90%)**	
2. Discourse/ Textual organizers	Linking	14 (4.95%)	15 (5.28%)	*in addition the* (NAC) *as well as* (PAC)
3. Stance/ interpersonal/ Personal Expressions	Attitudinal/modality stance	-	8 (2.82%)	- *to be more* (PAC)
4. Research topic-related		14 (4.95%)	-	*that the student* (NAC) -
	TOTAL	**283 (100%)**	**284 (100%)**	

C. Concordance analysis of m4

NAC: NP-based

NP with post-modifier

*(114) The **results** of the study revealed that the main factors that caused …*
*(115) **Most** of the participants agreed that TBLT provided English classes with …*
*(116) The **majority** of respondents had high degree of satisfaction towards....*

NAC: PP-based

Of + PP fragment

*(117) The result **of** the study revealed that grammatical errors were based on …*
*(118) The findings **of** the study indicated that listening problems are caused from…*

Other PPs

*(119) NESTs recorded a high degree of satisfaction from the respondents **in** terms of competency…*
*(120) **In** addition, the findings also revealed that …*
*(121) **In** addition, the data from COCA show more varieties of strategies, …*

NAC: VP-based

VP-active

*(122) The findings **revealed** that the level of ability about English collocations of the participants…*

*(123) The results obtained from the questionnaires **showed** that the factors causing ….*

NAC: Others

Miscellaneous

*(124) The results showed **that** the students frequently used past simple and present simple, while…*

*(125) Overall, the findings reveal **that** the students had positive attitudes about the benefits of …*

PAC: NP-based

NP with pre- modifier

*(126) ….that participants in the experimental group achieved significantly higher scores on the given task than the participants in the control **group**.*

*(127) As hypothesized, the experimental group significantly outperformed the control **group** on…*

NP with post-modifier

*(128) …interactional metadiscourse is best reflected through the **use** of hedges and boosters.*

*(129) The **results** of the study show that these learners are moderately interested in …*

*(130) Results and **analysis** of the instrument's context suitability and limitations are discussed below.*

PAC: PP-based

NP + V.

*(131) The **results showed** that French-English bilingual preschoolers were less accurate with …*

*(132) Our **findings suggest** that strong, local, durable networks are crucial to enabling …*

*(133) Our **results suggest** that contextual support from the stories may have surmounted much …*

Other PPs

(134) … .differences between the programs in expectations and genre sets, as well as illuminating ….

(135) … pointed to L2 writers' difficulties in lexical knowledge and use, as well as possible sources of …

(136) …professional information about the learners in terms of their future language requirements.

PAC: VP-based

VP-active

*(137) The results **show** that the effect of multimedia on the recall of …*

*(138) Inspection of pairwise comparisons seemed to **indicate** that the difference in gains between…*

Anticipatory it/there + verb

*(139) …, **it was found** that guessing behaviors can be influenced by factors such as ….*

*(140) **It was found** that fluency improved for both groups during training but …*

PAC: Others

NP + V

*(141) The **findings indicate** that it can be difficult for gay immigrant (and international) students to …*

*(142) **The results show** that knowledge of the most frequent 3500 word families and ….*

*(143) **The results showed** that low-education learners behaved differently in the classroom than…*

The concordance lines from the NAC show that a number of novice researchers start their m4 with one of the following NP clusters: *the result of*, *the results of the* and *results of the study*. These clusters are often a part of NP + V clusters such as *study revealed that* and *of the study revealed*, as well as a part of VP-based such as *revealed that the* and *showed that the*. In addition, *of*-phrases LBs are often used in references to the text or group such as *of the study* (examples 114, 117 and 118). In contrast, although 'NP + V' clusters are the most prevalent in the PAC, they are less often accompanied by *of*-phrases LBs, which can make the sentences more succinct. Compared with the NAC, in which the top two clusters are *of the study* and *the results of*, the top two most frequently occurring clusters in the PAC are *results show that* and *results showed that* (see Appendix M for more details). In terms of action verbs accompanying noun phrases, *revealed* and *showed* are among the most frequent in the NAC, while a greater variety of action verbs such as *show(ed), indicated, suggest* and *found* are among the top twenty LBs of the list. This also reflects the fact that the past simple tense is more likely to be found in the NAC when writers are describing their findings (see examples 114-120 and 122-124). In contrast, both the present simple and the past simple tenses are found in the PAC.

7.4.5 Structural and functional analysis of m5: Conclusion

A. Structural analysis of m5

As shown in Table 7.13, among the four structural patterns identified in m5 of the NAC, the most prevalent clusters are NP-based (38.96%), followed by PP-based (33.77%), VP-based (20.78%) and 'Others' (6.49%). Of the three NP-based subcategories, noun phrases with a post-modifier are most frequently found, which is similar to other four moves discussed above. Both *of*-phrases and other prepositional phrases are found in the same proportion (16.88%). In addition, three types of VP-based LBs – VP-active, VP-passive and (V) + inf – are found, though in slightly different proportions (5.84%, 7.14% and 7.79% respectively). Clausal LBs are less likely to be found in m5 of the NAC, accounting for only 6.49% of clusters.

Table 7.13 Proportional distribution of lexical bundles in m5 of the NAC and the PAC, categorized by structure

Category	Structural patterns	NAC	PAC	Examples
		m5	m5	
1. NP-based	NP with pre-modifier	4 (2.60%)	-	*the present study* (NAC) -
	NP with post-modifier	52 (33.77%)	95 (38.15%)	*findings of this study* (NAC) *the implications of* (PAC)
	Others	4 (2.60%)	-	*teaching and learning* (NAC) -
	Subtotal	**60 (38.96%)**	**95 (38.15%)**	
2. PP-based	*of* + NP fragment	26 (16.88%)	24 (9.64%)	*of the study* (NAC, PAC) *of the findings* (PAC)
	Other PPs	26 (16.88%)	30 (12.05%)	*for further research* (NAC) *as well as* (NAC, PAC)
	Subtotal	**52 (33.77%)**	**54 (21.69%)**	
3. VP-based	VP-active	9 (5.84%)	14 (5.62%)	*will be beneficial* (NAC) *suggest that the* (PAC)
	VP-passive	11 (7.14%)	8 (3.21%)	*can be used* (NAC) *based on the* (PAC)
	(V) + infinitive	12 (7.79%)	6 (2.41%)	*to improve their* (NAC) *need to be* (PAC)
	Subtotal	**32 (20.78%)**	**28 (11.24%)**	
4. Others	NP + V	10 (6.49%)	157 (34.20%)	*this study will* (NAC) *the article concludes with* (PAC)
	Total	**154 (100%)**	**249 (100%)**	

Similarly, the most prevalent LBs in the PAC are also NP-based (38.15%), followed by clausal LBs of the 'NP +V' type (34.20%), PP-based (21.69%) and VP-based (11.24%). Among the three subcategories of VP-based clusters, the writers represented in the PAC seem to rely on VP-active the most (5.62%), rather than VP-passive (3.21%) or (V) + inf (2.41%).

B. Functional analysis of m5

Table 7.14 shows that all four functional categories of LBs are present in the NAC: referential (74.03%), stance/interpersonal/personal expressions (14.29%), discourse/text organizers (9.09%), and research related-topic clusters (2.60%). Of all the referential LBs, the most prevalent subcategory is reporting/description/interpretation (36.36%) and those referring to text/group (26.62%). Similarly, referential LBs are the most frequently found clusters in the PAC, accounting for nearly 90%, while the remaining types are found in very low proportions: 5.62% for LBs functioning as discourse/textual organizers and 2.41% each for LBs functioning as stance/interpersonal/personal expressions and research topic-related.

Table 7.14 Proportional distribution of lexical bundles in m5 of the NAC and the PAC, categorized by function

Category	Functions	NAC m5	PAC m5	Examples
1. Referential	Intangible framing	-	13 (5.22%)	- *the nature of* (PAC)
	Referential text/group	41 (26.62%)	24 (9.64%)	*from this study* (NAC) *of these findings for* (PAC)
	Referential time/ sequence	5 (3.25%)	5 (2.01%)	*for further research* (NAC) *for future research* (PAC)
	Reporting/ description/ interpretation	56 (36.36%)	155 (62.25%)	*results of the study* (NAC) *indicated that the* (PAC)
	Multifunctional	12 (7.79%)	26 (10.44%)	*the development of* (NAC) *the use of* (NAC, PAC)
	Subtotal	**114 (74.03%)**	**223 (89.56%)**	
2. Discourse/ Textual organizers	Linking	14 (9.09%)	14 (5.62%)	*in order to* (NAC) *as well as* (NAC, PAC)
3. Stance/ Interpersonal/ Personal Expressions	Attitudinal/modality stance	22 (14.29%)	6 (2.41%)	*will be beneficial* (NAC) *need to be* (PAC)
4. Research topic-related		4 (2.60%)	6 (2.41%)	*students attitudes towards* (NAC)* *the acquisition of* (PAC)
TOTAL		**154 (100%)**	**249 (100%)**	

*Note: *Apostrophe is omitted in the clusters 'students attitudes towards', which makes the phrase a three-word cluster. This is because AntConc does not treat (') as part of a word (Anthony, 2013).*

C. Concordance analysis of m5

NAC: NP-based

NP with pre- modifier

(144) *The present **study** is believed to benefit the reinforcement of pragmatic competency…*

(145) *…prove that the M.6 student in the present **study** intentionally avoided using phrasal verbs ….*

NP with post-modifier

(146) *The results of the study indicated that the strategies conducted by SIIT staff will be beneficial …*

(147) *The **findings** of this study will give a clearer insight for pharmaceutical companies and*

(148) *This study also concludes with suggestions for the development of*

NP -Miscellaneous

(149) ***Teaching and learning** methods should concentrate on the vocabulary acquisition, ...*

(150) *…, some suggestions are discussed for the benefit of English **teaching and learning**.*

NAC: PP-based

Of + PP fragment

(151) *The discussion and suggestions at the conclusion **of** the study have been made…*

(152) *The findings **of** this study have significant pedagogical implications for EFL reading strategy instruction in academic context.*

(153) *Regarding the research findings, …. how material design should take into account the various learning styles **of** the students.*

Other PPs

(154) *…acknowledge the importance of English speaking ability and their problems **as** well as to make…*

(155) *The recommendation **from** this study was that the teacher should motivate students to…*

(156) *The results of the study and recommendations **for** further research are also discussed in detail in the paper.*

NAC: VP-based

VP-active

(157) *The information will **be** useful for English language teachers at …*

(158) *However, this study may **be** useful for both students and teachers because ...*

(159) *The findings from this study **will be** beneficial to both Hotel A as well as their employees.*

VP-passive

(160) *In addition, the findings **can be used** as a guideline for …*

*(161) … towards learning English can **be used** to improve the English learning achievement of the students.*

(V) + to infinitive

*(162) … students can make use of the findings to encourage themselves **to improve** their skills.*

*(163) … will be beneficial and useful to others who want **to improve** their English speaking skills effectively.*

NAC: Others

NP + V.

(164) The finding of this study will help teachers to recognize the effectiveness…
(165) The findings of this study will give a clearer insight for …

PAC: NP-based

NP with post- modifier

*(166) The **implications** of the study are discussed.*
*(167) The pedagogical implications of this study highlight the **importance** of ….*
*(168) The paper concludes with a **discussion** of the implication of this for …*

PAC: PP-based

Of + PP fragment

*(169) The implications **of** the findings are highlighted, with suggestions on how language trainers can….*
*(170) The pedagogical implications **of** this study highlight the importance of explicit instruction...*
*(171) We discuss the limitations and implications **of** these findings and point to research questions …*

Other PPs

*(172) …, this study offers insights into the participants' experiences **as** well as the application …*
*(173) The paper ends **with** a discussion of the usefulness and effectiveness of …*
*(174) …,which has various implications **for** future research with regard to the …*

PAC: VP-based

VP-active

*(175) The results **suggest** that the first language is inhibited during…*
*(176) Findings **suggest** that the FLCAS is likely to be measuring…*
*(177) The findings **have** implications for instructors who are interested in investigating…*

VP-passive

*(178) **Based** on the findings, implications are discussed and some possible changes to teaching practices proposed, …*

(179) ***Based*** *on the analysis of narratives selected from 28 hours of videotaped classroom interaction, …*

(V) + to infinitive

(180) It has been noted that learners ***need*** *to be given a more informed and realistic framework…of word recognition in reading, ….*

(181) This highlights the implications that EFL learners ***need*** *to be explicitly taught about …*

PAC: Others

NP + V.

(182) The ***article concludes with*** *a summary of the benefits and constraints of the …*

(183) ***Results are discussed*** *in view of the critical role of noun gender as ….*

(184) These ***findings suggest that the*** *corpus evidence, including frequency data, should be considered when…*

In considering the characteristics of the m5 conclusion moves represented in the NAC, two observations can be made. Firstly, it is notable that LBs such as *the results of (the), results of the study, this study will, findings of this study* and *findings from this study* are commonly found, as in examples 144, 146, 147 and 152. Secondly, modal verbs, such as *will, can* and *may*, are used to show how the findings can be applicable in terms of teaching and learning or future research, as seen in examples 147 and 157-163, while recommendations are made with *should*, as shown in examples 149 and 155.

In contrast, a variety of different observations can be made about the m5 moves contained in the PAC texts. Firstly, a greater variation of NP-based LBs are found at the top of the frequency list, such as *implications of the, the implications of, implications of the study* and *a discussion of*. Secondly, a greater variety of 'NP + V' clusters can be identified, such as *the article concludes (with), findings suggest that, results suggest that, results are discussed* and *findings suggest that the*. In addition, when it comes to phrases that refer to the opportunities that emerge from the study, the cluster *for future research* is found in PAC, which is different from the LB *for further research* which occurs in the NAC. This may suggest the stylistic difference. Moreover, LBs in the top 30 of the PAC list reveal a number of clusters related to implications, but these clusters are not among the LBs recorded in the list of most frequent clusters in the NAC. It is also noteworthy that present simple tense is more prevalent in m5 of the PAC than in the NAC. Finally, unlike the NAC, in which the LB *the study will…* is found, the verbs *conclude* and *suggest* are more common in m5 text of the PAC. This suggests that the professional researchers are more cautious than the novice researchers by avoiding to make a strong claim with *will*.

7.5 Summary

This chapter has reviewed some important characteristics of lexical bundles (LBs) and described the three criteria involved in the retrieval of LBs from the NAC and PAC texts. These criteria are the frequency cut-off thresholds, the length of the clusters and their distribution within the corpora. With respect to the frequency criterion, a number of methodological issues have been addressed. First, the commonly used normalized frequency thresholds are not applicable to this study. This is due to the relatively small sizes of the subcorpora. For example, if a normalized frequency of 40 occurrences per million words (pmw) was used, the corresponding raw frequency would be only 1 (rounding down from 1.33). Therefore, the decision was initially made to use ratio-based cut-offs (1000 words per 1 LB), rather than pmw cut-off thresholds. However, this led to only a very small number of LBs being retrieved from the PAC subcorpora (see Section 7.2.1, Table 7.3c). As a result, the approach that was finally settled on was to retrieve the top 20 to 30 LBs, in accordance with the number of LBs that had been identified in the NAC. The distribution criterion was used to guard against LBs being identified on the basis of their repeated occurrence in only a few texts, and the criterion for each move was set in accordance with the frequency cut-offs (see Section 7.2.3). As regards the length of the LBs, only three- and four-word clusters were examined in this study. Because some three-word LBs were found to be part of four-word LBs, the exclusion of some three-word LBs was necessary in the case of complete overlap. However, in cases of complete subsumption, where the frequency of apparently related three-word and four-word clusters might be different, the decision to retain both the three-word and the four-word LBs was made.

Having discussed and determined the LB retrieval criteria, this chapter then described the structural and functional categories into which LBs can be classified. After that, the LBs retrieved from the different moves in the NAC and the PAC were classified according to these structural and functional types, in order to highlight similarities and differences between the two corpora. Regarding the four structural categories (NP-based, PP-based, VP-based and Others) of the two corpora, the two most commonly found LBs tend to fall into either NP-based or ‘Others’ categories. The most prevalent LBs in m1 and m5 of the NAC were found to be NP-based (m1 = 57.53% and m5 =38.96%), while the most commonly found LBs in m2 and m3 appeared to be ‘Others’, or specifically ‘NP + V’

subcategory (m2 = 40.57% and m3 = 41.91%). In m4 of the NAC, however, NP-based LBs and those classified as 'Others' were found with relatively similar proportion (NP-based = 31.44% and Others = 33.44%). In contrast, the most prevalent types of LB in m1, m2, m3 were found to be NP-based, with the exception of m4, where the subcategory 'NP +V' under 'Others' was found to be more dominant (49.51%). In m5 of the PAC, NP-based LBs and 'NP +V' were found at somewhat a similar proportion (NP-based = 38.15% and NP +V =34.20%). With regard to the functional categories (Referential, Discourse/Textual Organizers, Stance/Interpersonal/Impersonal Expressions, and Research-related topic LBs), referential LBs were found to be the most salient in every move of the NAC (m1 = 67.65%, m2 =95.28%, m3 = 94.74%, m4 = 90.11%, and m5 = 74.03%). In contrast, referential LBs were the most commonly found LBs in m1(48.51%), m2 (61.99%), m4 (91.90%), and m5 (89.56%), with the exception of m3 where research topic-related LBs were the most pervasive (51.18%). The detailed summary can be found in Section 8.2.4. It should be noted that, in terms of frequency of occurrence, direct comparisons between what was found in the NAC and what was found in the PAC are not appropriate, because of the different raw frequencies that needed to be used in the identification of the LBs. However, the retrieval of the LBs from each subcorpus was useful in enabling the most frequently found structural and functional patterns in the NAC and the PAC to be identified, examined and compared. Evidence of contexts in which notable LBs were used in each move of the two subcorpora was provided in the form of concordance lines.

Chapter 8. Discussion and Conclusions

8.1 Introduction

This final chapter discusses the findings presented in Chapters 5-7 in relation to the issues identified in the literature review in Chapters 2 and 3. Section 8.2 begins with a discussion of the key findings, addressing each of the research questions (RQs). After that, Section 8.3 discusses some methodological issues associated with the corpus-based and corpus-driven approaches that were adopted in the study, including the identification of keywords and of lexical bundles (LBs). The limitations of the study are highlighted in Section 8.4, and the implications in Section 8.5. Section 8.6 summarizes the contributions of the study, while Section 8.7 presents some concluding remarks. Sections 8.8 provides reflections on the research process.

8.2 Discussion of Key Findings

The present study examined the rhetorical move patterns of abstracts written by novice and professional researchers in Education and Applied Linguistics and explored the different ways in which these researchers use lexis to achieve their communicative purposes. The main argument arising from the study is that the similarities and differences between the move patterns and the lexis found in the novice abstract corpus (NAC) and the professional abstract corpus (PAC) can best be analyzed by using a mixed-methods approach, combining a genre-based move analysis with a corpus-based analysis. A qualitative move analysis, based on Hyland's (2000) five move model, was used to answer research question 1. In terms of the lexis analysis, multiple corpus techniques, namely, the creation of wordlists, and the extraction of keywords and clusters/n-grams (LBs) were employed to address research questions 2-4. The following sections discuss the findings that relate to each of the research questions.

8.2.1 Similarities and differences between the rhetorical moves of abstracts written by novice and professional researchers

RQ1: What are the similarities and differences between the rhetorical move compositions and move patterns found in abstracts written by novice researchers and those written by professional researchers in the fields of Education and Applied Linguistics?

In this study, the novice abstract corpus (NAC) and the professional abstract corpus (PAC) were analyzed from three different perspectives: (1) move compositional types, (2) rhetorical move patterns, and (3) the status of each of the moves as optional, conventional or obligatory. The textual move analysis was based on Hyland's (2000) five-move model, consisting of Introduction (m1), Purpose (m2), Method (m3), Product (m4), and Conclusion (m5).

Perspective 1: Move compositional types

With regard to the issue of the move compositional types, some similarities were found between the four most frequently found compositional types in the two corpora, that is {2, 3, 4, 5}, {2, 3, 4}, {1, 2, 3, 4, 5} and {1, 2, 3, 4}. Only one difference between these four compositions could be identified, in that while composition {1, 2, 3, 4, 5} was found to be the second most common in the PAC, it was the third most common composition in the NAC. The results partially support Hyland's (2000) and Ren and Li's (2011) studies. While Hyland (2000) and Ren and Li (2011) found that composition type {2, 3, 4} is the most common in research abstracts (RABs), this was identified only as the second most common type in the NAC and the third in the PAC. This suggests that in abstract writing, where conciseness is deemed important, the Purpose, Method and Product moves are found to be very important in Education and Applied Linguistics fields, and in the other disciplines examined by Hyland (2000). Ren and Li (2011) claim that novice writers are likely to be more cautious than professional writers by including all the moves in their abstracts. As evidenced in this study, the composition {1, 2, 3, 4, 5} is identified as only the third most prevalent in the NAC, but the second most prevalent in the PAC. Considering that the proportion of this composition shows a slight difference between the NAC and the PAC (23.33%: 25.66%), it might not be possible to assume that the novice or professional writers are more vigilant in the way that they include all the five communicative purposes (Introduction-Purpose-Method-Product-Conclusion) in their abstract writing.

The results of this study are found to be in line with Saeeaw and Tangkiengsirisin's (2014) cross-disciplinary research on differences in the compositions in RABs. They found that the composition {2, 3, 4, 5} is the most frequently found among RABs in Applied Linguistics while it is the second most common composition in Environmental Sciences RABs.

Perspective 2: Rhetorical move patterns

In relation to the rhetorical move patterns, the results reveal a greater variety of move patterns found in the PAC than in the NAC with a ratio of 163: 37 patterns. Unlike other previous studies (e.g., Dos Santos, 1996; Hyland, 2000; Pho, 2008; Saeeaw and Tangkiengsirisin, 2014), in which embedded moves were only mentioned, but not quantified and included in the move patterns, the identification of move patterns with embedded moves was undertaken in this study. Therefore, the present study addresses a gap in those previous studies. The greater number and the variety of move patterns found both in the NAC and the PAC, in comparison with previous findings, is likely to result from the inclusion of embedded moves.

Close examination of the embedded moves of the PAC has found that the professional writers tend to compose their abstracts with more complex move structures than the novice writers, as shown in the following examples of move patterns found in the PAC but not the NAC: 2(3)-3-4-4(5)-5, 1-2(3)-3-3(5), 2(3)-3-3(4)-5(4), etc. (see Appendix G). In contrast, where their abstracts exhibit embedding, the majority of novice writers are likely to have the Method Move attached to the Purpose Move, as in patterns 2(3)-3-4-5, 2(3)-3-4, and 1-2(3)-3-4-5, with a much lower proportion combining the Product Move with the Conclusion Move, as in pattern 2(3)-3-4-4(5). Despite the embedded moves found in both NAC and PAC, the qualitative analysis reveals that the novice writers did not employ embedded moves to achieve conciseness, as the information presented in the embedded moves tends to be repeated in subsequent sentences. The linking of m3 and m2 is common because the aims of the study are likely to be combined with descriptions of the participants and the research tools. The combination of Product and Conclusion Moves is also found because abstracts can be considered as a summary of the whole article, where these two communicative purposes are likely to be found together in the same chapter of the full research article. According to Bhatia (2014, p. xii), expert members of the discourse community are likely to break the conventional practices to create a ‘private construct’. The issue of ‘private construct’ creation is reflected by the much larger number of move patterns found in the PAC than in the NAC where complex embedded structures occurred only once.

Perspective 3: Optional, conventional or obligatory status

With regard to the status of moves and the issue of whether they are optional, conventional or obligatory, in both corpora, m1 was classified as optional, m2 and m3 as obligatory, and m5 as conventional. Only one difference was found between the NAC and the PAC, in relation to m4. While m4 was identified as obligatory in the NAC, it was identified only as conventional in the PAC. This result provides partial support for Hyland (2000), in that m4 is conventional in both his study and the PAC. The possible explanation is that both the PAC abstracts and Hyland's abstracts were drawn from abstracts written by professional writers. It is possible that m4 was found to be obligatory in the NAC because novice writers consider presenting their findings in the abstract to be crucial.

Given that both m1 (Swales, 1990) and m4 (Hyland, 2000) can be used for promotional purposes, the identification of m4 as an obligatory move in the NAC, but only a conventional move in the PAC, is an interesting finding. First, it is possible that the length restrictions in submission guidelines that apply to the abstracts of the PAC play a part in the lack of m4 in some texts. Second, m1 was classified as an optional move in both corpora, suggesting that both novice and professional writers are more likely to foreground and marketize their research with the findings, rather than the introduction. If the results of the current study were compared with those that arise from the application of a four-move-model, such as Swales' IMRD model, where the Introduction Move is combined with the Purpose Move (and called 'Introduction'), the interpretation should be done with care, because it is possible that this kind of Introduction Move might be found to be conventional or obligatory given that it covers the aims of the study.

8.2.2 The most frequently found lexical words in the NAC and the PAC

RQ2: To what extent are the most frequent lexical words in the NAC similar to those in the PAC?

With regard to research question 2, the findings firstly show the top twenty lexical words found in the two corpora before using two statistical values (log-L and log-R) to determine the degree of similarities and differences among those words. Twelve words (i.e., *English*, *students*, *study*, *reading*, *language*, *learning*, *data*, *research*, *results*, *use*, *teachers*, and *findings*) are found in both. Among the twelve words, the log-likelihood value demonstrates that there is no statistically significant difference between the frequencies of

eight of the words, as shown by the keyness values below 6.63. The log-R values show a small degree of difference in the frequency of these eight words in the two corpora, ranging from -0.38 to 0.23 (see Tables 6.3-6.4, Section 6.2.2). In other words, these words were used with some degree of similarity between the two groups. The eight words include *learning*, *research*, *teachers*, *results*, *use*, *writing*, *findings*, and *teaching*. Recall that the closer the value is to zero, the higher the degree of similarity between the two corpora (see Section 4.8.2 for details of the interpretation of log-R values). These words that exhibit a high degree of similarity are examples of what Baker (2011) refers to as 'lockwords'.

Taylor (2018, p. 20) argues that similarity is often neglected in favor of a focus on difference, but that it is necessary to examine both similarity and difference to avoid bias. By focusing only on keywords, which are a reflection of the differences between two corpora, only a 180-degree, rather than a 360-degree visualization of the data is achieved. In this study, this sort of bias might reinforce an assumption that the patterns of lexis found in the NAC share no or only few similarities with those of the PAC. Nevertheless, the results from the wordlists showed that this is not an accurate assumption because the wordlists reveal some similarities between the NAC and the PAC. It was found that among the 20 most frequently found lexical words, 12 words or more than 50% appeared in both corpora. The comparisons of similarity between the NAC and the PAC confirm that the present study has considered not only the differences but also taken similarity into consideration, as suggested by Taylor (2018) and Marchi and Taylor (2018).

A closer analysis of the concordance lines also shows some differences between the two corpora in terms of the usage of these lexical items. Among the eight lockwords, four words (*learning*, *research*, *results* and *findings*) were selected for further analysis, because they were more clearly related to the academic nature of the abstracts than the remaining four (*teachers*, *use*, *writing* and *teaching*). It was found that, although some of the words can function as both nouns and verbs, they are more likely to be found as nouns than verbs in the PAC. This means that professional writers realize the importance of nominalization in academic writing, as found by Biber and Gray (2010). In contrast, novice writers tend to use these words with other functions, such as adjective or pre-noun modifier, rather than as nouns. It can be concluded, therefore, that the degree of similarity in the frequencies of these words is not a guarantee that these words are being used in the same way in the two corpora, for instance with respect to function.

8.2.3 Keywords associated with each move in the NAC and the PAC

RQ3: What keywords are associated with each move in (a) the NAC and (b) the PAC?

As we would expect, the keywords with the highest keyness values in one corpus are found to be those with zero occurrence in another corpus. The markedness of proper nouns, such as the name of the university (*Thammasat* and *Silpakorn*), those referring to nationality or place (*Bangkok, Thai, Thailand*), and keywords indicating the specific characteristics of the move (*spss, convenience, AEC* and *employees*) are in line with Scott's (1999 as cited in Baker, 2004, p. 347) observation that three kinds of keywords are usually found: (1) proper nouns; (2) keywords indicating the 'aboutness' of a specific text; and, (3) high-frequency words, such as *because, shall,* or *already*, that might be suggestive of an individual author's style. The findings are also consistent with Grabowski (2015), in that proper nouns are found among keywords in one type of different texts under analysis.

The use of some keywords, namely *spss* (raw freq =38), *deviation* (raw freq = 27) and *employees* (raw freq = 24), reflects some conventional practice(s) of the novice researchers under study or a small culture that the statistical program used in the research is likely to be mentioned, and the word *employees* reflects the participants involved in the research in the NAC, specifically in English for Careers (CR) program. However, this is not common practice in the PAC. The characteristics of the NAC reflected through these keywords corroborate Kilgarriff's (2009) remark that keyword lists are advantageous in terms of identifying characteristics of each corpus. In addition, the occurrence of *AEC* (ASEAN Economic Community) shows what are among the research interests of these novice researchers, although the raw frequency is only 5, with a range of 3. It is perhaps notable that the word *objectives* was not found in m2 of the PAC at all. A possible explanation could be that the professional writers use some other synonymous words such as *aim(s)* or *purpose(s)* in the Purpose Move. Concordance lines showed there were 13 occurrences of *aim* as a singular noun, not a plural noun. The synonymous words *purpose* appears 12 times and *purposes* 13 times. However, a closer look at the word *purposes* revealed that the word was not used to indicate the aim of the study because it occurs in such phrases as *academic and professional purposes*, *ESP pedagogic purposes* and *English for Specific Purposes*. Qualitative analysis of the abstracts revealed that in many instances of the abstracts, the professional used two types of verbs that indicate 'research act' or 'discourse act' as a main verb (Hyland and Tse, 2005, p. 130). LBs that are found to show the research

act include: *study **investigates** the*, *this study **examines*** and *this study **explores***, while LBs signify the discourse act are: *this article **reports*** and *this article **describes***. This differs from what the novice writers described their research aim(s) because they tend to use the word *aim* either as a main verb or as a subject followed by these two types of verbs as to infinitive such as *study aimed **to investigate**, **to examine** the* and *was **to investigate***, which contributes to longer sentences (see Move 2 clusters in Appendix M).

Apart from the zero occurrence keywords that can be seen to signify a small culture of the NAC, a qualitative analysis also shows that some words are literally translated from Thai, such as *Mathayom/Mattayom*, indicating the grade level of students. The explanation here might be that novice writers did not think or expect that their research articles would be of any interest to an international audience. However, in the Information Age it could be argued that texts are more likely to be widely available and are more easily disseminated. Therefore, these words should be avoided in scholarly publications.

8.2.4 Lexical bundles associated with each move in the NAC and the PAC

RQ4: What lexical bundles are associated with each rhetorical move in (a) the NAC and (b) the PAC?

In summary, lexical bundles (LBs) identified in each rhetorical move can be categorized according to their structures and functions. After the adjustment of the proportional ratio cut-off and the exclusion of overlap bundles, 125 LBs were identified in the NAC and 135in the PAC. The structural categorization was adapted from Chen and Baker (2010), while the functional categorization was adapted from Liu (2012). A summary of the most frequently found structural and functional categories and subcategories of each move in the NAC and the PAC are listed in Tables 8.1 and 8.2.

As shown in Table 8.1, NP-based LBs are prevalent in both the NAC and the PAC, especially in m1 and m5. The most common subcategory of NP-based LBs in both corpora is the NP with a post-modifier. Another type of structure, the 'NP + V' subcategory of 'Others', was found to be the most frequent kind of LB in m2 and m4 of both corpora, as well as in m3 of the NAC. This suggests that the assertion made by Biber *et al.* (1999) and Chen and Baker (2010) (i.e. that clausal bundles are more likely to be found in conversation than in academic prose) is not applicable to every move of abstract writing. Overall, while NP-based clusters are more commonly found in the PAC, the clausal 'NP

+ V' LBs are more frequently found in the NAC. On the other hand, the findings do support Biber and Gray (2010) with regard to nominalization in academic writing. The frequent use of 'NP + V' LBs is also in line with Liu's observations about the very high frequency of 'NP + action verb' clusters. However, Liu's findings suggested that the 'NP + V' construction is likely to be followed by a *that* clause. In the present study, although the 'NP + V' constructions in m4 and m5 correspond with Liu's finding, the 'NP + V' constructions in m2 and m3 differ in that they are typically 'NP + V to inf.'. These results highlight the fact that, although the same general construction of LBs might be found in various moves, the specific construction might not be exactly the same, depending on the communicative purposes of each move. Because the overall analyses were limited to textual analysis and corpus analysis without conducting any further interviews, it is impossible to know why each writer used some particular structures as shown in those abstracts. However, one possible explanation could be that these structural differences simply reflect the discrepancies in general English proficiency between the novice and the professional groups.

Table 8.1 A summary of the most frequently found structural category and the accompanying most frequent subcategory of each move

Move	Category	%	Structural patterns	%
NAC m1	NP-based	57.53	NP with post-modifier	39.73
PAC m1	NP-based	59.38	NP with post-modifier	42.19
NAC m2	Others	40.57	NP + V	40.57
PAC m2	Others	33.14	NP + V	33.14
NAC m3	Others	41.91	NP + V	41.91
PAC m3	NP-based	55.65	NP with post-modifier	35.70
NAC m4	Others	33.44	NP + V	28.76
PAC m4	Others	49.51	NP + V	49.51
NAC m5	NP-based	38.96	NP with post-modifier	33.77
PAC m5	NP-based	38.15	NP with post-modifier	38.15

Table 8.2 clearly shows that, in terms of function, referential LBs are the most prevalent type in nearly all the moves, with the exception of m1 and m3 of the PAC. In these two moves, research topic-related LBs are most frequently identified. The prevalence of referential LBs is in line with the findings of Ädel and Erman (2012), and in partial

agreement with Grabowski (2015), where referential LBs are identified as the most frequent in one of the four text types under analysis, and with Chen and Baker (2010), where referential LBs are found to be more common in a corpus of native expert writing than two other corpora (see Section 3.4). When looking at the subcategories of referential LBs, the identification/identity specification category is common in m1 of the NAC, but in other moves, in both corpora, the reporting/description/interpretation subcategory is more common. This suggests that the communicative purpose of each move might be linked with, or reflected in, the use of LBs with different kinds of functions.

Table 8.2 A summary of the most frequently found functional category and the accompanying most frequent subcategory of each move

Move	Category	%	Functional patterns	%
NAC m1	Referential	67.65	Identification/identify specification	28.43
			Quantity/ specification	28.43
PAC m1	Referential	48.51	Multifunctional	12.69
	Research topic-related	47.01	Research topic-related	47.01
NAC m2	Referential	95.28	Reporting/description/interpretation	74.15
PAC m2	Referential	61.99	Reporting/description/interpretation	47.76
NAC m3	Referential	94.74	Reporting/description/interpretation	65.01
PAC m3	Research topic-related	51.18	Research topic-related	51.18
NAC m4	Referential	90.11	Reporting/description/interpretation	62.90
PAC m4	Referential	91.90	Reporting/description/interpretation	70.42
NAC m5	Referential	74.03	Reporting/description/interpretation	36.36
PAC m5	Referential	89.56	Reporting/description/interpretation	62.25

8.2.5 Summary of linguistics patterns found in the research abstracts (RABs)

Overarching RQs:

- Are there any differences between the rhetorical moves and linguistic patterns of the NAC and those of the PAC? If yes, what are their pedagogical implications?

Differences between the rhetorical moves and linguistic patterns of the NAC and those of the PAC can be found. The distinctive characteristics in the RABs of the NAC and the PAC can be categorized by the use of different move patterns and lexical choices. As mentioned in Section 5.4, the three most prevalent move patterns in the NAC are found to be 2(3)-3-4-5, 2(3)-3-4 and 1-2(3)-3-4-5. These three patterns, combined together,

account for more than 66% of the total. Similarly, the top three most frequently found patterns are 2(3)-3-4-5, 1-2(3)-3-4-5, and 2(3)-3-4. The three patterns, however, represent 27.78% of the PAC abstracts overall. It is noticeable that unlike the NAC, where the move patterns cluster around the top three patterns, the PAC consists of a greater variety of patterns with more complex embedded moves. With regard to lexical choices, the multi-word constructions known as LBs signify the characteristics of each move of the two corpora. Structurally, the comparisons between NAC LBs and PAC LBs reveal that both groups tend to rely on NP-based LBs and 'N + V' construction LBs, as shown in Table 8.1 in Section 8.2.4. The difference is found only in m3, where 'N + V' LBs are the most prevalent in the NAC and NP-based LBs are the most frequently found in the PAC. In terms of the functional categories of the LBS, it was found that the most prevalent LBs for both groups are identified as referential LBs. However, in m1 and m3 of the PAC, research topic-related LBs are also prevalent.

8.3 Discussion of Methodological Issues

In this section, two methodological issues are discussed in relation to the literature reviewed in Chapters 3 and 4.

8.3.1 Keyword retrieval

The process of keyword retrieval, whereby words are identified that are significantly frequent in one corpus relative to the another as a reference corpus, can be problematic in some cases. Kilgariff (2009) argues that the statistical measures used in identifying keywords normally yield two possible keyword lists: one from the top high frequency list of Corpus A in relation to Corpus B, and the other from the bottom of the list with words found only once or zero occurrence in Corpus A in relation to Corpus B. In the present study, the higher keyness values are found with the zero occurrences rather than the high frequencies. The zero occurrence keywords reflect a methodological issue in the keyword retrieval before the incorporation of various statistical measures into the latest version of AntConc software that can filter out these zero occurrence keywords.

8.3.2 Lexical bundle retrieval

With regard to lexical bundle extraction, it was found that the normalized frequency-based cut-off thresholds are not applicable to the small subcorpora used in this study. The main

reason for different cut-off thresholds in previous studies might be caused by the sizes of different corpora as suggested by Bestgen (2018) and Biber and Barbieri (2007). A bundle which occurs with a raw frequency of 3 has a normalized frequency greater than 40 times per million words (pmw) in a small corpus of 50,000 words. In the present study, a raw frequency ratio cut-off of 1 LB per 1000 tokens was used, instead of a cut-off based on normalized frequencies. However, this resulted in a very small number of clusters being identified in the PAC. As discussed in Section 7.2.1, the final adjustment was to focus on the top twenty to thirty most frequent LBs from the both corpora, a number that was based on the number of LBs found in the NAC with a ratio of 1 LB per 1000 tokens. This yields a higher number of LBs from the PAC, suitable for analysis. Because of the different corpus sizes, extracting LBs in this way poses some problems, specifically in relation to making direct comparisons between the two corpora. Nevertheless, it was possible to compare which of the structural or functional categories of LBs are the most or the least commonly found within each corpus. It is also noteworthy that when the normalized frequency was initially used for LB retrieval, a much greater number of LBs were found in the NAC than in the PAC. This suggests that the novice writers rely more on these muli-word constructions than profession writers, which confirms previous studies (Hyland, 2008b; Qin, 2014; and Wray, 2002) that apprentice writers rely more heavily in using the formulaic patterns than the more experienced writers.

8.4 Limitations of the Study

8.4.1 Authenticity

In terms of the authenticity of the data, it should be acknowledged that abstracts are commonly proofread and revised prior to final publication. During the process of peer-reviewing, it is possible that the form and content of earlier drafts of the PAC abstracts might have been affected by the feedback from the rigorous peer-review process, as pointed out by Paltridge and Starfield (2016). Similarly, although in some cases it appears that the significance of the abstracts was overlooked by students who therefore paid little attention to revising them, many of the NAC abstracts are likely to have been reviewed by the student's supervisor and perhaps by a proofreader who was a native speaker of English. Therefore, there are some inescapable limits to the authenticity of the data. Nevertheless, the combination of abstracts from both groups still serves the aims of the present study.

8.4.2 Corpus sizes

In terms of the corpus sizes, the small specialized corpora used in this study posed some methodological problems with respect to the retrieval of lexical bundles. These problems might not have arisen if the analysis had been performed with the whole corpora instead of the move subcorpora. Given the differences in corpus size, the frequency of features, such as lexical bundles, in the two corpora are only directly comparable if normalized frequencies are used. Due to the adjustments required for meaningful LB retrieval in the present study, any comparisons of the findings in the two subcorpora should be made with care, as should any direct comparisons with different studies, owing to the inapplicability of the normalized frequency cut-off threshold in this case.

8.5 Implications of the Study

Several implications of the study can be drawn from the findings. These are grouped into pedagogical implications and implications for future research below.

8.5.1 Pedagogical implications

In terms of pedagogical implications, the findings have shown that despite some similarities in the most frequently found move compositional types, the notable differences were found with regard to the rhetorical move patterns and the lexical choices in the two corpora. The quantitative analysis of rhetorical move patterns clearly indicates the ratio difference of 163:37 patterns between the PAC and the NAC (see Section 8.2.1). The qualitative analysis reveals that embedded moves in the PAC were of greater variety and complexity than those in the NAC. While a greater variety and complexity of these move compositions might reflect a more advanced ability of the professional writers, it also reveals the language inferiority of those novices.

The overall findings therefore suggest that abstract writing should be taught explicitly as part of academic writing courses to help prepare Master's degree or doctorate students in disseminating their research. Genre-based pedagogy should be adopted to facilitate the teaching and learning process. In addition, Hyland's (2000) framework could also be adopted in helping students practice observing and analyzing their own abstracts and those written by others as shown in Appendix N. The commonly found move compositions in both NAC and PAC could be adopted as examples in order to: (1) raise the awareness of

both instructors and learners in terms of the variety of patterns found in both corpora, which is in line with the ESP approach; (2) raise the awareness of how similar or different the novice abstracts are in comparison with the professional abstracts; and, (3) design academic writing materials that focus specifically on abstract writing. In terms of conciseness, for example, a comparison of embedded moves in the NAC and the PAC would be useful in helping to pinpoint redundancies in abstracts by the novice writers.

In addition, the frequency lists of words and of multi-word constructions, or LBs, could be used to increase students' repertoire of academic vocabulary. With regard to the keyword lists, those with zero occurrence in the NAC in relation to the PAC could also be used in the development of learning materials and applied in classroom, as illustrations of terms that are not commonly found in abstracts in scholarly publications. Examples retrieved from the NAC, representing the authentic language use of their peers, might well be more engaging for learners than examples invented by their instructors.

8.5.2 Implications for further research

There are a number of directions which future research on this topic might explore. Firstly, embedded moves could be further investigated in texts related to different academic disciplines in order to compare and contrast the results with those of the present study. Secondly, in order to increase the generalizability of the study, the compilation of NAC data should include postgraduate students from other universities. Thirdly, the combined measures of log-L and log-R could be adopted in analyzing the degree of similarity found in two corpora. This combination could extend the current use of these two statistical measures, in order to promote understanding of aspects of corpus analysis other than keyword analysis.

8.6 Contributions of the Study

Despite some limitations, the present study has made several contributions to the growing body of research in ESP/EAP.

8.6.1 Bridging gaps in previous studies

Extensive literature review showed that few studies have focused on the abstracts written by novice and professional researchers. The majority of studies (e.g., Hu and Cao, 2011;

Kanoksilapatham, 2009; Hyland, 2000; Saeeaw and Tangkiengsirisin, 2014) focus their attention on abstracts written by professional researchers by compiling abstracts from international peer-reviewed journals. Therefore, the results of the study with regard to the comparisons of move patterns and lexical usage between novice researchers and professional researchers can be considered a contribution to existing knowledge. In particular, the findings increase our awareness in relation to the similarities and differences of move patterns used by novice and professional researchers in Education and Applied Linguistics. The frequently found move compositions and patterns can be used as model patterns for teaching and learning purposes.

Additionally, numerous previous studies (e.g., Golebiowski, 2009; Kanoksilapatham, 2007; Samraj, 2005) placed importance on steps (the internal organization of each move) but ignored the identification of embedded moves. To bridge this gap, the present study identified the various patterns of embedded moves found in the NAC and the PAC. The findings regarding the most prevalent embedded moves, where m3 is attached to m2 or m4 combines with m5, could be used to raise the awareness of teachers and students about the conventional practices of move embedding. Although embedded moves are found in both corpora, there are differences: they can be seen to contribute to the conciseness of the professional abstracts, while in the novice abstracts they are regularly found to be redundant because the information they contain is repeated in subsequent sentences.

8.6.2 Pedagogical contributions

The findings of the present study suggest that move patterns and lexical bundles could be used in training novice writers, helping them to acquire a larger repertoire of various move patterns and lexical choices in their abstract writing and to develop the conciseness of their writing. The identification of embedded moves in this study can also contribute to existing knowledge of how and where these types of moves commonly occur, and which types of embedded moves are commonly found, so that the patterns can be applied in the teaching and learning of academic writing more generally. Additionally, the NAC keywords with zero occurrence in relation to the PAC could be used in terms of raising awareness what words are not commonly found or are irrelevant in the scholarly publications.

8.6.3 Methodological contributions

Several methodological contributions can be identified. Firstly, the combination of genre-based analysis with corpus-based analysis enables us to have a large number of texts under scrutiny, which increases the generalizability of the study. Secondly, the log-likelihood and log-ratio statistics, which are normally applied to the identification of keywords, can be adopted to investigate the similarity of the words or 'lockwords' occurring in the wordlists of the two corpora. Thirdly, separating keyword lists into those having zero occurrence in relation to another corpus can reveal the lexical words that are specific to each group of the writers.

8.7 Concluding Remarks

The present study has examined the rhetorical move patterns of abstracts written by novice and professional writers in the fields of Education and Applied Linguistics. The ESP approach to genre-based analysis was adopted as a theoretical framework for analyzing the rhetorical move patterns of the abstracts in the NAC and the PAC. Hyland's five-move model was adopted as a framework for analyzing the main communicative purposes of the abstracts. Both Swales (1990) and Hyland (2000) assert that abstracts can be used as a promotional genre where writers promote their work through academic discourse. However, some discrepancy was found in terms of which move consists of the promotional attribute. While Swales (1990) suggests that m1 (Introduction) is likely to be considered as a promotional genre, Hyland (2000) advocates m4 (Product) as having this attribute. However, in the present study m1 was found to be an optional move in both of the corpora, suggesting that the promotional genre is likely to be found in m4. It should also be noted that many of the studies that adopted Swales' (1990) I-M-R-D model for analysis might found that m1 are commonly present. This might be attributed to the fact that in his model, the Introduction Move includes m1 and m2 in this study. Therefore, the interpretation of the findings when comparing with another model or framework should take this issue into consideration.

With regards to lexis, the frequent words, keywords and lexical bundles found in the NAC and the PAC could be applied in a program of learning to help develop students' vocabulary and also to raise their awareness of how they can make lexical choices to suit

their communicative purposes. The keyword lists could be introduced in terms of the significant differences found in the NAC against the PAC or vice versa.

The synergy of genre-based and corpus-based approaches have provided richer insights into academic discourse, specifically research abstracts. Through a labor-intensive genre-based approach, research abstracts written by novice and professional researchers were examined according to their communicative purposes or moves, leading to a better understanding of the complexity of the move compositional types and move patterns used by these researchers. Embedded moves, which have been mentioned in previous studies (e.g. Dos Santos, 1996; Hyland, 2000; Pho, 2008; Saeeaw and Tangkiengsirisin, 2014) but left underexplored, are shown to signify some differences between the two groups, as discussed in Section 5.4. Some methodological issues arising from keywords analysis not only confirmed Scott's (1999. cited in Baker, 2004, p. 347) remark regarding the findings of three typical types of keywords (as discussed in Section 8.2.3), but also made the researcher aware that these keywords do not serve practical teaching purposes. As only textual analysis was carried out without an accompanying ethnomethodology in this study, further research involving interview-based narrative accounts from writers and gatekeepers would be highly recommended. With the interviews of writers, some aspects why these researchers write in a particular way are revealed, and thus promote the understanding of not only the product (texts) but also the process of the text construction. With the interviews of gatekeepers, how important research abstracts are important can be substantiated.

8.8 Reflections on the Research Process

Conducting PhD research has been one of the most challenging tasks I have ever undertaken. Along the journey, I have learned to expect the unexpected, push myself beyond my comfort zone, and devise alternative solutions when things went wrong. I discovered that no matter how hard I tried to control external and internal variables to keep my work on track, a number of unpredictable factors caused me to veer either slightly or considerably off track. However, knowing that a clearer and brighter sky would always await me after a cloudy day, I kept moving forward with patience and determination. Occasionally, an immediate solution could be reached effortlessly if I opened up and shared my problems with others.

On reflection, during the data collection, an alternative plan was adopted when some of the novice abstracts were not retrievable online. It will be more complicated if my future research includes novice abstracts from other universities. Therefore, the harmonious and effective relationships with other researchers that I had developed during my study would help. Although the move tagging was a tiring and laborious process, with a total of 528 abstracts for move division and tagging to be completed, it was an excellent opportunity to notice and learn different styles of research abstracts written by novice and professional researchers. All of them were authentic, not the materials specifically produced for teaching purposes. One of the most challenging tasks for me was how to analyze and interpret my corpus findings, with overwhelming data and three new released versions of the corpus software, resulting in the re-analysis. To illustrate, a decision to change some parts of the research focus occurred when the findings did not appear to support my research aim, which was pedagogically motivated. The re-analysis also happened every time I realized that a new version of the corpus software I was using was launched. The latest version provided a set of effect size statistics that I had already done the calculations using Excel spreadsheet. This made me realize that, when dealing with technology, it is imperative that you keep up-to-date with it and with current trends in the field. In sum, what I have learned reaches far beyond my research topic. The research process has broadened my worldview and gradually transformed me into a more independent, flexible and resilient researcher who is able to deal with uncertainties.

References

Ädel, A. and 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), pp. 81-92.

Anthony, L. (2013) 'A critical look at software tools in corpus linguistics', *Linguistic Research*, 30(2), pp. 141-161.

Anthony, L. (2015). AntFileConverter (1.2.0) [Computer Software]. Tokyo, Japan: Waseda University. Available at: http://www.laurenceanthony.net/

Anthony, L. (2018). AntConc (Version 3.5.2) [Computer Software]. Tokyo, Japan: Waseda University. Available at: http://www.laurenceanthony.net/

Atkinson, D. (2004) 'Contrasting rhetorics/contrasting cultures: Why contrastive rhetoric needs a better conceptualization of culture', *Journal of English for Academic Purposes*, 3(4), pp. 277-289.

Baker, P. (2004) Querying keywords: Questions of difference, frequency, and sense of keywords analysis, *Journal of English Linguistics*, 32(4), pp. 346-359.

Baker, P. (2006) *Using corpora in discourse analysis*. London: Continuum.

Baker, P. (2011) 'Times may change but we'll always have money: A corpus driven examination of vocabulary change in four diachronic corpora', *Journal of English Linguistics*, 39, pp. 65-88.

Baker, P. (2013) 'Corpora and discourse analysis', in Hyland, K. (ed.) *Discourse Studies Reader: Essential Excerpts*. London: Bloomsbury Academic, pp. 11-34.

Baker, P. and Egbert, J. (2016) 'Introduction', in Baker, P. and Egbert, J. (eds.) *Triangulating Methodological Approaches in Corpus Linguistic Research.* New York: Routledge, pp. 1-19.

Baker, P., Gabrielatos, C., KhosraviNick, M., Krzyzanowski, M., McEnery, T. and Wodak, R. (2008) 'A useful methodological synergy? Combining critical

discourse analysis and corpus linguistics to examine discourses of refugees and asylum seekers in the UK press', *Discourse Society*, 19, pp. 273-306.

Baker, P. and McEnery, T. (2015) 'Introduction', in Baker, P. and McEnery, T. (eds.) *Corpora and Discourse Studies: Integrating Discourse and Corpora*. New York: Palgrave Macmillan, pp. 1-19.

Baker, W. (2013) 'Interpreting the culture in intercultural rhetoric: A critical perspective from English as a lingua franca studies, in Belcher, D. and Nelson, G. (eds.) *Critical and Corpus-based Approaches to Intercultural Rhetoric.* Ann Arbor: University of Michigan Press, pp. 22-45.

Bawarshi, A.S. and Reiff, M.J. (2010) *Genre: An introduction to history, theory, research and pedagogy*. Indiana: Parlor Press.

Bestgen, Y. (2018) 'Evaluating the frequency threshold for selecting lexical bundles by means of an extension of the Fisher's exact test', *Corpora*, 13(2), pp. 215-228.

Bhatia, V.K. (1993) *Analyzing genre: language use in professional settings*. Essex: Longman Group Ltd.

Bhatia, V.K. (2012) 'Critical reflections on genre analysis', *Ibérica*, 24, pp. 17-28.

Bhatia, V.K. (2014) *Worlds of written discourse: A genre-based view*. London: Bloomsbury.

Biber, D. (2012). Register as a predictor of linguistic variation. *Corpus Linguistics and Linguistic Theory*, 8(1), pp. 9-37.

Biber, D. and Barbieri, F. (2007) 'Lexical bundles in university spoken and written registers', *English for Specific Purposes*, 26, pp. 263-286.

Biber, D. and Conrad, S. (2009) *Register, genre and style*. Cambridge: Cambridge University Press.

Biber, D., Connor, U. and Upton, T. A. (2007) *Discourse on the move: Using corpus analysis to describe discourse structure*. Philadelphia: John Benjamins Publishing.

Biber, D., Conrad, S. and Cortes, V. (2004) 'If you look at…: Lexical bundles in university teaching and textbooks', *Applied Linguistics,* 25, pp. 371-405.

Biber, D., Johansson, S., Leech, G., Conrad, S. and Finegan, E. (1999*) Longman grammar of spoken and written English*. Essex: Pearson Education Limited.

Biber, D., Reppen, R. and Friginal, E. (2012) 'Research in corpus linguistics', in Kaplan, R. (ed.) *The Oxford Handbook of Applied Linguistics* [Online]. Available at: doi: 10.1093/oxfordhb/9780195384253.013.0038 (Accessed: 10 November 2017).

Bloor, M. and Wood, F. (2006) *Keywords in qualitative methods: A vocabulary of research concepts*. London: Sage.

Bryman, A. (2012) *Research methods* (4th edn.). Oxford: Oxford University Press.

Charles, M. and Pecorari, D. (2016) *Introducing English for academic purposes*. London: Routledge.

Chen, Y. and Baker, P. (2010) 'Lexical bundles in L1 and L2 academic writing', *Language Learning & Technology*, 14(2), pp. 30-49.

Connor, U. (1996) *Contrastive rhetoric: Cross-cultural aspects of second language writing*. Cambridge: Cambridge University Press.

Connor, U., Upton, T.A. and Kanoksilapatham, B. (2007) 'Introduction to move analysis', in Biber, D., Connor, U. and Upton, T. A. (eds.) *Discourse on the Move: Using Corpus Analysis to Describe Discourse Structure*. Amsterdam/ Philadelphia: John Benjamins Publishing, pp 23-42.

Corbetta, P. (2003) *Social research: Theory, methods and techniques*, London: Sage.

Cortes, V. (2004) ‘Lexical bundles in published and student disciplinary writing: Examples from history and biology’, *English for Specific Purposes*, 23(4), pp.397-423.

Cortes, V. (2011) ‘Genre analysis in the academic writing class: with or without corpora?’, *Quaderns de Filologia-Estudis Lingüístics*, 16, pp. 65-80.

Cortes, V. (2013) ‘The purpose of this study is to: Connecting lexical bundles and moves in research article introductions’, *Journal of English for Academic Purposes,*12(1), pp. 33-43.

Creswell, J. W. (2009). *Research design: Qualitative, quantitative, and mixed methods approaches* (3rd edn.) London: Sage publications.

Creswell, J. W. (2012) *Educational research: Planning, conducting, and evaluating quantitative and qualitative research* (4th edn.) London: Pearson.

Dong, H.L. and Huan, X.U.E. (2010) ‘Generic structure of research article abstracts’, *Cross-Cultural Communication*, 6(3), pp. 36-44.

Dörnyei, Z. (2007). *Research methods in applied linguistics*. New York: Oxford University. Press.

Dos Santos, M.B. (1996) 'The textual organization of research paper abstracts in applied linguistics', *Text-Interdisciplinary Journal for the Study of Discourse,* 16(4), pp. 481-500.

El-Dakhs, D.A.S. (2018) ‘Comparative genre analysis of research article abstracts in more and less prestigious journals: Linguistics journals in focus’, *Research in Language*, 16(1), pp. 47-63.

Esfandiari, R. and Barbary, F. (2017) ‘A contrastive corpus-driven study of lexical bundles between English writers and Persian writers in psychology research articles’, *Journal of English for Academic Purposes,* 29, pp. 21-42.

Evison, J. (2010) ‘What are the basics of analysing a corpus?’, in O'Keeffe, A. and McCarthy, M. (eds.) *The Routledge Handbook of Corpus Linguistics.* London: Routledge, pp. 122-135.

Feak, C. and Swales, J. (2009) *Telling a research story: Writing a literature review*. Ann Arbor: Michigan Publishing.

Flowerdew, J. (1999) ‘Problems in writing for scholarly publication in English: The case of Hong Kong’, *Journal of Second Language Writing,* 8(3), pp. 243-264.

Flowerdew, J. (2002) ‘Genre in the classroom: A linguistic approach’, in Johns, A.M. (ed.) *Genre in the Classroom: Multiple Perspectives*. London: Lawrence Erlbaum, pp. 91-102.

Flowerdew, J. (2013) *Discourse in English language education*. London: Routledge.

Flowerdew, J. and Forest, R. W. (2009) ‘Schematic structure and lexico-grammatical realization in corpus-based genre analysis: The case of research in the PhD literature review’, in Charles, M., Pecorari, D. and Hunston, S. (eds.) *Academic Writing: At the Interface of Corpus and Discourse.* London: Continuum, pp. 15-36.

Gablasova, D., Brezina, V. and McEnery, T. (2017) ‘Exploring learner language through corpora: Comparing and interpreting corpus frequency information’, *Language Learning,* 67, pp. 130-154.

Gabrielatos, C., and Marchi, A. (2012) ‘Keyness: Appropriate metrics and practical issues’, CADS International Conference 2012, *Corpus-assisted Discourse Studies: More than the sum of Discourse Analysis and computing?*, 13-14 September, University of Bologna, Italy.

Golebiowski, Z. (2009) 'Prominent messages in education and applied linguistic abstracts: How do authors appeal to their prospective readers?', *Journal of Pragmatics*, 41(4), pp. 753-769.

Grabowski, L. (2015) 'Keywords and lexical bundles within English pharmaceutical discourse: A corpus-driven description', *English for Specific Purposes*, 38, pp. 23-33.

Gries, S. T. (2005) 'Null-hypothesis significance testing of word frequencies: A follow-up on Kilgarriff', *Corpus Linguistics and Linguistic Theory*, 1(2), pp. 277-294.

Gries, S. T. (2009) 'What is corpus linguistics?', *Language and Linguistics Compass*, *3*(5), pp. 1225-1241.

Gries, S.T. (2011) 'Methodological and interdisciplinary stance in corpus linguistics', in Viana, V., Zyngier, S. and Barnbrook, G. (eds.) *Perspectives on Corpus Linguistics: Connections and Controversies*, Amsterdam: John Benjamins Publishing, pp. 81-98.

Grix, J. (2002) 'Introducing students to the generic terminology of social science', *Politics*, 22(3), pp. 175-186.

Guba, E. G. and Lincoln, Y. S. (1994) 'Competing paradigms in qualitative research' in Denzin, N. K. and Lincoln, Y. S. (eds.) *Handbook of Qualitative Research.* London: Sage, pp. 163-194.

Guinda, C.S. (2015) 'Genres on the move: Currency and erosion of the genre moves construct', *Journal of English for Academic Purposes*, 19, pp. 73-87.

Hardie, A. (2014) Log Ratio: An informal introduction. [Online] Available at: http://cass.lancs.ac.uk. (Accessed: 10 April 2017).

Hewings, M. (2006) *Academic writing in context: Implications and applications*, London: Continuum.

Hu, G. and Cao, F. (2011) 'Hedging and boosting in abstracts of applied linguistics articles: A comparative study of English-and Chinese-medium journals', *Journal of Pragmatics*, 43(11), pp. 2795-2809.

Hunston, S. (2002) *Corpora in applied linguistics*. Cambridge: Cambridge University Press.

Hyland, K. (1994) 'Hedging in academic writing and EAP textbooks', *English for Specific Purposes*, 13(3), pp. 239-256.

Hyland, K. (2000) *Disciplinary discourses: Social interactions in academic writing*. Reprint, Michigan: Michigan Publishing, 2004.

Hyland, K. (2004) *Genre and second language writing*. Ann Arbor: University of Michigan Press.

Hyland, K. (2005) 'Stance and engagement: A model of interaction in academic discourse', *Discourse Studies*, 7(2), pp. 173-192

Hyland, K. (2007) 'Genre pedagogy: Language, literacy and L2 writing instruction', *Journal of Second Language Writing*, 16(3), pp. 148-164.

Hyland, K. (2008a) 'As can be seen: Lexical bundles and disciplinary variation', *English for Specific Purposes*, 27(1), pp. 4-21.

Hyland, K. (2008b) 'Academic clusters: text patterning in published and postgraduate writing', *International Journal of Applied Linguistics*, 18(1), pp. 41-62.

Hyland, K. (2009a) *Teaching and researching writing* (2nd edn.). Essex: Pearson Education.

Hyland, K. (2009b) *Academic discourse: English in a global context*. London: Bloomsbury Academic.

Hyland, K. (2012a) 'Bundles in academic discourse', *Annual Review of Applied Linguistics*, 32, pp. 150-169.

Hyland, K. (2012b) 'Genre and discourse analysis in language for specific purposes', in Chapelle, C.A. (ed.) *The Encyclopedia of Applied Linguistics*. Hoboken, N.J.: Wiley-Blackwell Publishing Ltd. [Online]. Available at:

https://onlinelibrary.wiley.com/doi/pdf/ 10.1002/ 9781405198431.wbeal0452 (Accessed: 20 March 2017).

Hyland, K. and Guinda, C.S (2012) *Stance and voice in written academic genres*. Houndmills: Palgrave Macmillan.

Hyland, K. and Milton, J., (1997) 'Qualification and certainty in L1 and L2 students' writing', *Journal of Second Language Writing*, 6(2), pp. 183-205.

Hyland, K. and Tse, P. (2005) 'Hooking the reader: A corpus study of evaluative *that* in abstracts', *English for Specific Purposes*, 24(2), pp. 123-139.

Hyon, S. (1996) 'Genre in three traditions: Implications for ESL', *TESOL Quarterly*, 30(4), pp. 693-722.

Kachru, Y. and Nelson, C. L. (2006) *World Englishes in Asian contexts* (Vol. 1). Hong Kong: Hong Kong University Press.

Kanoksilapatham, B. (2005) 'Rhetorical structure of biochemistry research articles', *English for Specific Purposes*, 24(3), pp. 269-292.

Kanoksilapatham, B. (2007) 'Writing scientific research articles in Thai and English: Similarities and differences', *Silpakorn University International Journal*, 7, pp. 172-203.

Kanoksilapatham, B. (2009) 'Generic structure of research article abstracts in sciences', *Journal of English Studies*, 4, [no pagination].

Kaplan, R.B. (1966) 'Cultural thought patterns in inter-cultural education', *Language Learning*, 16, pp. 1-20.

Kaplan, R.B. (1987) 'Cultural thought patterns revisited', in Connor, U. and Kaplan, R. (eds.) *Writing Across Languages: Analysis of L2 Text*, Reading, MA: Addison-Wesley, pp. 9-21.

Kaur, A., Young, D. and Kirkpatrick, R. (2016) ‘English education policy in Thailand: Why the poor results?’, in Kirkpatrick, R. (ed.) *English Language Education Policy in Asia*, New York: Springer, pp. 345-361.

Khedri, M., Heng, C.S. and Ebrahimi, S.F. (2013) 'An exploration of interactive metadiscourse markers in academic research article abstracts in two disciplines', *Discourse Studies*, 15(3), pp. 319-331.

Kilgarriff, A. (2005) ‘Language is never, ever, ever, random’, *Corpus Linguistics and Linguistic Theory*, 1(2), pp. 263–276.

Kilgarriff, A. (2009) ‘Simple maths for keywords’, in *Proceedings of the Corpus Linguistics Conference. July, Liverpool, UK*.

Kirkpatrick, A. (2014) ‘Redesigning the linguistic ecology of Southeast Asia: English and/or local languages?’, *Journal of English Studies*, 8, pp. 8-28.

Klimova, B.F. (2013) 'Common mistakes in writing abstracts in English', *Procedia-Social and Behavioral Sciences*, 93, pp. 512-516.

Lao, R. (2015) *A critical study of Thailand’s higher education reforms: The culture of borrowing*, New York: Routledge.

Liu, D. (2012) ‘The most frequently-used multi-word constructions in academic written English: A multi-corpus study’, *English for Specific Purposes*, 31, pp. 25-35.

Lorés, R. (2004) 'On RA abstracts: From rhetorical structure to thematic organisation', *English for Specific Purposes*, 23(3), pp. 280-302.

Mack, L. (2010) ‘The philosophical underpinnings of educational research’, *Polyglossia, 19*, pp. 5-11.

Mackey, A. and Gass S.M. (2005) Second language research: Methodology and design. London: Lawrence Erlbaum Associates.

Mahboob, A. and Paltridge, B. (2015) ‘In this issue’, *TESOL Quarterly*, 49(4), pp. 623–625.

Marchi, A. and Taylor, C. (2018) ‘Introduction: Partiality and reflexivity’, in Taylor, C. and Marchi, A. (eds.) *Corpus Approach to Discourse: A Critical Review*. London/ New York: Routledge, pp. 1-15.

Martín, P.M. (2003) ‘A genre analysis of English and Spanish research paper abstracts in experimental social sciences’, *English for Specific Purposes*, 22(1), pp. 25-43.

Matsuda, P.K., Canagarajah, A.S., Harklau, L., Hyland, K., Warschauer, M. (2003) ‘Changing currents in second language writing research: A colloquium’, *Journal of Second Language Writing*, 12, pp. 151-179.

McCarthy, M. (2001) *Issues in applied linguistics*. Cambridge: Cambridge University Press.

McEnery, T. and Wilson, A. (1996) *Corpus linguistics*. Edinburgh: Edinburgh University Press.

McEnery, T., Xiao, R. and Tono, Y. (2006) *Corpus-based language studies: An advanced resource book*. London: Routledge.

Melander, B., Swales, J.M. and Fredrickson, K.M. (1997) 'Journal abstracts from three academic fields in the United States and Sweden: National or disciplinary proclivities?', in Duszak, A. (ed.) *Culture and Styles of Academic Discourse*. New York: Mouton de Gruyter, pp. 251-272.

Miller, C. R. (1984) ‘Genre as social action’, *Quarterly Journal of Speech*, 70(2), pp. 151-167.

Miller, C. R. (2015) ‘Genre as social action (1984), revisited 30 years later (2014)’, *Letras & Letras*, 31(3), pp. 56-72.

Northcott, J. (2013) ‘Methods for language for specific purposes’, in Chapelle, C.A. (ed.) *The Encyclopedia of Applied Linguistics*, Hoboken, N.J.: Wiley-Blackwell Publishing Ltd. [Online]. Available at: https://onlinelibrary-wiley-

com.libproxy.ncl.ac.uk/ doi/ 10.1002/9781405198431.wbeal0764. (Accessed: 1 April 2018).

Office of the Higher Education Commission (OHEC) (2016) ‘Thai higher education: Fact and figures’, [Online]. Available at: http://inter.mua.go.th/wp-content/uploads/2016/07/Thai-HE-fact-figures_edited.pdf. (Accessed: 15 April 2018).

Paltridge, B. (1994) ‘Genre analysis and the identification of textual boundaries’, *Applied linguistics*, 15(3), pp. 288-299.

Paltridge, B. (1997) *Genre, frames and writing in research settings*. Amsterdam/Philadelphia: John Benjamins Publishing.

Paltridge, B. (2012) *Discourse analysis: An introduction*. London: Bloomsbury Publishing.

Paltridge, B. and Phakiti, A. (2015) *Research methods in applied linguistics: A practical resource*. http://www.amazon.co.uk/kindle-ebooks (Downloaded: 20 November 2017).

Paltridge, B. and Starfield, S. (2016) *Getting published in academic journals*. Ann Arbor: The University of Michigan Press.

Paquot, M. and Bestgen, Y. (2009) ‘Distinctive words in academic writing: A comparison of three statistical tests for keyword extraction’ in Jucker, A. H., Schreier, D., & Hundt, M. (eds). *Corpora: Pragmatics and Discourse.* Leiden: Brill/ Rodopi, pp. 247-269.

Peacock, M. (2012) ‘High-frequency collocations of nouns in research articles across eight disciplines’, *Ibérica*, 23, pp. 29-46.

Pérez-Llantada, C. (2013) 'Glocal' rhetorical practices in academic writing: An intercultural rhetoric approach to L2 English discoursal hybridisation’, *European Journal of English Studies* 17(3), pp. 261-278.

Pho, P.D. (2008) 'Research article abstracts in applied linguistics and educational technology: A study of linguistic realizations of rhetorical structure and authorial stance', *Discourse Studies*, 10(2), pp. 231-250.

Ponterotto, J.G. (2005) 'Qualitative research in counseling psychology: A primer on research paradigms and philosophy of science', *Journal of Counseling Psychology*, 52(2), pp. 126-136.

Qin, J. (2014) 'Use of formulaic bundles by non-native English graduate writers and published authors in applied linguistics', *System*, 42, pp. 220-231.

Rayson, P. and Garside, R. (2000) 'Comparing corpora using frequency profiling', in proceedings of the *workshop on Comparing Corpora, held in conjunction with the 38th annual meeting of the Association for Computational Linguistics*, 1-8 October 2000, Hong Kong, pp. 1-6.

Ren, H. and Li, Y. (2011) 'A comparison study on the rhetorical moves of abstracts in published research articles and Master's foreign-language thesis', *English Language Teaching*, 4(1), pp. 162-166.

Risager, K. (2006) *Language and culture: Global flows and local complexity*. Bristol: Multilingual Matters.

Saeeaw, S. and Tangkiengsirisin, S. (2014) 'Rhetorical variation across research article abstracts in environmental science and applied linguistics', *English Language Teaching*, 7(8), pp. 81-93.

Sala, M. (2015) 'Knowledge construction and knowledge promotion in academic communication: The case of research article abstracts: A corpus-based study', in Thompson, P. and Diani, G. (eds.) *English for Academic Purposes: Approaches and Implications*, Newcastle: Cambridge Scholars Publishing, pp. 103-125.

Salager-Meyer, F. (1992) 'A text-type and move analysis study of verb tense and modality distribution in medical English abstracts', *English for Specific Purposes*, 11(2), pp. 93-113.

Salazar, D. (2014) *Lexical bundles in native and non-native scientific writing: Applying a corpus-based study to language teaching*. Amsterdam: John Benjamins Publishing.

Samraj, B. (2002) 'Introductions in research articles: Variations across disciplines', *English for Specific Purposes*, 21(1), pp. 1-17.

Samraj, B. (2005) 'An exploration of a genre set: Research article abstracts and introductions in two disciplines', *English for Specific Purposes*, 24(2), pp. 141-156.

Sánchez, P. (2014) 'Lexical bundles in Biology: Differences between textbooks and research articles', *Revista Española de Lingüística Aplicada/Spanish Journal of Applied Linguistics*, 27(2), pp. 492-513.

Schmitt, N. and Carter, R. (2004) 'Formulaic sequences in action', in Schmitt, N. (ed.) *Formulaic Sequences: Acquisition, Processing, and Use.* Amsterdam: John Benjamin Publishing, pp. 1-22.

Scotland, J. (2012) 'Exploring the philosophical underpinnings of research: Relating ontology and epistemology to the methodology and methods of the scientific, interpretive, and critical research paradigms', *English Language Teaching*, 5(9), pp. 9–16.

Shahriari, H. (2017) 'Comparing lexical bundles across the introduction, method and result sections of the research article', *Corpora*, 12(1), pp. 1-22.

Simpson-Vlach, R. and Ellis, N.C. (2010) 'An academic formulas list: New methods in phraseology research', *Applied Linguistics*, 31(4), pp. 487-512.

Soler-Monreal, C., Carbonell-Olivares, M. and Gil-Salom, L. (2011) 'A contrastive study of the rhetorical organisation of English and Spanish PhD thesis introductions', *English for Specific Purposes*, 30(1), pp. 4-17.

Supatranont, P. (2012) 'Developing a writing template of research article abstracts: A corpus-based method', *Procedia - Social and Behavioral Sciences*, 66, pp. 144-156.

Swales, J.M. (1990) *Genre analysis: English in academic and research settings*. Cambridge: Cambridge University Press.

Swales, J. M. (2016) 'Reflections on the concept of discourse community', *ASp. la revue du GERAS*, 69, pp. 7-19.

Swales, J.M. and Feak, C.B. (2009) *Abstracts and the writing of abstracts*. Michigan: The University of Michigan Press.

Swales, J.M. and Feak, C.B. (2012). *Academic writing for graduate students: Essential tasks and skills*. Michigan: The University of Michigan Press.

Tankó, G. (2017) 'Literary research article abstracts: An analysis of rhetorical moves and their linguistic realizations', *Journal of English for Academic Purposes*, 27, pp. 42-55.

Taylor, C. (2008) 'What is corpus linguistics? What the data says', *ICAME Journal*, 32, pp. 179-200.

Taylor, C. (2018) 'Similarity', in Taylor, C. and Marchi, A. (eds.) *Corpus Approaches to Discourse: A Critical Review*. London: Routledge, pp. 19-37.

Tessuto, G. (2015) 'Generic structure and rhetorical moves in English-language empirical law research articles: Sites of interdisciplinary and interdiscursive cross-over', *English for Specific Purposes*, 37, pp. 13-26.

Tognini-Bonelli, E. (2001) *Corpus linguistics at work*. Philadelphia, Amsterdam: John Benjamins.

Tognini-Bonelli, E. (2010) 'Theoretical overview of the evolution of corpus linguistics', in O'Keeffe, A. and McCarthy, M. (eds.) *The Routledge Handbook of Corpus Linguistics*, London: Routledge, pp. 42-56.

The British Council (2013) ‘The English effect’, [Online]. Available at: https://www.britishcouncil.org/sites/default/files/english-effect-report-v2.pdf. (Accessed: 22 May 2018).

Upton, T.A. and Cohen, M.A. (2009) 'An approach to corpus-based discourse analysis: The move analysis as example', *Discourse Studies*, 11(5), pp. 585-605.

van Bonn, S. and Swales, J.M. (2007) ‘English and French journal abstracts in the language sciences: Three exploratory studies’, *Journal of English for Academic Purposes*, 6(2), pp. 93-108.

van Dijk, T.A. (2009) *Society and discourse: How social contexts influence text and talk*. Cambridge: Cambridge University Press.

Wei, Y. and Lei, L. (2011). Lexical bundles in the academic writing of advanced Chinese EFL learners. *RELC journal*, 42(2), pp. 155-166.

Wellein, V. (2008) *Cross-cultural differences: American, German, Japanese and Chinese negotiation styles*. diplom. de.

Wray, A. (2000) ‘Formulaic sequences in second language teaching: Principle and practice’, *Applied Linguistics*, 21(4), pp. 463-489.

Wray, A. (2002) *Formulaic language and the lexicon*. Cambridge: Cambridge University Press.

Appendixes

A. Master of Arts Program in Teaching English as a Foreign Language

Academic Program
Master of Arts Program in Teaching English as a Foreign Language
Degree Offered
Master of Arts (Teaching English as a Foreign Language)
M.A. (Teaching English as a Foreign Language)
Objectives
1. To develop the quality and standards of English teachers at all levels.
2. To offer a curriculum of effective English teaching theory and practice for those who are interested in this career.
3. To help solve the problem of a shortage of qualified English teachers at primary, secondary and university levels.
Admission
1. Applicants must hold a Bachelor's degree or equivalent in any field of study from a local or foreign institution recognized by the Thammasat University Council.
2. Applicants may be graduates of the Graduate Diploma Program in Teaching English as a foreign Language from the language Institute, Thammasat University.
3. Applicants must meet the requirements in "Number 8: Admission" of Thammasat University's Graduate School under the 1998 Master of Arts Programs Section.
4. Applicants also must successfully pass an English language entrance examination and an interview in English.
Academic System
1. The M.A. Program in Teaching English as a Foreign Language is a program where English is the medium of instruction.
2. The academic year is divided into two 16-week semesters (first and second semesters) and one 8-week summer session. Courses are offered on weekends: Saturdays and Sundays (9:00 – 4:00 p.m.).
3. There are two options available:
Plan A Plan A consists of course work and a thesis.
Plan B Plan B consists of course work, a comprehensive examination, and independent study instead of a thesis.
3.1 Plan A: Thesis
Students can register for a thesis after they have completed at
least 30 credits with a minimum cumulative GPA of 3.00.
*To graduate from the program, students must have their theses published in an academic journal.
3.2 Plan B: Comprehensive Examination
Students who are qualified to take a comprehensive examination must have completed all credits required for graduation, including 3 credits of independent study, and attained a GPA of at least 3.00. The comprehensive examination will be offered in the second semester of the year in which a student is expected to graduate.
4. Students must pass a foreign language test with a "P" grade as required by Thammasat University's Regulations for Graduate Students (1996) before they are allowed to defend their thesis or take the comprehensive examination.

5. The process of preparing in carrying out the thesis for Plan A students must follow the Regulations of Thammasat University concerning Master's Thesis Writing.
6. Students from other faculties may take courses in this program with permission of the Language Institute Graduate Committee.

Duration of Study

The maximum study period must not exceed 14 semesters.

Registration

Registration is under the requirements in "Number 10: Registration" of Thammasat University's Graduate School under the 1998 Master of Arts Programs Section.

Grading System and Graduation

Grading System

1. Academic performance is evaluated as follows:

Grade	A	A-	B+	B	B-	C+	C	D	F
Grade point	4.00	3.67	3.33	3.00	2.67	2.33	2.00	1.00	0.00

2. Only credit courses with grades of S (Satisfactory) or at least C will be counted for graduation. However, the GPA and cumulative GPA will be computed from all courses taken including those for which the student received a D or F.
3. For Plan A, the thesis is given a grade of S (Satisfactory) of U (Unsatisfactory). The thesis with S is equivalent to passing. and can be counted for graduation.
4. For Plan B, the comprehensive examination is given a grade of P (Pass) of N (Not pass). The comprehensive examination with P is equivalent to passing, and can be counted for graduation.
5. The Foreign Language Test is graded P (Pass) or N (Not Pass).
6. Other requirements must comply with Thammasat University's Regulation "Number 11: Grading system' of the 1998 Graduate Studies.

Graduation

Students who graduate from this program must have:
1. earned 42 credits in either Plan A or B
2. a cumulative GPA of not less than 3.00
3. received P in the foreign Language Test
4. received S for the thesis in Plan A or received P for Independent Study in Plan B
5. Other requirements must comply with Thammasat University's Regulations for Graduate Students.

B. Master of Arts Program in English for Careers

Language Institute
Thammasat University Revised 2008

Academic Program

Master of Arts Program in English for Careers

Degree Offered

Master of Arts (English for Careers)

M.A. (English for Careers)

Objectives

1. To provide opportunities for graduates in various fields to study and practice English for specific purposes using all four skills and enable them to use English in their work effectively and efficiently.
2. To enable graduates in various fields to practice English in order to enhance their analytical and research skills in the English language, facilitating their experience and professionalism in their respective fields.
3. To produce graduates at an advanced level of English for Careers to meet the increasing needs of the job market for fully-qualified professionals.

Admission

1. Applicants must hold a Bachelor's degree or equivalent in any field of study from local or foreign institutions which are recognized by the Thammasat University Council.
2. Applicants may be graduates of the Graduate Diploma Program in English for Careers from the Language Institute, Thammasat University.
3. Applicants must meet the requirements in "Number 8: Admission" of Thammasat University's Regulations on Graduate Studies under the 1988 and 2002 Master of Arts Programs Section.
4. Applicants also must successfully pass an English language entrance examination and an interview in English.

Academic System

1. The M.A. Program in English for Careers is a program where English is the medium of instruction.
2. The academic year is divided into two 16-week semesters (first and second semesters) and one 8-week summer session. Courses are offered on weekends: Saturdays and Sundays (9.00 a.m. - 4.00 p.m.).
3. There are two options available:

Plan A Plan A consists of course work and a thesis.

Plan B Plan B consists of course work, a comprehensive examination, and independent study instead of a thesis.

3.1 **Plan A: Thesis**

Students can register for a thesis after they have completed at least 21 credits with a minimum cumulative GPA of 3.00.

3.2 **Plan B: Comprehensive Examination**

Students who are qualified to take a comprehensive examination must complete all credits required for graduation, including 6 credits for independent study, and earned a GPA of at least 3.00.

4. Students must pass a foreign language test (English) with a "P" (Pass) grade as required by Thammasat University's Regulations on Graduate Studies (1998 and 2002) before they are allowed to defend their thesis or take the comprehensive examination.

5. The process of preparing the thesis for "Plan A" students must follow the Regulations of Thammasat University concerning Master's Thesis Writing.
6. Students from other faculties may take courses in this program with permission of the Language Institute Graduate Committee.

Duration of Study

The maximum study time must not exceed 5 years.

Registration

Registration is under the requirements in "Number 10: Registration" of Thammasat University's Regulations on Graduate Studies under the 1988 and 2002 Master of Arts Programs Section.

Grading System and Graduation

Grading System

1. Academic performance is evaluated as follows:

Grade	A	A-	B+	B	B-	C+	C	D	F
Grade point	4.00	3.67	3.33	3.00	2.67	2.33	2.00	1.00	0.00

2. Only credit courses with grades of S (Satisfactory) or at least C will be counted for graduation. However, the GPA and cumulative GPA must be computed for a D or F in any particular course.
3. For Plan A, the thesis is given a grade of S (Satisfactory) or U (Unsatisfactory). The thesis with S is equivalent to passing and can be counted for graduation.
4. For Plan B, the comprehensive examination is given a grade of P (Pass) or N (Not pass). The comprehensive examination with P is equivalent to passing, and can be counted for graduation.
5. The Foreign Language Test (English) is graded P (Pass) or N (Not pass).
6. Other requirements must comply with Thammasat University's Regulation "Number 13: Grading System" of the 1988 and 2002 Graduate Studies.

Graduation

Students who graduate from this program must:
1. earn 42 credits in either Plan A or B
2. have a cumulative GPA of not less than 3.00
3. receive P in the Foreign Language Test (English): TU-GET 550, TOEFL IBT 80, or IELTS 6.5
4. receive S for the thesis in Plan A or received P for Independent Study in Plan B
5. have their thesis, or part of it, published in an academic journal or presented at a conference
6. meet other requirements set in Thammasat University's Regulations on Graduate Studies
7. make sure all their debts or obligations are paid to the University

C. Lists of Research Title in Novice Abstract Corpus (NAC)

No.	English for Careers Program (CR)/ Research Title	Code
1	Barriers to effective communication of Thai cabin crew working in international airlines where English is not the primary language	N_CR_2010_1
2	Learning problems of Assumption Commercial College students in classes of native English-speaking language teachers	N_CR_2010_2
3	Request strategies used by American and Thai office workers in Bangkok: A cross-cultural pragmatics study	N_CR_2010_3
4	The effect of code-mixing (Thai English in print advertisements on product knowledge and understanding, attitudes and purchase intention: Consumers' educational background as a moderator	N_CR_2010_4
5	English-Thai mixing on international women magazine's covers and how this trend has developed and changed in the past six years: A case study of Thai edition of Marie Claire Magazine	N_CR_2010_5
6	Satisfaction of teenage audience in Bangkok with English in program advertising on True Visions	N_CR_2010_6
7	Students' opinions on the impact of using Facebook for English reading and writing skill improvement	N_CR_2010_7
8	A research study of needs and wants of English skills used by local employees working in an international organization	N_CR_2010_8
9	A study of collocational errors in Thai EFL learner's writing	N_CR_2010_9
10	Attitude towards English learning of Matthayom Suksa 2 students at Muangpranburi School	N_CR_2010_10
11	Avoidance of phrasal verbs by M.6 students at Thiamudomsuksapattanakarn Nonthaburi School	N_CR_2010_11
12	A study of Thai EFL learners' metacognitive awareness of reading strategies	N_CR_2010_12
13	Listening comprehension abilities and strategies of Mini-English program (MEP) and Non-Mini English Program (Non-MEP) students: A study of Suankularb Wittayalai Nonthaburi School	N_CR_2010_13
14	A survey study of vocabulary learning strategies used by M.3 students at Satit Suansunandha School in academic year 2010	N_CR_2010_14
15	A study of the english listening comprehension problems for efl learners at the faculty of ICT, Silpakorn University	N_CR_2010_15
16	An error analysis of the use of past simple and past continuous tenses by first year students at Silpakorn University	N_CR_2010_16
17	How leadership affects expatriates' executive secretaries' oral communication skills	N_CR_2010_17
18	Factors affecting motivation in studying English of the Thai employees at Siam DENSO Manufacturing Co., Ltd.	N_CR_2010_18
19	A study on using songs in teaching English through listening and singing	N_CR_2010_19
20	A study of academic factors affecting the English reading comprehension of students of master of arts in English for careers, Batch 13 of Thammasat University	N_CR_2011_20
21	A survey of oral presentation techniques by high-proficiency graduate students	N_CR_2011_21
22	Students' anxiety in learning English at Wall Street Institute Pinklao	N_CR_2011_22
23	A survey study on motivation in attending English classes among employees of an engineering services company	N_CR_2011_23
24	Problems in reading scientific texts of grade seven students in Patumwan Demonstration School English Program	N_CR_2011_24
25	Genre analysis of scientific research proposal: Analysis of moves in Malaria research proposal	N_CR_2011_25
26	Corpus linguistics and systematic functional grammar of Harry Potter and the Deadly Hallows	N_CR_2011_26

No.	English for Careers Program (CR)/ Research Title	Code
27	Grammatical collocation problems among Thai students in the medical science field: A study on the use of prepositions	N_CR_2011_27
28	Factor affecting M.2 students' performance in learning English at Traimudomsuksapattanakarn Ratchada School	N_CR_2011_28
29	The study of intercultural barriers between Thai flight attendants and foreign passengers: Comparing regional routes and intercontinental routes	N_CR_2011_29
30	The relationship between attitudes and English reading proficiency of undergraduate students at Dhurakij Pundit University	N_CR_2011_30
31	Thai staff attitudes towards speaking English at the British Embassy Bangkok	N_CR_2011_31
32	A study of communication apprehension. Self-esteem, and willingness to communicate among MEC13 students at the Language Institute, Thammasat University	N_CR_2011_32
33	A study of public speaking anxiety (PSA) among the graduate students at the Language Institute, Thammasat University	N_CR_2011_33
34	Success factors and barriers to cross-cultural communication of the Thai and Indian employees working in information technology firms	N_CR_2011_34
35	English listening comprehension problems of international conferences attendees: a study of military officers at the directorate of joint communications Royal Thai Armed Forces Headquarters	N_CR_2011_35
36	A study of Thai EFL learners' production of polysyllabic English word stress	N_CR_2011_36
37	The classification of grammatical errors of written business messages among M.A. students at Language Institute, Thammasat University	N_CR_2011_37
38	Rhetorical move analysis: A comparative study of U.S. and Thai CEO's letter to the shareholders during the 2008	N_CR_2011_38
39	A survey study of attitudes of employees towards English courses at Maxim Integrated Products Thailand (MIPT)	N_CR_2012_39
40	Attitudes of Thai undergraduate students toward non-native English teachers from the ASEAN Community	N_CR_2012_40
41	A survey of first year ICT students' needs in English usage and wants regarding an English online learning program	N_CR_2012_41
42	The motivation for reading in a foreign language and reading achievement of MEC students at Thammasat University	N_CR_2012_42
43	A corpus-based study of sematic prosody of the words cause and commit	N_CR_2012_43
44	An error analysis on English infinitive usage of Thai graduate students	N_CR_2012_44
45	Narrative plot structure analysis in 'the Curious Enlightenment of Professor Caritat'	N_CR_2012_45
46	A corpus-based analysis of English synonyms: Remember, recall, recollect	N_CR_2012_46
47	Pattern of obligation and necessity in speeches on climate change to the United Nations framework convention	N_CR_2012_47
48	Attitudes towards English loanwords in Japanese and their effects on spoken English among Japanese workers in Thailand	N_CR_2012_48
49	Attitude of undergraduate students towards English computer loanwords and Thai terminology	N_CR_2012_49
50	Grammatical errors and the role of schema in written business communication	N_CR_2012_50
51	A corpus-based study on the synonymous English verbs of "give", "provide" and "offer"	N_CR_2012_51
52	Job satisfaction of Thai EFL teachers: A study of Benchama Maharat School, Ubonratchathani	N_CR_2012_52
53	Reflections on using simulations in an EFL classroom: a case study of students in a postgraduate program in Thammasat University	N_CR_2012_53
54	Student's attitudes towards teacher's error correction in written work	N_CR_2012_54
55	Investigating the comprehension of chatting abbreviations: a study of line messenger users	N_CR_2012_55

No.	English for Careers Program (CR)/ Research Title	Code
56	A survey study on the reasons of Thai high school student for studying at English tutorial schools	N_CR_2012_56
57	Temporal English preposition errors: a case study of Thai undergraduate students	N_CR_2012_57
58	The use of vocabulary learning strategies by postgraduate students at Thammasat University in the academic year 2014	N_CR_2013_58
59	The attitudes of officers at Haco Group (1991) Co., Ltd. towards English oral communication in their workplace	N_CR_2013_59
60	Students' satisfaction toward the developing English skill program of Singburi Vocational College	N_CR_2013_60
61	Students' satisfaction with the services from administrative officers in the Language Institute, Thammasat University	N_CR_2013_61
62	A survey of problems and needs in English communication of clinical research associates	N_CR_2013_62
63	Speaking anxiety of Thai salespeople in the jewelry business in Bangkok when communicating with foreign customers	N_CR_2013_63
64	The use of business English e-mails of hotel employees	N_CR_2013_64
65	Thai EFL learners' understanding of past simple tense and present perfect tense	N_CR_2013_65
66	Writing English business e-mails by engineers in ETA (Thailand) Limited	N_CR_2013_66
67	Attitude towards learning English online	N_CR_2013_67
68	Problem in oral communication skill of Thai journalists when interviewing foreigners	N_CR_2013_68
69	A survey study of motivation and preparation of Debsirin High School students in Matthayomsuksa 6 to pass English examinations	N_CR_2013_69
70	Petrochemical employees' motivation for learning English	N_CR_2013_70
71	Communication strategies used by Mattayom 3 student at Nawamintrachinuthi Satriwittaya 2 School	N_CR_2013_71
72	A survey of perception and understanding of mosquito repellent product labeling	N_CR_2013_72
73	A corpus-based analysis of speech act of thanking in the Corpus of Contemporary American English (COCA)	N_CR_2013_73
74	A study of speaking strategies of SIIT employees when communicating with foreign students at SIIT, Thammasat University	N_CR_2013_74
75	English reading level of grade 12 students (Arts-English section), Rajavinit Bangkhen School and Wat Sangwet School	N_CR_2013_75
No.	**Teaching English as a Foreign Language (TE)/ Research Title**	**Code**
1	An evaluation of the textbook "Language leader: Intermediate" for first-year students at Thai-Nichi Institute of Technology	N_TE_2010_76
2	A Survey of The Application of Sociocultural Concepts to EFL Thai Classrooms at The Faculty of Archaeology, Silpakorn University	N_TE_2010_77
3	A survey of the needs of English proficiency of five-star hotels' staff and representatives in Bangkok and the aspect of needs referring to hotel English and general English	N_TE_2010_78
4	Needs analysis: English training course design for Thai sales staff in English bookstore in Bangkok	N_TE_2010_79
5	A survey study of the first year TEP and TEPE students' opinions towards English language needs and problems in English usage	N_TE_2010_80
6	A survey study of motivational factors affecting English extensive reading of high school students at Benjamarachanusorn School	N_TE_2010_81
7	A survey study of English reading strategies used by able students and less able students majoring in general business management at Silpakorn University	N_TE_2010_82
8	Vocabulary learning strategies of Matthayom 6 student at Assumption Convent School	N_TE_2010_83

No.	Teaching English as a Foreign Language (TE)/ Research Title	Code
9	A survey study of M 4 students' attitudes towards using wordless picture books to improve their writing skill	N_TE_2010_84
10	A survey study of reading strategies utilized by able English users and less able English users of M.4 Triamudom Suksa students	N_TE-2010_85
11	A correlational study of linguistic competence vs performance of the English plural morpheme by Matthayom students	N_TE_2010_86
12	Effects of speed reading on Thai primary school EFL students' reading rate and comprehension	N_TE_2010_87
13	The production of /r/ and /l/ of M.6 students in the English program at Suankularbwittayalai Rangsit School	N_TE_2010_88
14	The use of authentic materials to enhance vocabulary for listening comprehension of Matthayomsuksa 6 students in the English and Japanese program	N_TE_2010_89
15	Students' attitudes towards benefits of assigned external reading: A case study of first-year undergraduate students in the major of business management and English, at Silapakorn University, Petchaburi branch	N_TE_2010_90
16	A survey study of export staff's needs and problems in English communication at Thanulux Public Co., Ltd.	N_TE_2010_91
17	A survey study of English reading comprehension problems in academic texts as perceived by Matthayom 5 students in the English program at Thai Christian School	N_TE_2010_92
18	A comparison of English reading comprehension ability of students from Prakdeang Pittayakom School	N_TE_2010_93
19	Vocabulary learning strategies used by the first year Thai undergraduate students of the Bachelor of Economics international program at Thammasat university	N_TE_2010_94
20	A study on the use of English reading strategies by Mattayom 4 students at princess Chulabhorn's College Chonburi (regional science school)	N_TE_2010_95
21	A survey study of students' attitudes towards the English textbook, World Club 1, students' book at Watdusitaram Secondary School, Bangkok	N_TE_2010_96
22	A survey study of the opinions of the lower secondary students toward using the Student English Access (SEAR) at Roi-Et Wittiyalai School	N_TE_2010_97
23	News writing problem in a news writing course	N_TE_2011_98
24	An error analysis of English writing of eleventh grade students at Thanyaburi School	N_TE_2011_99
25	The effect of social network language on learners' writing style: The study of fourth year university students in Thailand	N_TE_2011_100
26	An error analysis of voyage reports written by pilots of Thai Airways International Public Company Limited	N_TE_2011_101
27	An analysis of problems and needs in English oral presentation skills for occupational purposes of Thai engineers in Denso Thailand companies	N_TE_2011_102
28	Needs analysis for English for logistics course: A survey study of undergraduates at Silpakorn University	N_TE_2011_103
29	Needs analysis for English course for trainer in fundamental skill training course at the leading automotive company	N_TE_2011_104
30	Need analysis for in-bound tour guide in Bangkok	N_TE_2011_105
31	Vocabulary learning strategies in comprehending the company's written documents utilized by Thai finance staff at Unilever Thai Trading	N_TE_2011_106
32	Needs analysis of English language skills for Royal Thai Air Force personnel	N_TE_2011_107
33	The effects of a use of collocation teaching on speaking ability of grade 10 students at Ban Mii Wittayakom School	N_TE_2011_108
34	Refusal strategies of Thai hotel staff	N_TE_2011_109
35	L1 interference on communicative English in Japanese company: A focus on speaking skill	N_TE_2012_110
36	Measuring Thai EFL learners' awareness of reading strategies	N_TE_2012_111
37	English listening-speaking Exemption Test: Test reliability and validity	N_TE_2012_112

No.	Teaching English as a Foreign Language (TE)/ Research Title	Code
38	A study on the use of English relative clause in English reading class	N_TE_2012_113
39	A survey study of the motivational factors affecting English as a second language learning of the students at Sirindhorn International Institute of Technology (SIIT), Thammasat university	N_TE_2012_114
40	Thai learners' acquisition of English passive voice	N_TE_2012_115
41	The study of 10th grade students' achievement and retention of English vocabulary using KSVP memorization model at Phol School in Phol District of Khon-Kaen	N_TE_2012_116
42	A survey study: What EFL teacher want to include in vocabulary game application to help students learning vocabulary	N_TE_2012_117
43	Factors affecting Thai EFL students' English oral communication skills in 4th years student in computer-science at King Mongkut's University of Technology Thonburi (KMUTT)	N_TE_2012_118
44	Online Learning Experiences and Attitudes of Autonomous Thai EFL Learners Using Facebook-based Community in 2012 (Non-English major Thai undergraduate students at Thammasat University)	N_TE_2012_119
45	Students' perceptions of learning English with native English-speaking teachers	N_TE_2012_120
46	Attitudes and perceptions of EFL learners in Thailand for the use of problem-based learning instruction to improve English language skills: A Survey Research	N_TE_2012_121
47	Exploring characteristics for effective EFL teachers from the perceptions of high school students at Mahidol Wittayanusorn School.	N_TE_2012_122
48	Task-based language teaching for intermediate level students	N_TE_2012_123
49	Parents' Teaching English Through "Whole Language Approach": A Case Study of Thai Elementary Homeschooled Children	N_TE_2012_124
50	Cognitive reading strategies in EFL adult learners	N_TE_2012_125
51	A survey study of needs for improving English language proficiency of the employees at Booking.com (Thailand) Company Limited.	N_TE_2012_126
52	Effects of scaffolding on reading comprehension	N_TE_2012_127
53	The survey study of reading comprehension problems of Thai learners in English academic Text	N_TE_2012_128
54	EFL teachers' use of motivational strategies in secondary schools in Thailand	N_TE_2013_129
55	A study of the use of participial phrases - present participles and past participles - as post-nominal modifiers among high proficiency learners at Mathayom 6 level in Bangpakokwitthayakom School	N_TE_2013_130
56	A study of the use of cohesive devices of Thai students in a bilingual school	N_TE_2013_131
57	The survey study of motivation in L2 extensive reading of undergraduate students in the major of English and International Engineering Management at Thammasat University	N_TE_2013_132
58	Needs analysis in English use of taxi drivers	N_TE_2013_133
59	Need analysis of English for Thai teacher assistant in international school	N_TE_2013_134
60	A survey of Thai EFL undergraduate first year students' preferences about teacher's feedback on writing	N_TE_2013_135
61	An investigation into Thai high school learners in Muang District, Khon Kaen perceptions towards learner autonomy in language learning	N_TE_2013_136
62	Needs analysis for English skills of Thai Teachers who teach Thai as a foreign language (ESP)	N_TE_2013_137
63	Hotel sales staff perceptions toward the use of English skills	N_TE_2013_138
64	The Benefit of cognitive development in vocabulary of Thai-English bilinguals in elementary school	N_TE_2013_139
65	Aspects of unaccusative Verbs and Thai EFL learners' overpassivization	N_TE_2013_140
66	Validity and reliability: The English written test of content subjects for regular program.	N_TE_2013_141

No.	Teaching English as a Foreign Language (TE)/ Research Title	Code
67	Teachers' experience in English language teaching and learning involving young THAI EFL learners with Attention Deficit Hyperactivity Disorder	N_TE_2013_142
68	The perception of Thai EFL teachers on using translanguaging in classroom context	N_TE_2014_143
69	Discourse analysis of graded readers: A case of Jane Austen's Pride and Prejudice	N_TE_2013_144
70	The effectiveness of authentic materials to enhance reading comprehension for high school students	N_TE_2013_145
71	English conversation challenges: A Case study of postal school students at Postal School, Thailand Post Headquarter	N_TE_2013_146
72	Needs analysis of using English of Muay Thai trainers in Thailand	N_TE_2013_147
73	Learning style preferences of young Thai EFL learners	N_TE_2013_148
74	The effect of theme based instruction on intensive English program students at primary level	N_TE_2013_149
75	The attitudes of the parents of international school students towards the use of code switching in classroom	N_TE_2013_150

D. Lists of Research Title in Professional Abstract Corpus (PAC)

No.	English Language Teaching (EL)/ Research Title	Code
1	Dealing with learner reticence in the speaking class	P_EL_2010_001
2	Why are students quiet? Looking at the Chinese context and beyond	P_EL_2010_002
3	Help seeking in language learning	P_EL_2010_003
4	Student views on learning grammar with web- and book- based materials	P_EL_2010_004
5	Balancing the dual functions of portfolio assessment	P_EL_2010_005
6	Views on creativity from an Indonesian perspective	P_EL_2010_006
7	Formulation as evidence of understanding in teacher–student talk	P_EL_2010_007
8	Using the linguistic landscape as a pedagogical resource	P_EL_2010_008
9	Pre-task syntactic priming and focused task design	P_EL_2010_009
10	Attribution and learning English as a foreign language	P_EL_2010_010
11	Bumping into someone: Japanese students' perceptions and observations	P_EL_2010_011
12	Critical Friends Group for EFL teacher professional development	P_EL_2010_012
13	Direct teaching of vocabulary after reading: is it worth the effort?	P_EL_2010_013
14	An exploration of on-task language policy and student satisfaction	P_EL_2010_014
15	'Would you perhaps consider …': Hedged comments in ESL writing	P_EL_2010_015
16	A focus on purpose: Using a genre approach in an EFL writing class	P_EL_2010_016
17	Pragmatics, awareness raising, and the Cooperative Principle	P_EL_2010_017
18	Reducing student reticence through teacher interaction strategy	P_EL_2010_018
19	Augmenting learner autonomy through blogging	P_EL_2010_019
20	The Listening Log: Motivating autonomous learning	P_EL_2010_020
21	The effect of pair work on a word-building task	P_EL_2010_021
22	Teaching English as a friendly language: Lessons from the SCOTS corpus	P_EL_2010_022
23	The role of EFL teachers' emotional intelligence in their success	P_EL_2010_023
24	A mindset for EFL: Learners' beliefs about the role of natural talent	P_EL_2010_024
25	Students transcribing tasks: Noticing fluency, accuracy, and complexity	P_EL_2010_025
26	Feedback revolution: What gets in the way?	P_EL_2010_026
27	Explicit grammar rules and L2 acquisition	P_EL_2011_027
28	If only it were true: The problem with the four conditionals	P_EL_2011_028
29	Using the L1 'errors' of native speakers in the EFL classroom	P_EL_2011_029
30	Closing the loop between theory and praxis: New models in EFL teaching	P_EL_2011_030
31	Pausing patterns: Differences between L2 learners and native speakers	P_EL_2011_031
32	Politics without pedagogy: Questioning linguistic exclusion	P_EL_2011_032
33	Learner self-assessment: An investigation into teachers' beliefs	P_EL_2011_033
34	Quality self-reflection through reflection training	P_EL_2011_034
35	Product–process distinctions in ELT curriculum theory and practice	P_EL_2011_035
36	Introducing diversity of English into ELT: Student teachers' responses	P_EL_2011_036
37	Using translation exercises in the communicative EFL writing classroom	P_EL_2011_037
38	A PBLT approach to teaching ESL speaking, writing, and thinking skills	P_EL_2011_038
39	A case study of students' perceptions of peer assessment in Hong Kong	P_EL_2011_039
40	Turn allocation patterns and learning opportunities	P_EL_2011_040
41	Reading discussion groups for teachers: Connecting theory to practice	P_EL_2011_041
42	Classroom silence: Voices from Japanese EFL learners	P_EL_2011_042
43	L1 to teach L2: Complexities and contradictions	P_EL_2011_043
44	Student perceptions of student interaction in a British EFL setting	P_EL_2011_044
45	Interjections in the EFL classroom: Teaching sounds and sequences	P_EL_2011_045
46	From classroom to workplace: Tracking socio-pragmatic development	P_EL_2011_046
47	The first term at university: Implications for EAP	P_EL_2011_047
48	The potential of dual-language cross-cultural peer review	P_EL_2011_048
49	Metacognitive instruction for helping less-skilled listeners	P_EL_2011_049
50	Teachers researching their own practice	P_EL_2011_050
51	Talk or chat? Chatroom and spoken interaction in a language classroom	P_EL_2011_051
52	Weblogs: a tool for EFL interaction, expression, and self-evaluation	P_EL_2011_052
53	Children's perceptions of learning with an interactive whiteboard	P_EL_2011_053
54	The representation of foreign cultures in English textbooks	P_EL_2011_054

No.	English Language Teaching (EL)/ Research Title	Code
55	What learners get for free: learning of routine formulae in ESL and EFL environments	P_EL_2011_055
56	'Small Talk': Developing fluency, accuracy, and complexity in speaking	P_EL_2012_056
57	Developing multiliteracies in ELT through telecollaboration	P_EL_2012_057
58	Learner negotiation of L2 form in transcription exercises	P_EL_2012_058
59	Reflections on a transnational peer review of teaching	P_EL_2012_059
60	Improving teacher talk through a task-based approach	P_EL_2012_060
61	L1–L2 sentence translation in classroom grammar tests	P_EL_2012_061
62	Speechant: A vowel notation system to teach English pronunciation	P_EL_2012_062
63	Implementing the European Language Portfolio in a Turkish context	P_EL_2012_063
64	Promoting reflective practice in continuing education in France	P_EL_2012_064
65	A genre-based approach to teaching EFL summary writing	P_EL_2012_065
66	The value of targeted comic book readers	P_EL_2012_066
67	The dark side of motivation: teachers' perspectives on 'unmotivation'	P_EL_2012_067
68	Literature Circles in ELT	P_EL_2012_068
69	Primary school English teachers' research engagement	P_EL_2012_069
70	Teachers' beliefs and practices regarding learner autonomy	P_EL_2012_070
71	Learner code-switching versus English only	P_EL_2012_071
72	ESL readers' comprehension performance: the Chinese secondary context	P_EL_2012_072
73	Using task repetition to direct learner attention and focus on form	P_EL_2012_073
74	Is linguistic ability variation in paired oral language testing problematic?	P_EL_2012_074
75	Shared delivery: Integrating ELT in Australian vocational education	P_EL_2012_075
76	The mediating role of textbooks in high-stakes assessment reform	P_EL_2012_076
77	Trainer talk: Levels of intervention	P_EL_2012_077
78	Transcription and the IELTS speaking test: Facilitating development	P_EL_2013_078
79	Time is of the essence: Factors encouraging out-of-class study time	P_EL_2013_079
80	'Storyline': A task-based approach for the young learner classroom	P_EL_2013_080
81	Posters, self-directed learning, and L2 vocabulary acquisition	P_EL_2013_081
82	Questions arising from the assessment of EFL narrative writing	P_EL_2013_082
83	'A basic starter pack': The TESOL Certificate as a course in survival	P_EL_2013_083
84	Reflective feedback sessions using video recordings	P_EL_2013_084
85	Memorization versus semantic mapping in L2 vocabulary acquisition	P_EL_2013_085
86	Interest and interest-enhancing strategies of adolescent EFL learners	P_EL_2013_086
87	Students' and teachers' ideals of effective Business English teaching	P_EL_2013_087
88	Designing tasks for the Business English classroom	P_EL_2013_088
89	The six-category intervention analysis: A classroom observation reference	P_EL_2013_089
90	Listening for needles in haystacks: How lecturers introduce key terms	P_EL_2013_090
91	Scaffolding language learning in an academic ESL classroom	P_EL_2013_091
92	Effective and practical critical thinking-enhanced EFL instruction	P_EL_2013_092
93	'Hi doctor, give me handouts': Low-proficiency learners and requests	P_EL_2013_093
94	Online collaborative learning on an ESL teacher education programme	P_EL_2013_094
95	Vocabulary explanation in English-major university classrooms in China	P_EL_2013_095
96	The relationship between assessment types and text revision	P_EL_2013_096
97	It's never too late: An overview of e-learning	P_EL_2013_097
No.	**English for Specific Purposes (ES)/ Research Title**	**Code**
1	Speaking rate and information content in English lingua franca oral presentations	P_ES_2010_098
2	Two ESP projects under the test of time: The case of Brazil and Tunisia	P_ES_2010_099
3	A corpus-based evaluation of metaphors in a business English textbook	P_ES_2010_100
4	Ethics and integrity in proofreading: Findings from an interview-based study	P_ES_2010_101
5	The linguistic and the contextual in applied genre analysis: The case of the company audit report	P_ES_2010_102
6	Directives in office hour consultations: A corpus-informed investigation of learner and expert usage	P_ES_2010_103
7	Devising and implementing a business proposal module: Constraints and compromises	P_ES_2010_104

No.	English for Specific Purposes (ES)/ Research Title	Code
8	Collaborative writing: Bridging the gap between the textbook and the workplace	P_ES_2010_105
9	Transfer climate and EAP education: Students' perceptions of challenges to learning transfer	P_ES_2010_106
10	Business as usual: The use of English in the professional world in Hong Kong	P_ES_2010_107
11	English communicative events and skills needed at the workplace: Feedback from the industry	P_ES_2010_108
12	Conflict or cooperation: The use of backchannelling in ELF negotiations	P_ES_2010_109
13	The effects of the use of English in Polish product advertisements: Implications for English for business purposes	P_ES_2010_110
14	Language needs analysis from a perspective of international professional mobility: The case of French mountain guides	P_ES_2010_111
15	English in the context of European integration: A corpus-driven analysis of lexical bundles in English EU documents	P_ES_2010_112
16	Meeting the challenges of European student mobility: Preparing Italian Erasmus students for business lectures in English	P_ES_2010_113
17	Academic research networks: Accessing resources for English-medium publishing	P_ES_2010_114
18	A contrastive study of the rhetorical organisation of English and Spanish PhD thesis introductions	P_ES_2010_115
19	"You don't say what you know, only what you can": The perceptions and practices of senior Spanish academics regarding research dissemination in English	P_ES_2011_116
20	Perception vs. evidence: An analysis of this and these in academic prose	P_ES_2011_117
21	Wikis and academic writing: Changing the writer–reader relationship	P_ES_2011_118
22	A function-first approach to identifying formulaic language in academic writing	P_ES_2011_119
23	Lexicogrammar in the international construction industry: A corpus-based case study of Japanese–Hong-Kongese on-site interactions in English	P_ES_2011_120
24	How a corpus-based study of the factors which influence collocation can help in the teaching of business English	P_ES_2011_121
25	Changes to English as an Additional Language writers' research articles: From spoken to written register	P_ES_2011_122
26	Comparative and contrastive observations on scientific titles written in English and Spanish	P_ES_2011_123
27	Discursive hierarchical patterning in Economics cases	P_ES_2011_124
28	The Discussion section as argument: The language used to prove knowledge claims	P_ES_2011_125
29	Asserting or deflecting expertise? Exploring the rhetorical practices of master's theses in the philosophy of education	P_ES_2011_126
30	Exploring perception and use of everyday language and medical terminology among international medical graduates in a medical ESP course in Australia	P_ES_2011_127
31	Meeting the challenges of English-medium higher education: The first-year experience in Hong Kong	P_ES_2011_128
32	Lecturing undergraduate science in Danish and in English: A comparison of speaking rate and rhetorical style	P_ES_2011_129
33	A corpus-based approach to online materials development for writing research articles	P_ES_2011_130
34	The vocabulary thresholds of business textbooks and business research articles for EFL learners	P_ES_2011_131
35	A pragmatic approach to the macro-structure and metadiscoursal features of research article introductions in the field of Agricultural Sciences	P_ES_2011_132
36	The influence of learner strategies on oral presentations: A comparison between group and individual performance	P_ES_2011_133
37	The presentation of self in scholarly life: Identity and marginalization in academic homepages	P_ES_2011_134
38	Legal Problem Question Answer Genre across jurisdictions and cultures	P_ES_2011_135

No.	English for Specific Purposes (ES)/ Research Title	Code
39	English for airport ground staff	P_ES_2011_136
40	Developing an English for specific purpose curriculum for Asian call centres: How theory can inform practice	P_ES_2012_137
41	The most frequently-used multi-word constructions in academic written English: A multi-corpus study	P_ES_2012_138
42	Translation competence and translation performance: Lexical, syntactic and textual patterns in student translations of a specialized EU genre	P_ES_2012_139
43	Academic literacy socialization of first year doctoral students in US: A micro-ethnographic perspective	P_ES_2012_140
44	Educating Chinese scientists to write for international journals: Addressing the divide between science and technology education and English language teaching	P_ES_2012_141
45	Recurrent word combinations in academic writing by native and non-native speakers of English: A lexical bundles approach	P_ES_2012_142
46	Proper vocabulary and juicy collocations': EAP students evaluate do-it-yourself corpus-building	P_ES_2012_143
47	The use of questions by professors in lectures given in English: Influences of disciplinary cultures	P_ES_2012_144
48	Modifiers in BASE and MICASE: A matter of academic cultures or lecturing styles?	P_ES_2012_145
49	Business and legal case genre networks: Two case studies	P_ES_2012_146
50	A reception study of the articles published in "English for Specific Purposes" from 1990-1999	P_ES_2012_147
51	Structural patterns in empirical research articles: A cross-disciplinary study	P_ES_2012_148
52	Stance and engagement in pure mathematics research articles: Linking discourse features to disciplinary practices	P_ES_2012_149
53	Modeling the relationships between test-taking strategies and test performance on a graph-writing task: Implications for EAP	P_ES_2012_150
54	Evaluating prior scholarship in literature reviews of research articles: A comparative study of practices in two research paradigms	P_ES_2012_151
55	Designing email tasks for the Business English classroom: Implications from a study of Hong Kong's key industries	P_ES_2012_152
56	Just remember this: Lexicogrammatical relevance markers in lectures	P_ES_2012_153
57	"That's it for today": Academic lecture closings and the impact of class size	P_ES_2012_154
58	"Convenience Editing" in action: Comparing English teachers' and medical professionals' revisions of a medical abstract	P_ES_2012_155
59	Epistemological and interpersonal stance in a data description task: Findings from a discipline-specific learner corpus	P_ES_2012_156
60	Multimodal evaluation in academic discussion sessions: How do presenters act and react?	P_ES_2012_157
61	Business students' beliefs about language learning in a university context	P_ES_2012_158
62	"Just spoke to …": The types and directionality of intertextuality in professional discourse	P_ES_2013_159
63	Evaluation of oral performance in outsourced call centres: An exploratory case study	P_ES_2013_160
64	Multimedia as a means to enhance teaching technical vocabulary to physics undergraduates in Rwanda	P_ES_2013_161
65	Chemistry journal articles: An interdisciplinary approach to move analysis with pedagogical aims	P_ES_2013_162
66	(Critical) Language awareness in business communication	P_ES_2013_163
67	Self-mention and the projection of multiple identity roles in TESOL graduate student presentations: The influence of the written academic genres	P_ES_2013_164

No.	English for Specific Purposes (ES)/ Research Title	Code
68	Graduate learners' approaches to genre-analysis tasks: Variations across and within four disciplines	P_ES_2013_165
69	Engineering English and the high-tech industry: A case study of an English needs analysis of process integration engineers at a semiconductor manufacturing company in Taiwan	P_ES_2013_166
70	Student and teacher reflections on indirectness as a pragmatic feature of pharmacist–patient simulations	P_ES_2013_167
71	The discursive construction of professional expertise: Appeals to authority in barrister's opinions	P_ES_2013_168
72	Business English students learning to write for international business: What do international business practitioners have to say about their texts?	P_ES_2013_169
73	Non-canonical grammar in Best Paper award winners in engineering	P_ES_2013_170
74	A corpus-assisted comparative genre analysis of corporate earnings calls between Korean and native-English speakers	P_ES_2013_171
75	"Just wanna give you guys a bit of an update": Insider perspectives on business presentations in Hong Kong	P_ES_2013_172
76	Developing research report writing skills using corpora	P_ES_2013_173
77	The Article of the future: Strategies for genre stability and change	P_ES_2013_174
78	The lexical needs of ESP students in a professional field	P_ES_2013_175
79	An exploration of the impact of students' prior genre knowledge on their constructions of 'audience' in a Marketing course at postgraduate level	P_ES_2013_176
No.	**Language Learning (LL)/ Research Title**	**Code**
1	Nonnative processing of verbal morphology: in search of regularity	P_LL_2010_177
2	Verbal inflectional morphology in L1 and L2 Spanish: A frequency effects study examining storage versus composition	P_LL_2010_178
3	Cognitive predictors of generalization of Russian grammatical gender categories	P_LL_2010_179
4	Second language acquisition of gender agreement in explicit and implicit training conditions: An event-related potential study	P_LL_2010_180
5	Processing English compounds in the first and second language: The influence of the middle morpheme	P_LL_2010_181
6	Sociolinguistic variation in the speech of learners of Chinese as a second language	P_LL_2010_182
7	A corpus-based analysis of the discourse functions of ser/estar + adjective in three levels of Spanish as FL learners	P_LL_2010_183
8	Cross-Language priming of word meaning during second language sentence comprehension	P_LL_2010_184
9	Teaching L2 learners how to listen does make a difference: An empirical study	P_LL_2010_185
10	Exploring the relationship between modified output and working memory capacity	P_LL_2010_186
11	Data-Driven Learning: Taking the computer out of the equation	P_LL_2010_187
12	The development of polysemy and frequency use in English second language speakers	P_LL_2010_188
13	Attribution and self-efficacy and their interrelationship in the Korean EFL context	P_LL_2010_189
14	Environmental influence on language acquisition: Comparing second and foreign language acquisition of Swedish	P_LL_2010_190
15	Bilingual children's acquisition of English verb morphology: Effects of language exposure, structure complexity, and task type	P_LL_2010_191
16	Working memory and reactivity	P_LL_2010_192
17	Disfluency markers in L1 attrition	P_LL_2010_193
18	Perceptual learning style matching and L2 vocabulary acquisition	P_LL_2010_194
19	Willingness to Communicate in English: A model in the Chinese EFL classroom context	P_LL_2010_195

No.	Language Learning (LL)/ Research Title	Code
20	The Role of L1 transfer in the interpretation of articles with definite plurals in L2 English	P_LL_2010_196
21	Modeling the development of L1 and EFL writing proficiency of secondary school students	P_LL_2011_197
22	Proficiency and animacy effects on L2 gender agreement processes during comprehension	P_LL_2011_198
23	The time factor in EFL classroom practice	P_LL_2011_199
24	Learner-generated noticing of written L2 input: What do learners notice and why?	P_LL_2011_200
25	Effects of stress and working memory capacity on foreign language readers' inferential processing during comprehension	P_LL_2011_201
26	A comparison of the effects of reading and listening on incidental vocabulary acquisition	P_LL_2011_202
27	Negative transfer from Spanish and English to Portuguese pronunciation: The roles of inhibition and working memory	P_LL_2011_203
28	Does foreign language writing benefit from increased lexical fluency? Evidence from a classroom experiment	P_LL_2011_204
29	Deliberate learning and vocabulary acquisition in a second language	P_LL_2011_205
30	Learning to express motion events in an L2: the case of Chinese directional complements	P_LL_2011_206
31	Processability theory and German case acquisition	P_LL_2011_207
32	Effects of negative and positive evidence on adult word learning	P_LL_2011_208
33	Fluency training in the esl classroom: An experimental study of fluency development and proceduralization	P_LL_2011_209
34	Processing advantages of lexical bundles: Evidence from self-paced reading and sentence recall tasks	P_LL_2011_210
35	Semantic categories and context in L2 vocabulary learning	P_LL_2011_211
36	Verb-noun collocations in second language writing: a corpus analysis of learners' English	P_LL_2011_212
37	Testing second language oral proficiency in direct and semidirect settings: a social-cognitive neuroscience perspective	P_LL_2011_213
38	Influence of second language Cherokee immersion on children's development of past tense in their first language, English	P_LL_2011_214
39	Le carte blanc or la carte blanche? Bilingual children's acquisition of French adjective agreement	P_LL_2011_215
40	An exploration of L2 teachers' use of pedagogical interventions devised to draw L2 learners' attention to form	P_LL_2011_216
41	"I prefer not text": Developing Japanese learners' communicative competence with authentic materials	P_LL_2011_217
42	Social-cultural-historical contradictions in an L2 listening lesson: a joint activity system analysis	P_LL_2011_218
43	Effect of explicit and implicit instruction on free written response task performance	P_LL_2011_219
44	The effect of L2 proficiency and study-abroad experience on pragmatic comprehension	P_LL_2011_220
45	Morphological congruency and the acquisition of L2 morphemes	P_LL_2011_221
46	Self-consciousness and assertiveness as explanatory variables of L2 oral ability: A latent variable approach	P_LL_2011_222
47	Legitimate peripheral participation and language learning: two Quichua learners in a transnational community	P_LL_2011_223
48	The roles of structured input activities in processing instruction and the kinds of knowledge they promote	P_LL_2011_224
49	Effectiveness of different pinyin presentation formats in learning Chinese characters: A cognitive load perspective	P_LL_2011_225
50	Learning a tonal language by attending to the tone: An in vivo experiment	P_LL_2011_226

No.	Language Learning (LL)/ Research Title	Code
51	Second language learning benefits from similarity in word endings: Evidence from Russian	P_LL_2011_227
52	Characteristics of academic language register occurring in caretaker-child interaction: Development and validation of a coding scheme	P_LL_2011_228
53	Using raters from India to score a large-scale speaking test	P_LL_2011_229
54	Deletion of /t, d/ and the acquisition of linguistic variation by second language learners of English	P_LL_2011_230
55	When is a difference really different? Learners' discrimination of linguistic contrasts in American Sign Language	P_LL_2011_231
56	Contextual dynamics in foreign language learning motivation	P_LL_2011_232
57	Evidence on the effectiveness of comprehensive error correction in second language writing	P_LL_2012_233
58	Relearning in the elderly: Age-related effects on the size of savings	P_LL_2012_234
59	Morphological family size effects in young first and second language learners: Evidence of cross-language semantic activation in visual word recognition	P_LL_2012_235
60	Working memory and the observed effectiveness of recasts on different L2 outcome measures	P_LL_2012_236
61	From dichotomy to divergence: number/gender marking on Hebrew nouns and adjectives across school ages	P_LL_2012_237
62	Acquiring regular and irregular past tense morphemes in English and French: Evidence from bilingual children	P_LL_2012_238
63	A comparison of rereading benefits in first and second language reading	P_LL_2012_239
64	Comparing playful and nonplayful incidental attention to form	P_LL_2012_240
65	Statistical learning and language: An individual differences study	P_LL_2012_241
66	L2 negation constructions at work	P_LL_2012_242
67	Acquiring a variable structure: an interlanguage analysis of second language mood use in Spanish	P_LL_2012_243
68	The effects of input-based practice on pragmatic development of requests in L2 Chinese	P_LL_2012_244
69	The role of task complexity, modality, and aptitude in narrative task performance	P_LL_2012_245
70	Do L1 reading achievement and L1 print exposure contribute to the prediction of L2 proficiency?	P_LL_2012_246
71	Reinvestigating the noticing function of output	P_LL_2012_247
72	Socio-interactive practices and personality in adult learners of English with little formal education	P_LL_2012_248
73	Teacher motivational practice, student motivation, and possible L2 selves: An examination in the Iranian EFL context	P_LL_2012_249
74	Effects of form-focused instruction and corrective feedback on L2 pronunciation development of /ɹ/ by Japanese learners of English	P_LL_2012_250
75	Constraints on implicit learning of grammatical form-meaning connections	P_LL_2012_251
76	The effect of pleasure reading on Japanese university EFL learners' reading rates	P_LL_2012_252
77	Older learners in SLA research: a first look at working memory, feedback, and L2 development	P_LL_2012_253
78	Optimizing language instruction: Matters of explicitness, practice, and cue learning	P_LL_2012_254
79	English language learners and academic achievement: revisiting the threshold hypothesis	P_LL_2012_255
80	Investigating Chinese university students' knowledge of and attitudes toward plagiarism from an integrated perspective	P_LL_2012_256
81	Tasks, teacher feedback, and learner modified output in naturally occurring classroom interaction	P_LL_2012_257
82	The etiology of individual differences in second language acquisition in Australian school students: A behavior-genetic study	P_LL_2012_258
83	Syntax and discourse in near-native French: Clefts and focus	P_LL_2012_259
84	Structural priming and second language learning	P_LL_2012_260

No.	Language Learning (LL)/ Research Title	Code
85	Effects of input properties, vocabulary size, and l1 on the development of third person singular –s in child L2 English	P_LL_2012_261
86	A self-system perspective on young adolescents' motivation to learn English in urban and rural settings	P_LL_2012_262
87	Investigating deaf children's vocabulary knowledge in British Sign Language	P_LL_2012_263
88	Learning vowel categories from maternal speech in Gurindji Kriol	P_LL_2012_264
89	Acquiring noun plurals in Palestinian Arabic: Morphology, familiarity, and pattern frequency	P_LL_2012_265
90	Written languaging, direct correction, and second language writing revision	P_LL_2012_266
91	The relative effects of explicit correction and recasts on two target structures via two communication modes	P_LL_2012_267
92	Direct and indirect roles of morphological awareness in the English reading comprehension of native English, Spanish, Filipino, and Vietnamese speakers	P_LL_2012_268
93	Intercultural conflict and mediation: An intergroup perspective	P_LL_2012_269
94	Improving l2 listeners' perception of English vowels: A computer-mediated approach	P_LL_2012_270
95	Expanding the role of connectionism in SLA theory	P_LL_2013_271
96	The effects of teachers' motivational strategies on learners' motivation: A controlled investigation of second language acquisition	P_LL_2013_272
97	Age effects in a study abroad context: Children and adults studying abroad and at home	P_LL_2013_273
98	Explicit and implicit lexical knowledge: Acquisition of collocations under different input conditions	P_LL_2013_274
99	The development of L2 oral language skills in two L1 groups: A 7-year study	P_LL_2013_275
100	Effects of referent token variability on L2 vocabulary learning	P_LL_2013_276
101	Investigating the role of verbal working memory in young children's sentence comprehension	P_LL_2013_277
102	Dynamic adaptation in child–adult language interaction	P_LL_2013_278
103	Motivation, gender, and possible selves	P_LL_2013_279
104	What a Bayesian analysis can do for SLA: new tools for the sociolinguistic study of subject expression in l2 Spanish	P_LL_2013_280
105	Elementary school EP_LLS' reading skill profiles using cognitive diagnosis modeling: roles of length of residence and home language environment	P_LL_2013_281
106	Motivation and vision: an analysis of future L2 self images, sensory styles, and imagery capacity across two target languages	P_LL_2013_282
107	Stability and change in one adult's second language English negation	P_LL_2013_283
108	The acquisitional value of recasts in instructed second language speech learning: teaching the perception and production of English /ɹ/ to adult Japanese learners	P_LL_2013_284
109	Hi-lab: a new measure of aptitude for high-level language proficiency	P_LL_2013_285
110	Doing planning and task performance in second language acquisition: an ethnomethodological respecification	P_LL_2013_286
111	Individual differences in sequence learning ability and second language acquisition in early childhood and adulthood	P_LL_2013_287
112	Language production and reception: a processability theory study	P_LL_2013_288
113	Processing of regular and irregular past-tense verb forms in first and second language reading acquisition	P_LL_2013_289
114	An event-related potential (ERP) investigation of filler-gap processing in native and second language speakers	P_LL_2013_290
115	Transition probabilities and different levels of prominence in segmentation	P_LL_2013_291
116	Participation structure and incidental focus on form in adult ESL classrooms	P_LL_2013_292
117	Relationship of L1 skills and L2 aptitude to L2 anxiety on the foreign language classroom anxiety scale	P_LL_2013_293

No.	TESOL Quarterly (TQ) / Research Title	Code
1	Going beyond patterns: Involving cognitive analysis in the learning of collocations	P_TQ_2010_294
2	Do ESL essay raters' evaluation criteria change with experience? A mixed-methods, cross-sectional study	P_TQ_2010_295
3	Duty and service: Life and career of a Tamil teacher of English in Sri Lanka	P_TQ_2010_296
4	Effects of dynamic corrective feedback on ESL writing accuracy	P_TQ_2010_297
5	Using the dash method to measure reading comprehension	P_TQ_2010_298
6	Group processes and EFL learners' motivation: a study of group dynamics in EFL classrooms	P_TQ_2010_299
7	Should vocabulary instruction be integrated or isolated?	P_TQ_2010_300
8	Learning phrasal verbs through conceptual metaphors: A case of Japanese EFL learners	P_TQ_2010_301
9	To be or not to be “part of them”: Micropolitical challenges in mainland Chinese students' learning of English in a multilingual university	P_TQ_2010_302
10	Toward a taxonomy of written errors: Investigation into the written errors of Hong Kong Cantonese ESL learners	P_TQ_2010_303
11	A dynamic metacognitive systems account of Chinese university students' knowledge about EFL reading	P_TQ_2010_304
12	Onwards and upwards: Space, placement, and liminality in adult ESOL classes	P_TQ_2010_305
13	A gay immigrant student's perspective: Unspeakable acts in the language class	P_TQ_2010_306
14	Agency in the making: Adult immigrants' accounts of language learning and work	P_TQ_2010_307
15	Narrating America: socializing adult esl learners into idealized views of the united states during citizenship preparation classes	P_TQ_2010_308
16	From refugee to transformer: a Bourdieusian take on a Hmong learner's trajectory	P_TQ_2010_309
17	“I learn more at school”: a critical perspective on workplace-related second language learning in and out of school	P_TQ_2010_310
18	L1 influence on the acquisition of l2 collocations: Japanese esl users and EFL learners acquiring English collocations	P_TQ_2010_311
19	Word meanings matter: cultivating English vocabulary knowledge in fifth-grade Spanish-speaking language minority learners	P_TQ_2010_312
20	Causal relationships between communication confidence, beliefs about group work, and willingness to communicate in foreign language group work	P_TQ_2010_313
21	Immigrant success stories in esl textbooks	P_TQ_2010_314
22	Influence of teacher-contact time and other variables on esl students' attitudes towards native- and nonnative-English-speaking teachers	P_TQ_2010_315
23	Native speaker norms and china English: From the perspective of learners and teachers in China	P_TQ_2010_316
24	Should we use characteristics of conversation to measure grammatical complexity in L2 writing development?	P_TQ_2011_317
25	A corpus-based evaluation of syntactic complexity measures as indices of college-level esl writers' language development	P_TQ_2011_318
26	Rater sensitivity to qualities of lexis in writing	P_TQ_2011_319
27	Effects of varying lengths of study-abroad experiences on Japanese EFL students' L2 writing ability and motivation: A longitudinal study	P_TQ_2011_320
28	An intensive look at intensity and language learning	P_TQ_2011_321
29	Responding to L2 students in college writing classes: Teacher perspectives	P_TQ_2011_322
30	Effects of working memory capacity and content familiarity on literal and inferential comprehension in L2 reading	P_TQ_2011_323
31	Effect of frequency and idiomaticity on second language reading comprehension	P_TQ_2011_324

No.	TESOL Quarterly (TQ) / Research Title	Code
32	A corpus-based study of the linguistic features and processes which influence the way collocations are formed: Some implications for the learning of collocations	P_TQ_2011_325
33	English textbooks in parallel-language tertiary education	P_TQ_2011_326
34	Researcher identity, narrative inquiry, and language teaching research	P_TQ_2011_327
35	TESOL and TESD in remote aboriginal Australia: The “true” story?	P_TQ_2011_328
36	Harnessing storytelling as a Sociopragmatic skill: Applying narrative research to workplace English courses	P_TQ_2011_329
37	Course-taking patterns of Latino ESL students: Mobility and mainstreaming in urban community colleges in the United States	P_TQ_2011_330
38	Evaluating the validity of a high-stakes ESL Test: Why teachers' perceptions matter	P_TQ_2011_331
39	The Most Frequently Used English Phrasal Verbs in American and British English: A Multicorpus examination	P_TQ_2011_332
40	Narrow viewing: the vocabulary in related television programs	P_TQ_2011_333
41	The perception of English speech sounds by Cantonese esl learners in Hong Kong	P_TQ_2011_334
42	Effects of instruction on adolescent beginners’ acquisition of request modification	P_TQ_2012_335
43	Glossing and second language vocabulary learning	P_TQ_2012_336
44	Second language writers’ strategy use and performance on an integrated reading-listening-writing task	P_TQ_2012_337
45	The discursive positioning of teachers: native-speaking English teachers and educational discourse in Hong Kong	P_TQ_2012_338
46	“I am never afraid of being recognized as an NNES”: one teacher's journey in claiming and embracing her nonnative-speaker identity	P_TQ_2012_339
47	English as a transcultural language in Swedish policy and practice	P_TQ_2012_340
48	Teacher development in a global profession: an autoethnography	P_TQ_2012_341
49	Advantages and disadvantages of native- and nonnative-English-speaking teachers: student perceptions in Hong Kong	P_TQ_2012_342
50	“The school of life”: differences in U.S. and CANADIAN settlement policies and their effect on individual Haitian immigrants’ language learning	P_TQ_2012_343
51	Developing academic language in English language learners through sheltered instruction	P_TQ_2012_344
52	Participation in peer response as activity: An examination of peer response stances from an activity theory perspective	P_TQ_2012_345
53	TESOL teacher education: novice teachers' perceptions of their preparedness and efficacy in the classroom	P_TQ_2012_346
54	The role of mentoring in supporting novice English language teachers in Hong Kong	P_TQ_2012_347
55	Furnished imagination: the impact of preservice teacher training on early career work in TESOL	P_TQ_2012_348
56	Novice ESOL teachers' perceptions of social support networks	P_TQ_2012_349
57	“It cannot be done alone”: the socialization of novice English teachers in South Korea	P_TQ_2012_350
58	Oral fluency, accuracy, and complexity in formal instruction and study abroad learning contexts	P_TQ_2012_351
59	Diary of a dabbler: ecological influences on an EFL teacher's efforts to study Japanese informally	P_TQ_2012_352
60	Authenticity and TV shows: A multidimensional analysis perspective	P_TQ_2012_353
61	Comparing multidimensional and continuum models of vocabulary acquisition: An empirical examination of the vocabulary knowledge scale	P_TQ_2012_354
62	“Any questions?”: Investigating the nature of understanding-checks in the language classroom	P_TQ_2012_355
63	Challenges in critical language teaching	P_TQ_2012_356
64	Reconsidering compulsory English in developing countries in Asia: English in a community of northeast Thailand	P_TQ_2012_357
65	Effects of pretask modeling on attention to form and question development	P_TQ_2013_358

No.	TESOL Quarterly (TQ) / Research Title	Code
66	The effect of focus on form and focus on forms instruction on the acquisition of productive knowledge of l2 vocabulary by young beginning-level learners	P_TQ_2013_359
67	Translanguaging, TEXMEX, and bilingual pedagogy: Emergent bilinguals learning through the vernacular	P_TQ_2013_360
68	English language learners' access to and attainment in postsecondary education	P_TQ_2013_361
69	Desire in translation: white masculinity and TESOL	P_TQ_2013_362
70	Foreign language anxiety: understanding its status and insiders' awareness and attitudes	P_TQ_2013_363
71	Comparing the effects of test anxiety on independent and integrated speaking test performance	P_TQ_2013_364
72	Chinese college English teachers' research engagement	P_TQ_2013_365
73	Is text written for children useful for L2 extensive reading?	P_TQ_2013_366
74	Contributions of individual differences and contextual variables to reading achievement of English language learners: An empirical investigation using hierarchical linear modeling	P_TQ_2013_367
75	An investigation of working memory effects on oral grammatical accuracy and fluency in producing questions in English	P_TQ_2013_368
76	Effects of the manipulation of cognitive processes on EFL writers' text quality	P_TQ_2013_369
77	The ESL Teacher as plurilingual: An Australian perspective	P_TQ_2013_370
78	2B or not 2B plurilingual? Navigating languages literacies, and plurilingual competence in postsecondary education in Canada	P_TQ_2013_371
79	Plurilingual Pedagogical practices in a policy-constrained context: A northern Ugandan case study	P_TQ_2013_372
80	Culture from the bottom up	P_TQ_2013_373
81	Culture in English as a Foreign Language (EFL) textbooks: A semiotic approach	P_TQ_2013_374
82	Teacher language background, codeswitching, and English-only instruction: Does age make a difference to learners' attitudes?	P_TQ_2013_375
83	The effects of explicit instruction on the reading performance of adolescent English language learners with intellectual disabilities	P_TQ_2013_376
84	The influence of second language experience and accent familiarity on oral proficiency rating: A qualitative investigation	P_TQ_2013_377
85	The I don't know option in the Vocabulary Size Test	P_TQ_2013_378

E. Pilot Study Sample and Answer Sheet

Pilot 1: CR_52_10

(s1) The purpose of this research was to analyze the effect of code-mixing (Thai-English) in print advertisements on product knowledge and understanding, attitudes and purchase intention.

(s2) In addition, another objective was to explore the relationship between code mixing and the consumers' educational background which was as a moderator.

(s3) The subjects of this study were 120 Thai females living in Bangkok and neighboring provinces.

(s4) The subjects consisted of respondents holding qualifications lower than a bachelor's degree and people who obtained a bachelor's and postgraduate degree.

(s5) They were selected by using stratified random sampling to complete the self-administered questionnaire with closed-ended questions for demographic information.

(s6) In addition, they answered closed-ended questions with semantic differential scales for information about the effect of code-mixing (Thai-English) in print advertisements on product knowledge and understanding, attitude, as well as purchase intention.

(s7) Moreover, two facial skincare whitening products were selected as samples for this research study.

(s8) Both of them were adjusted into two types of language used in the advertisement: Thai monolingual and code mixed (Thai-English) language attached with the questionnaire.

(s9) The findings were analyzed using the SPSS to find descriptive statistics such as frequency, percentage and mean.

(s10) The results obtained showed the majority of respondents believed that advertisements containing code-mixing could convey messages effectively, provide better knowledge and understanding than Thai monolingual advertisements, as well as create a positive attitude toward language in advertising and product image.

(s11) However, code-mixing in advertisements had no effect on attitudes toward advertisements as a whole and could not convince them to buy the products or had no effect on their purchase intention.

(s12) There is a correlation between the use of code-mixing and consumers' educational background; code-mixing (Thai-English) in advertisements created better knowledge, understanding, attitudes towards language and product image than not using mixed English terms for high educated consumers.

(s13) In contrast, the use of code-mixing had a significant effect only on product image for low educated consumers.

(s14) However, code-mixing (Thai-English) did not cause better attitudes toward advertisements as a whole and had no influence on the purchase intentions for both groups of respondents who were at different educational levels.

Pilot 2: CR_52_33

(s1) This study had two main purposes: (1) to investigate attitudes of students towards English learning at school; and (2) to study relationship between independent variables (gender and grade in previous English course) and attitudes of students towards English learning at school.

(s2) The subjects participating in the study were Mtthayom Suksa 2 at Muangpranburi School.

(s3) There were 183 students without being randomly selected.

(s4) The research instrument employed in the study was a questionnaire modified from previous researchers conducting studies in the same area.

(s5) The data was collected by using questionnaires which consist of two parts: the general information of the respondents and the students' opinions towards English learning.

(s6) The obtained data was analyzed by using the Statistical Package for Social Sciences (SPSS) version 12.0 to calculate the frequency, percentage and mean.

(s7) The research questions were tested by using t-test and F-test.

Pilot 3: CR_53_66

(s1) At present, reading plays an important role in all EFL/ESL students' academic lives.

(s2) Most students, sometimes, still cannot completely understand reading materials and texts because there are some significant factors influencing their reading comprehension.

(s3) Teachers and researchers have found that most students fail to read English effectively (Aebersold & Field, 1997).

(s4) Previous studies have suggested that most of the second language readers have difficulties in reading English texts because of the factors of readers' background knowledge, the reading materials, the teaching techniques, and the teacher's personality (Jamornmarn & Ruangtakul, 1995; Lenz, 2005)

(s5) Therefore, this study aims to investigate whether academic factors including teachers, reading materials, and students have an effect on the English reading comprehension of second language readers.

(s6) On theoretical grounds, we predicted that the factor of students' ability affects their English reading comprehension more than the factors of teachers and reading materials.

(s7) Data was collected from 62 students of the Master of Arts in English for Careers program, batch 13 of Thammasat University.

(s8) The questionnaire was divided into 4 parts: the background information of the respondents, the teacher, the reading materials, and the students' ability.

(s9) All data was fed into the computer and the result was computed by using the SPSS program (Statistical Package for the Social Sciences) in order to obtain descriptive statistics; frequency, percentage, and means.

(s10) The Likert's scale was used to rank the degree of the effect of each factor.

(s11) The findings showed that the academic factors affect the second language reading comprehension differently.

(s12) Moreover, the factor of students' ability mostly affects reading achievement and comprehension. (s13) However, this study may be useful for both students and teachers because students can make use of the finding to encourage themselves to improve their skills.

(s14) In addition, teachers can choose the appropriate reading materials which match the level of students' abilities and adapt their teaching tactics to help students develop their reading ability.

Pilot11: EL_10_7

(s1) With an increasing emphasis on creativity in education and language teaching, it is important for teachers and students to examine their own views on and perceptions of creativity.

(s2) What is regarded as creativity may vary from one context to another.

(s3) This paper examines the perceptions of creativity reflected in the evaluation of creative poems by students and teachers in an Indonesian context.

(s4) The participants evaluated short poems written by a group of Indonesian university students.

(s5) This resulted in several characteristics of creativity being highlighted: honesty, reality, truthfulness, and personal value.

(s6) The poems were voted on and the ones which received the greatest number of votes were also analysed.

(s7) The more creative poems demonstrated higher degrees of language play and creative language use. (s8) This paper suggests several activities which could be conducted by teachers and students in other contexts to help them examine their own explicit and implicit views on creativity.

Pilot 12: EL_12_66

(S1) This study utilizes a pre-test/post-test assessment to investigate the instructional efficacy of a genre-based approach to teaching summary writing.

(s2) Forty-one EFL university students in Taiwan were asked before and after the instruction to summarize a simplified version of The Adventures of Tom Sawyer in a maximum of 500 words.

(s3) All the students' summaries on the pre- and post-tests were evaluated against content, organization, vocabulary, and language use.

(c4.1) The statistical results showed that

(c4.2) such an approach was effective in improving students' overall summarization performance of a narrative source text and that the students benefited to a greater extent in content and organization than in vocabulary and language use.

(c5.1) The results were also supported

(p5.2) by the students' interview comments.

Pilot 13: LL_11_27

(s1) We examined negative transfer from English and Spanish to Portuguese pronunciation.

(s2) Participants were native English speakers, some of whom spoke Spanish.

(s3) Participants completed a computer-based Portuguese pronunciation tutorial and then pronounced trained letter-to-sound correspondences in unfamiliar Portuguese words; some shared orthographic form with their translation in Spanish or Spanish and English.

(s4) Spanish-speaking participants were more accurate and made more Spanish-like than English-like errors.

(s5) Contrary to predictions, non-Spanish speakers made more Spanish-like than English-like errors on cognates.

(s6) Participants with higher working memory were more accurate and made more Spanish-sounding errors on cognates.

(s7) The results suggest that the first language is inhibited during second-language production and that higher working memory is associated with an improved ability to inhibit the first language.

Answer Sheet

Directions: Please read the sample abstracts and decide which move each unit belongs to according to Hyland's (2000) analytical framework below. Please note that prior to the coding, the abstracts were initially divided into units, i.e. sentences, clauses, and phrases. If embedded moves are found in any sentence, the sentence will be sectioned into clauses or phrases as seen in the sample below. However, in case that you did not agree with any preliminary division, this could be resolved through discussion.

Sample Abstract

(s1) Motivation has long been identified as one of the main factors affecting English language learning (Gardner, 1985).

(c2.1) This study examined the type and level of English language learning motivation (instrumental or integrative)

(p2.2) of 30 first-year undergraduate students at an international institute of engineering and technology in Thailand.

(s3) A modified motivational survey of 20 items adapted from Gardner's Attitude/ Motivation Test Battery (AMTB) was conducted.

(s4) The data was analyzed by means of frequency, percentage, arithmetic mean and standard deviation.

Note: *'s' stands for sentence, 'c' stands for clause, and 'p' stands for phrase. In the sample above, the second sentence was divided into a clause and a phrase: c2.1 and p2.2. According to Hyland's Framework, c2.1 was categorized as move 2 (Purpose), while p2.2 was coded as move 3 (Methods).*

Hyland's (2000, p. 67) five-move model of abstract analysis

Introduction	Establishes context of the paper and motivates the research
Purpose	Indicates purpose, thesis or hypothesis, outlines the intention behind the paper.
Methods	Provides information on design, procedures, assumptions, approach, data, etc.
Product	States main findings or results, the argument, or what was accomplished.
Conclusion	Interprets or extends results beyond the scope of the paper, draws inferences, points to applications, or wider applications.

Interrater	Intro	Purpose	Method	Product	Conclusion	Others (pls. indicate)
Pilot 1						
Pilot 2						
Pilot 3						
Pilot 4						
Pilot 5						
Pilot 6						
Pilot 7						
Pilot 8						
Pilot 9						
Pilot 10						
Pilot 11						
Pilot 12						
Pilot 13						
Pilot 14						
Pilot 15						
Pilot 16						
Pilot 17						
Pilot 18						
Pilot 19						
Pilot 20						

F. Complete List of Rhetorical Move Patterns of the NAC Subcorpora Organized by Move Compositions

No.	Move Patterns (CR)	Freq.	%	Move Patterns (TE)	Freq.	%
1	**2, 3, 4, 5**	**29**	**38.67**	**2, 3, 4, 5**	**28**	**37.33**
1.1	2(3)-3-4-5	25	33.33	2(3)-3-4-5	19	25.33
1.2	2(3)-4-5	1	1.33	2(3)-4-5	2	2.67
1.3	2(3)-2-3-4-5	1	1.33	2(3)-3-3(4)-4-5	2	2.67
1.4	2(3)-3-4-4(5)-5	1	1.33	2-3-4-5	1	1.33
1.5	2(3)-3-5-4	1	1.33	2(3)-3-4-(4)5	1	1.33
1.6				2(3)-3-4-(3-4)5	1	1.33
1.7				2(3)-3-2-3-4-5	1	1.33
1.8				2(3)-3-(3)4-4-5	1	1.33
2	**2, 3, 4**	**23**	**30.67**	**2, 3, 4**	**16**	**21.33**
2.1	2(3)-3-4	18	24.00	2(3)-3-4	13	17.33
2.2	2-3-4	2	2.67	2(3)-2-3-4	1	1.33
2.3	2(3)-2-3-4	1	1.33	2(3)-3-3(4)-4	1	1.33
2.4	2(3)-3-4-3(2)-4	1	1.33	2-3-4	1	1.33
2.5	2(3)-4	1	1.33			
3	**1, 2, 3, 4, 5**	**16**	**21.3**	**1, 2, 3, 4, 5**	**19**	**25.33**
3.1	1-2(3)-3-4-5	11	14.67	1-2(3)-3-4-5	13	17.33
3.2	1-1(2-3)-3-4-5	1	1.33	1-2-3-4(5)	1	1.33
3.3	1-2(3)-2-3-4-5	1	1.33	1-2(3)-4-5	1	1.33
3.4	1-2(3)-3-4-5-4	1	1.33	1-2(3)-3-4(5)	1	1.33
3.5	1-2(3)-5-3-4	1	1.33	1-2(3)-3-3(4)-4-5	1	1.33
3.6	1-2-3-4-5	1	1.3	1-2(3)-3(4)-4-5	1	1.33
3.7				1-2(3)-2-3-4-5	1	1.33
4	**1, 2, 3, 4**	**6**	**8.00**	**1, 2, 3, 4**	**10**	**13.33**
4.1	1-2(3)-3-4	5	6.67	1-2(3)-3-4	8	10.67
4.2	2(3)-1-3-4	1	1.33	2(3)-1-3-4	1	1.33
4.3				1-3-3(4)-2(3)	1	1.33
5	**1, 2, 3, 5**	**0**	**0.00**	**1, 2, 3, 5**	**1**	**1.33**
5.1	1-2-3-5			1-2-3-5	1	1.33
6	**2, 3**	1	**1.33**	**2, 3**	**0**	**0.0**
6.1	2(3)-3	1	1.33			
7	**2, 3, 5**	**0**	**0.0**	**2, 3, 5**	**1**	**1.3**
7.1				2(3)-3-5	1	1.3

G. Complete List of Rhetorical Move Patterns in the PAC Subcorpora Organized by Move Compositions

No.	Move Patterns (EL)	Freq.	%	Move Patterns (ES)	Freq.	%	Move Patterns (LL)	Freq.	%	Move Patterns (TQ)	Freq.	%
1	2, 3, 4, 5	23	23.71	2, 3, 4, 5	26	32.91	2, 3, 4, 5	56	47.86	2, 3, 4, 5	26	30.59
1.1	2(3)-3-4-5	12	12.37	2(3)-3-4-5	6	7.59	2(3)-3-4-5	24	20.51	2(3)-3-4-5	13	15.29
1.2	2(3)-3-3(4)-5	1	1.03	2(3)-3-3(4)-5	1	1.27	2(3)-3-3(4)-5	4	3.42	2(3)-3-3(4)-5	2	2.35
1.3	2(3)-3(4)-5	2	2.06	2(3)-4-5	4	5.06	2(3)-4-5	3	2.56	2(3)-4-5	1	1.18
1.5	2(3)-2-4-5	1	1.03	2(3)-2-4-5	1	1.27	2(3)-3-4-4(5)	2	1.71	2(3)-3-4-4(5)	2	2.35
1.6	2(3)-3(4)-4-5	1	1.03	2(3)-3-3(4)-4-5	1	1.27	2(3)-3-3(4)-4-5	2	1.71	2(3)-2-3-4-5	2	2.35
1.7	2-3-4-5	1	1.03	2(3)-3-3(4)-3-5	1	1.27	2(3)-3-4-3(4)-5	2	1.71	2(3)-3-3(4)-4-5	2	2.35
1.8	2(3)-3(4)-4-4(3)-5	1	1.03	2(3)-3(4)-4-5	1	1.27	2-2(3)-4-5	2	1.71	2-2(3)-4-5	1	1.18
1.9	2(3)-3-2(3)-4-5	1	1.03	2(3)-3-4-3-4-5	1	1.27	2(3)-2-3(4)-4(3-5)-5	1	0.85	2(3)-3(4)-4-5	1	1.18
1.10	2(3)-3-4(3)-4-5	1	1.03	2(3)-5(3)-3-4-5	1	1.27	2(3)-2-3(4)-5	1	0.85	2(3)-3-3(4)-4-4(3)-5	1	1.18
1.11	3(2)-3(4)-5	1	1.03	2-3-3(4)-5	1	1.27	2(3)-2-3-4-4(5)	1	0.85	2(3)-3-3(4)-5(4)	1	1.18
1.12	3-2(3)-3-4-5	1	1.03	2(3)-3-4-4(5)	1	1.27	2(3)-3(4)-4-5	1	0.85			
1.13				2(3)-3-4-4(5)-3	1	1.27	2(3)-3(4)-5	1	0.85			
1.14				2(3)-3-4-4(5)-5	1	1.27	2(3)-3-4(3)-5	1	0.85			
1.15				2(3)-3-5-4	1	1.27	2(3)-3-4(5)	1	0.85			
1.16				2(3)-4(3)-4-5	1	1.27	2(3)-3-4-3-4-5	1	0.85			
1.17				2(3)-4-3(4)-4-3(4)-5	1	1.27	2(3)-4(3)-4-5	1	0.85			
1.18				2(3)-2-3(4)-4-5	1	1.27	2(3)-4-4(5)	1	0.85			
1.19				2(3)-3-4(5)	1	1.27	2-3(2)-3-4-5	1	0.85			
1.20							2-3-4-3(4)-5	1	0.85			

No.	Move Patterns (EL)	Freq.	%	Move Patterns (ES)	Freq.	%	Move Patterns (LL)	Freq.	%	Move Patterns (TQ)	Freq.	%
1.21							2-3-4-5	1	0.85			
1.22							2-4(3)-5	1	0.85			
1.23							3(2)-3-3(4)-4-5	1	0.85			
1.24							3(2)-3-4(5)-5	1	0.85			
							3(2)-3-4-4(5)-3-4-5	1	0.85			
2	1, 2, 3, 4, 5	21	21.65	1, 2, 3, 4, 5	28	35.44	1, 2, 3, 4, 5	19	16.2	1, 2, 3, 4, 5	29	34.12
2.1	1-2(3)-3-4-5	5	5.15	1-2(3)-3-4-5	9	11.39	1-2(3)-3-4-5	5	4.27	1-2(3)-3-4-5	7	8.24
2.2	1-2(3)-4-5	6	6.19	1-2(3)-4-5	2	2.53	1-2(3)-4-5	2	1.71	1-2(3)-4-5	4	4.71
2.3	1-2(3)-3(4)-5	1	1.03	1-2(3)-3-3(4)-5	3	3.80	1(2-3)-3-4-5	1	0.85	1-2(3)-3-3(4)-5	2	2.35
2.4	1-2(3)-3(4)-4-5	1	1.03	1-2(1-3)-3-4-5	2	2.53	1-1(2-3)-3-4-4(3)-4-5	1	0.85	2(3)-1-3-4-5	2	
2.5	1-2(3)-3-4(5)	1	1.03	1-(1)2(3)-3-(3)4-5	1	1.27	1-2(1-3)-3-4-5	1	0.85	1(2-3)-1-2-4(3)-5	1	1.18
2.6	1-2(3)-4(3)-5	1	1.03	1-2(1-3)-2-4-4(5)	1	1.27	1-2(3)-3(4)-5	1	0.85	1(2-3)-2-3(4)-4-4(5)	1	1.18
2.7	1-2(3)-4(5)	1	1.03	1-2(1-3)-3-4-3-5	1	1.27	1-2(3)-3(4)-5-4(5)	1	0.85	1(2-3)-3(4)-4-5	1	1.18
2.8	1-2(3)-4-3(4-5)	1	1.03	1-2(3)-2-4(5)	1	1.27	1-2(3)-3-3(4)-4-5	1	0.85	1-1(3-2)-3-4-5	1	1.18
2.9	1-2(3)-5(4)	1	1.03	1-2(3)-3 -4-5	1	1.27	1-2(3)-3-3(4)-5	1	0.85	1-2-(3)2(1)-4-5	1	1.18
2.10	1-3(2)-4-5	1	1.03	1-2(3)-3(4)-4-3-5	1	1.27	1-2(3)-3-4(5)	1	0.85	1-2(3)-3(4)-5	1	1.18
2.11	1-3-2(3)-4-5	1	1.03	1-2(3)-3-3(4)-5	1	1.27	1-2-3-4-5	1	0.85	1-2(3)-3-4(3)-4-5	1	1.18
2.12	2(3)-1-3(4)-5	1	1.03	1-2(3)-3-4(5)	1	1.27	1-2-4(3)-5	1	0.85	1-2(3)-3-4-3-5	1	1.18
2.13				1-2(3)-4(5)-5	1	1.27	1-3(2)-3-4-4(5)-5	1	0.85	1-2(3)-4(3)-5	1	1.18
2.14				1-2-3-4-5	1	1.27	3(2-1)-3-4-5	1	0.85	1-3(1-2)-3-4-5	1	1.18
2.15				2(3)-1-3-4-3-5	1	1.27				1-3(1-2)-4(3)-4-5	1	1.18

No.	Move Patterns (EL)	Freq.	%	Move Patterns (ES)	Freq.	%	Move Patterns (LL)	Freq.	%	Move Patterns (TQ)	Freq.	%
2.16				2(3)-1-3-4-5	1	1.27				1-3(2)-4(5)-3(4-5)	1	1.18
2.17				2(3)-2(1)-4-4(5)-5	1	1.27				2(3)-1-4-5	1	1.18
2.18										3(1-2)-3-4(3)-5	1	1.18
2.19												
3	2, 3, 4	8	8.25	2, 3, 4	6	7.59	2, 3, 4	26	22.22	2, 3, 4	8	9.41
3.1	2(3)-3-4	2	2.06	2(3)-3-4	2	2.53	2(3)-3-4	17	14.53	2(3)-3-4	3	3.53
3.2	2(3)-3(4)-4	1	1.03	2(3)-4-3	1	1.27	2(3)-3-3(4)-4	2	1.71	2(3)-3(4)	1	1.18
3.3	2(3)-3-2(3)-4	1	1.03	2(3)-3(4)-4	1	1.27	2(3)-2-3-4	1	0.85	2(3)-3(4)-4	1	1.18
3.4	2(3)-3-3(4)	1	1.03	2(3)-3-4(3)	1	1.27	2(3)-3-4(3)-4	1	0.85	2(3)-3-3(4)-4	1	1.18
3.5	2(3)-4(3)-4	1	1.03	2(3)-3-4-3-4	1	1.27	2(3)-3-4-3(4)	1	0.85	2(3)-3-4(3)-4	1	1.18
3.6	2(3-4)	1	1.03				2(3)-3-4-3(4)-4-3(4)	1	0.85	2(3)-4	1	1.18
3.7	2-2(3)-3-4	1	1.03				2(3)-4	1	0.85			
							2(3)-4(3)-4	1	0.85			
							3(2)-3-3(4)	1	0.85			
No.	Move Patterns (EL)	Freq.	%	Move Patterns (ES)	Freq.	%	Move Patterns (LL)	Freq.	%	Move Patterns (TQ)	Freq.	%
4	1, 2, 3, 4	14	14.43	1, 2, 3, 4	8	10.13	1, 2, 3, 4	11	9.40	1, 2, 3, 4	9	10.59
4.1	1-2(3)-4	4	4.12	1-2(3)-3-4	3		1-2(3)-4	3	2.56	1-2(3)-4	1	1.18
4.2	1-2(3)-3-4	3	3.09	1-1(2-3)-3-4	1	1.27	1-2(1-3)-3-3(4)	1	0.85	1-2(3)-3-4	1	1.18
4.3	1(3)-2(3)-3(4)	1	1.03	1-2(1-3)-3-4	1	1.27	1-2(1-3)-3-4	1	0.85	1-(2)3-3(4)	1	1.18
4.4	1-2(3)-3(4)-4	1	1.03	1-2(3)-3-3(4)	1	1.27	1-2(3)-3-3(4)	1	0.85	1-2(3)-4-3(4)	1	1.18
4.5	1-2(3)-3-3(4)	1	1.03	1-2(3)-4	1	1.27	1-2(3)-3-4	1	0.85	2(3)-3-1-3(4)	1	1.18
4.6	1-2(3)-3-3(4)-4	1	1.03	1-3(4)-4-2	1	1.27	1-2(3)-3-4-3(4)	1	0.85	2(3)-1-3-4	1	1.18
4.7	1-2(3)-4(3)	1	1.03				1-3(2)-3-4	1	0.85	1-2(3)-3-4(3)-3	1	1.18

No.	Move Patterns (EL)	Freq.	%	Move Patterns (ES)	Freq.	%	Move Patterns (LL)	Freq.	%	Move Patterns (TQ)	Freq.	%
4.8	1-2-2(3)-3(4)	1	1.03				1-3-4-2	1	0.85			
4.9	1-3(2)-4-3-4-3-4	1	1.03				2(3)-1-3-4	1	0.85			
5	1, 2, 3, 5	5	5.15	1, 2, 3, 5	5	6.33	1, 2, 3, 5	2	1.71	1, 2, 3, 5	5	5.88
5.1	1-2(3)-5	4	4.12	1-2(3)-3-5	2	2.53	1-2(3)-3-5	1	0.85	1-2(3)-5	2	2.35
5.2	1-2(3)-3-5	1	1.03	1-2(3)-5	1	1.27	1-2(3)-3-(3)5-5	1	0.85	1-2(3)-3-5	2	2.35
5.3				3(2)-1-3-5	2	2.53				1-2(3)-3-3(5)	1	1.18
6	1, 2, 3	11	11.34	1, 2, 3	0	0	1, 2, 3	0	0	1, 2, 3	0	0
6.1	1-2(3)	7	7.22									
6.2	1(2)-3	1	1.03									
6.3	1-2(3)-3	1	1.03									
6.4	1-1(3)-2(3)	1	1.03									
6.5	1-2(3)-2-2(3)	1	1.03	-	0	0	-	0	0	-	0	0
7	2, 3, 5	7	7.22	2, 3, 5	1	1.27	2, 3, 5	1	0.85	2, 3, 5	2	2.35
7.1	2(3)-3-5	3	3.09	2(3)-3-5	1	1.27	2(3)-3-5	1	0.85	2(3)-3-5	1	1.18
7.2	2(3)-2-3-5	1	1.03							2(3)-3-3/2-5	1	1.18
7.3	2(3)-(3)5	1	1.03									
7.4	2(3)-3-2(3)-5	1	1.03									
7.5	2(3)-2(3)-3-5	1	1.03									
8	2, 3	5	5.15	2, 3	2	2.53	2, 3	0	0	2, 3	1	1.18
8.1	2(3)	1	1.03	2(3)-3	1	1.27				2(3)-3-2-3	1	1.18
8.2	2(3)-3	1	1.03	2(3)-2-3	1	1.27						
8.3	2(3)-2	1	1.03									

No.	Move Patterns (EL)	Freq.	%	Move Patterns (ES)	Freq.	%	Move Patterns (LL)	Freq.	%	Move Patterns (TQ)	Freq.	%
8.4	3-2(3)-3	1	1.03									
8.5	3(2)-2-2(3)-3	1	1.03									
9	1, 3, 4, 5	1	1.03	1, 3, 4, 5	0	0	1, 3, 4, 5	2	1.71	1, 3, 4, 5	2	2.35
9.1	1-4-3-4(3)-5	1	1.03				1-3-4-5	2	1.71	1-3-4-5	1	1.18
										3(1)-3-5-4	1	1.18
10	1, 3, 5	1	1.03	1, 3, 5	0	0	1, 3, 5	0	0	1, 3, 5	0	0
10.1	1-3-5	1	1.03									
11	1, 2, 4	0	0	1, 2, 4	1	1.27	1, 2, 4	0	0	1, 2, 4	0	0
11.1				1-2(4)-4-2=4	1	1.27						
12	1, 3, 4	0	0	1, 3, 4	1	1.27	1, 3, 4	0	0	1, 3, 4	0	0
12.1				1-(1)3-3-4	1	1.27						
		97	100		79	100		117	100		85	100

H. Complete List of Rhetorical Move Patterns in the PAC

Variations of Move Patterns in PAC (Arranged by Frequency)

No	Patterns	Freq.	%	No	Patterns	Freq.	%
1	2(3)-3-4-5	55	14.55	57	2-3(2)-3-4-5	1	0.26
2	1-2(3)-3-4-5	26	6.88	58	2-2(3)-3-4	1	0.26
3	2(3)-3-4	24	6.35	59	2(3-4)	1	0.26
4	1-2(3)-4-5	14	3.44	60	2(3)-5(3)-3-4-5	1	0.26
5	1-2(3)-4	9	2.38	61	2(3)-5(3)	1	0.26
6	2(3)-3-3(4)-5	8	2.12	62	2(3)-4-4(5)	1	0.26
7	2(3)-4-5	8	2.12	63	2(3)-4-3(4)-4-3(4)-5	1	0.26
8	1-2(3)-3-4	8	2.12	64	2(3)-4-3	1	0.26
9	1-2(3)-5	7	1.85	65	2(3)-3-5-4	1	0.26
10	1-2(3)-3-5	6	1.59	66	2(3)-3-5(3)-5	1	0.26
11	1-2(3)	6	1.59	67	2(3)-3-4-4(5)-5	1	0.26
12	1-2(3)-3-4-5	6	1.32	68	2(3)-3-4-4(5)-3	1	0.26
13	2(3)-3-5	5	1.32	69	2(3)-3-4-3-4	1	0.26
14	2(3)-3-4-4(5)	5	1.32	70	2(3)-3-4-3(4)-4-3(4)	1	0.26
15	2(3)-3-3(4)-4-5	5	1.32	71	2(3)-3-4-3(4)	1	0.26
16	2(3)-3(4)-4-5	4	1.06	72	2(3)-3-4(3)-5	1	0.26
17	2-2(3)-4-5	3	0.79	73	2(3)-3-4(3)-4-5	1	0.26
18	2(3)-3-3(4)-4	3	0.79	74	2(3)-3-4(3)	1	0.26
19	2(3)-3(4)-5	3	0.79	75	2(3)-3-3(4)-5(4)	1	0.26
20	2(3)-3(4)-4	3	0.79	76	2(3)-3-3(4)-4-4(3)-5	1	0.26
21	2(3)-2-4-5	3	0.79	77	2(3)-3-3(4)-3-5	1	0.26
22	2(3)-1-3-4-5	3	0.79	78	2(3)-3-3(4)	1	0.26
23	1-3-4-5	3	0.79	79	2(3)-3-3(2)-5	1	0.26
24	1-2(3)-3-4(5)	3	0.79	80	2(3)-3-2-3	1	0.26
25	1-2(3)-3-3(4)	3	0.79	81	2(3)-3-2(3)-4-5	1	0.26
26	1-2(3)-3(4)-5	3	0.79	82	2(3)-3-2(3)-4	1	0.26
27	1-2(1-3)-3-4-5	3	0.79	83	2(3)-3-2(3)-2-5	1	0.26
28	2-3-4-5	2	0.53	84	2(3)-3-1-3(4)	1	0.26
29	2(3)-4(3)-4-5	2	0.53	85	2(3)-3(4)-4-4(3)-5	1	0.26
30	2(3)-4(3)-4	2	0.53	86	2(3)-3(4)	1	0.26
31	2(3)-4	2	0.53	87	2(3)-3	1	0.26
32	2(3)-3-4-3-4-5	2	0.53	88	2(3)-2-3-4-4(5)	1	0.26
33	2(3)-3-4-3(4)-5	2	0.53	89	2(3)-2-3-4	1	0.26
34	2(3)-3-4(5)	2	0.53	90	2(3)-2-3-1-2	1	0.26
35	2(3)-3-4(3)-4	2	0.53	91	2(3)-2-3(4)-5	1	0.26
36	2(3)-2-3-4-5	2	0.53	92	2(3)-2-3(4)-4-5	1	0.26
37	2(3)-1-5	2	0.53	93	2(3)-2-3(4)-4(3-5)-5	1	0.26
38	1-2-3-4-5	2	0.53	94	2(3)-2-3(2)-3-5	1	0.26
39	1-2(3)-4(3)-5	2	0.53	95	2(3)-2(1)-4-4(5)-5	1	0.26
40	1-2(1-3)-3-4	2	0.53	96	2(3)-2	1	0.26
41	3-3(4)-1-2(3)-3-4	1	0.26	97	2(3)-1-4-5	1	0.26
42	3-3(2)-3	1	0.26	98	2(3)-1-3-4-3-5	1	0.26
43	3-2(3)-3-4-5	1	0.26	99	2(3)-1-3-4	1	0.26
44	3(2-1)-3-4-5	1	0.26	100	2(3)-1-3-3(5)	1	0.26
45	3(2)-3-4-4(5)-3-4-5	1	0.26	101	2(3)-1-3-3(4)-3-3(4)-4	1	0.26
46	3(2)-3-4(5)-5	1	0.26	102	2(3)-1-3(4)-5	1	0.26
47	3(2)-3-3(4)-4-5	1	0.26	103	2(3)-1-3-(3)2-3(4)	1	0.26
48	3(2)-3-3(4)	1	0.26	104	2(3)-1- 2	1	0.26
49	3(2)-3(4)-5	1	0.26	105	2(3)-1	1	0.26
50	3(2)-1-3-5	1	0.26	106	2(3)	1	0.26
51	3(2)-1-2(3)-3	1	0.26	107	1-4-3-4(3)-5	1	0.26
52	3(1-2)-3-4(3)-5	1	0.26	108	1-3-4-2	1	0.26
53	3(1)-3-5-4	1	0.26	109	1-3-2(3)-4-5	1	0.26
54	2-4(3)-5	1	0.26	110	1-3(4)-4-2	1	0.26
55	2-3-4-3(4)-5	1	0.26	111	1-3(2)-4-5	1	0.26
56	2-3-3(4)-5	1	0.26	112	1-3(2)-4-3-4-3-4	1	0.26

No	Patterns	Freq.	%	No	Patterns	Freq.	%
113	1-3(2)-4(5)-3(4-5)	1	0.26	139	1-2(3)-3-3(4)-5	1	0.26
114	1-3(2)-3-4-4(5)-5	1	0.26	140	1-2(3)-3(4)-5-4(5)	1	0.26
115	1-3(2)-3-4	1	0.26	141	1-2(3)-3(4)-4-5	1	0.26
116	1-3(2)-3(4)	1	0.26	142	1-2(3)-3(4)-4-3-5	1	0.26
117	1-3(1-2)-4(3)-4-5	1	0.26	143	1-2(3)-3(4)-4	1	0.26
118	1-3(1-2)-3-4-5	1	0.26	144	1-2(3)-3	1	0.26
119	1-3(1)-3-4	1	0.26	145	1-2(3)-2-4(5)	1	0.26
120	1-2-4(3)-5	1	0.26	146	1-2(3)-2-2(3)	1	0.26
121	1-2-2(3)-3(4)	1	0.26	147	1-2(3)-2(5)	1	0.26
122	1-2(3)-5(4)	1	0.26	148	1-2-(3)2(1)-4-5	1	0.26
123	1-2(3)-5(3)-5	1	0.26	149	1-2(1-3)-3-4-3-5	1	0.26
124	1-2(3)-5(3)	1	0.26	150	1-2(1-3)-3-3(4)	1	0.26
125	1-2(3)-4-3(4-5)	1	0.26	151	1-2(1-3)-2-4-4(5)	1	0.26
126	1-2(3)-4-3(4)	1	0.26	152	1-1(3-2)-3-4-5	1	0.26
127	1-2(3)-4(5)-5	1	0.26	153	1-1(2-3)-3-4-4(3)-4-5	1	0.26
128	1-2(3)-4(5)	1	0.26	154	1-1(2-3)-3-4	1	0.26
129	1-2(3)-4(3)	1	0.26	155	1(3)-2(3)-3(4)	1	0.26
130	1-2(3)-3-5(3)-5	1	0.26	156	1(2-3)-3-4-5	1	0.26
131	1-2(3)-3-5(3)	1	0.26	157	1(2-3)-3(4)-4-5	1	0.26
132	1-2(3)-3-4-3-5	1	0.26	158	1(2-3)-2-3(4)-4-4(5)	1	0.26
133	1-2(3)-3-4-3(4)	1	0.26	159	1(2-3)-1-2-4(3)-5	1	0.26
134	1-2(3)-3-4(3)-4-5	1	0.26	160	1(2)-3	1	0.26
135	1-2(3)-3-4(3)-3	1	0.26	161	1-(1)2(3)-3-(3)4-5	1	0.26
136	1-2(3)-3-3(5)	1	0.26	162	3-4	1	0.26
137	1-2(3)-3-3(4)-4-5	1	0.26	163	1-3-5	1	0.26
138	1-2(3)-3-3(4)-4	1	0.26				

I. Keywords with Zero Occurrence in Relation to the Reference Corpus

Table 1 M1 keyword list of NAC in relation to zero occurrence in M1 of PAC, arranged by log-L filtered by range

(Criteria: Log likelihood ≥ 6.63, Range ≥ 3)

	Word Form	NAC Raw freq.	PAC Raw freq.	Log-L	Range		Word Form	NAC Raw freq.	PAC Raw freq.	Log-L	Range
1	thai	13	0	27.18	11	5	due	5	0	10.45	5
2	thailand	13	0	27.18	11	6	effectively	5	0	10.45	5
3	employees	7	0	14.64	4	7	economic	4	0	8.36	3
4	aec	5	0	10.45	3	8	foreigners	4	0	8.36	3

Table 2 M2 keyword list of NAC in relation to zero word occurrence in m1 of PAC, arranged by log-L filtered by range

(Criteria: Log likelihood ≥ 6.63, Range ≥ 5)

	Word Form	PAC Raw freq.	NAC Raw freq.	Log-L	Range		Word Form	PAC Raw freq.	NAC Raw freq.	Log-L	Range
1	thai	72	0	150.55	51	5	objectives	8	0	16.73	8
2	bangkok	13	0	27.18	11	6	thammasat	8	0	16.73	8
3	satisfaction	13	0	27.18	9	7	analyze	7	0	14.64	7
4	employees	11	0	23.00	9	8	silpakorn	6	0	12.55	5

Table 3 M3 keyword list of NAC in relation to zero occurrence in m3 of PAC, arranged by log-L filtered by range

(Criteria: Log likelihood ≥ 6.63, Range ≥ 5)

	Word Form	NAC Raw freq.	PAC Raw freq.	Log-L	Range		Word Form	NAC Raw freq.	PAC Raw freq.	Log-L	Range
1	thai	116	0	242.55	58	10	officers	11	0	23.00	5
2	spss	38	0	79.46	38	11	convenience	9	0	18.82	9
3	percentage	32	0	66.91	32	12	majoring	8	0	16.73	6
4	bangkok	28	0	58.55	22	13	objectives	7	0	14.64	7
5	deviation	27	0	56.46	27	14	silpakorn	7	0	14.64	6
6	thammasat	25	0	52.27	21	15	checklist	6	0	12.55	6
7	employees	24	0	50.18	12	16	december	5	0	10.45	5
8	likert	21	0	43.91	20	17	technology	5	0	10.45	5
9	ltd	13	0	27.18	6						

Table 4 M4 keyword list of NAC in relation to zero occurrence in m4 of PAC, arranged by log-L filtered by range
(Criteria: Log likelihood ≥ 6.63, Range ≥ 5)

	Word Form	NAC Raw freq.	PAC Raw freq.	Log-L	Range		Word Form	NAC Raw freq.	PAC Raw freq.	Log-L	Range
1	thai	47	0	98.28	26	9	ranked	7	0	14.64	6
2	problem	26	0	54.36	17	10	wanted	7	0	14.64	5
3	needed	18	0	37.64	9	11	physical	6	0	12.55	6
4	satisfied	14	0	29.27	11	12	communicate	5	0	10.45	5
5	rated	12	0	25.09	10	13	intrinsic	5	0	10.45	5
6	having	10	0	20.91	9	14	media	5	0	10.45	5
7	satisfaction	9	0	18.82	8	15	suggestions	5	0	10.45	5
8	mail	8	0	16.73	3	16					

Table 5 M5 keyword list of NAC in relation to zero occurrence in m5 of PAC, arranged by log-L filtered by range
(Criteria: Log likelihood ≥ 6.63, Range ≥ 3)

	Word Form	NAC Raw freq.	PAC Raw freq.	Log-L	Range		Word Form	NAC Raw freq.	PAC Raw freq.	Log-L	Range
1	thai	21	0	43.91	15	7	recommended	5	0	10.45	5
2	employees	6	0	12.55	6	8	techniques	5	0	10.45	5
3	able	5	0	10.45	3	9	thailand	5	0	10.45	4
4	company	5	0	10.45	3	10	hotel	4	0	8.36	3
5	encourage	5	0	10.45	5	11	reliable	4	0	8.36	4
6	program	5	0	10.45	3	12	respondents	4	0	8.36	3

Table 6 M1 keyword list of PAC in relation to zero word occurrence in m1 of NAC, arranged by log-L filtered by range
(Criteria: Log likelihood ≥ 6.63, Range ≥ 6)

	Word Form	PAC Raw freq.	NAC Raw freq.	Log-L	Range		Word Form	PAC Raw freq.	NAC Raw freq.	Log-L	Range
1	development	19	0	16.46	17	17	g	10	0	8.66	7
2	into	19	0	16.46	17	18	genre	10	0	8.66	7
3	based	18	0	15.59	16	19	yet	10	0	8.66	10
4	between	17	0	14.73	16	20	learner	9	0	7.80	9
5	attention	15	0	12.99	13	21	new	9	0	7.80	8
6	education	15	0	12.99	15	22	paper	9	0	7.80	9
7	linguistic	15	0	12.99	12	23	task	9	0	7.80	8
8	while	15	0	12.99	15	24	tesol	9	0	7.80	7
9	but	14	0	12.13	14	25	themselves	9	0	7.80	8
10	assessment	13	0	11.26	6	26	article	8	0	6.93	8
11	feedback	12	0	10.40	8	27	competence	8	0	6.93	6
12	input	12	0	10.40	9	28	critical	8	0	6.93	7
13	where	12	0	10.40	12	29	despite	8	0	6.93	8
14	esp	11	0	9.53	9	30	develop	8	0	6.93	8
15	literature	11	0	9.53	11	31	rarely	8	0	6.93	8
16	focused	10	0	8.66	9						

Table 7 M2 keyword list of PAC in relation to zero word occurrence in m1 of NAC, arranged by log-L filtered by range
(Criteria: Log likelihood ≥ 6.63, Range ≥ 10)

	Word Form	PAC Raw freq.	NAC Raw freq.	Log-L	Range		Word Form	PAC Raw freq.	NAC Raw freq.	Log-L	Range
1	article	115	0	99.62	101	17	drawing	14	0	12.13	14
2	investigates	31	0	26.85	31	18	forms	14	0	12.13	13
3	education	23	0	19.92	21	19	investigation	14	0	12.13	13
4	context	22	0	19.06	20	20	role	14	0	12.13	13
5	genre	19	0	16.46	12	21	discusses	13	0	11.26	11
6	chinese	17	0	14.73	14	22	practice	13	0	11.26	12
7	describes	17	0	14.73	17	23	presents	13	0	11.26	13
8	spanish	17	0	14.73	15	24	processes	12	0	10.40	10
9	word	17	0	14.73	15	25	related	12	0	10.40	12
10	languages	16	0	13.86	15	26	attention	11	0	9.53	11
11	current	15	0	12.99	15	27	children	11	0	9.53	10
12	hong	15	0	12.99	14	28	interaction	11	0	9.53	10
13	individual	15	0	12.99	14	29	experiences	10	0	8.66	10
14	kong	15	0	12.99	14	30	results	10	0	8.66	10
15	may	15	0	12.99	14	31	small	10	0	8.66	10
16	over	15	0	12.99	13	32	time	10	0	8.66	10

Table 8 M3 keyword list of PAC in relation to zero occurrence in m3 of NAC, arranged by log-L filtered by range
(Criteria: Log likelihood ≥ 6.63, Range ≥ 10)

	Word Form	PAC Raw freq.	NAC Raw freq.	Log-L	Range		Word Form	PAC Raw freq.	NAC Raw freq.	Log-L	Range
1	n	40	0	34.65	25	58	lectures	11	0	9.53	9
2	spanish	36	0	31.19	26	59	processing	11	0	9.53	9
3	chinese	34	0	29.45	23	60	recognition	11	0	9.53	8
4	investigates	31	0	26.85	31	61	states	11	0	9.53	11
5	articles	28	0	24.26	18	62	times	11	0	9.53	11
6	interaction	26	0	22.52	23	63	translation	11	0	9.53	7
7	model	25	0	21.66	17	64	video	11	0	9.53	11
8	analyses	24	0	20.79	23	65	whereas	11	0	9.53	9
9	hong	23	0	19.92	21	66	account	10	0	8.66	9
10	input	23	0	19.92	13	67	audio	10	0	8.66	10
11	kong	23	0	19.92	21	68	capacity	10	0	8.66	7
12	children	22	0	19.06	17	69	complex	10	0	8.66	7
13	describes	22	0	19.06	21	70	disciplinary	10	0	8.66	8
14	languages	22	0	19.06	20	71	discussed	10	0	8.66	9
15	may	22	0	19.06	21	72	engagement	10	0	8.66	7
16	drawing	20	0	17.33	20	73	gains	10	0	8.66	9
17	measured	20	0	17.33	19	74	gap	10	0	8.66	8
18	processes	20	0	17.33	16	75	his	10	0	8.66	6
19	small	20	0	17.33	18	76	making	10	0	8.66	9
20	across	18	0	15.59	17	77	observed	10	0	8.66	10
21	conditions	17	0	14.73	12	78	professionals	10	0	8.66	10
22	experiences	17	0	14.73	13	79	relevant	10	0	8.66	10
23	new	17	0	14.73	15	80	shows	10	0	8.66	10
24	system	17	0	14.73	12	81	treatment	10	0	8.66	8
25	talk	17	0	14.73	9	82	vs	10	0	8.66	6
26	tesol	17	0	14.73	8	83	wm	10	0	8.66	6
27	authors	16	0	13.86	15	84	adults	9	0	7.80	7
28	feedback	16	0	13.86	11	85	associated	9	0	7.80	9
29	modeling	16	0	13.86	11	86	direct	9	0	7.80	6
30	noun	16	0	13.86	9	87	evaluated	9	0	7.80	9
31	planning	16	0	13.86	10	88	experiments	9	0	7.80	6
32	complexity	15	0	12.99	11	89	initial	9	0	7.80	8
33	peer	15	0	12.99	6	90	insights	9	0	7.80	9
34	fluency	14	0	12.13	12	91	irregular	9	0	7.80	6
35	our	14	0	12.13	6	92	long	9	0	7.80	8
36	presents	14	0	12.13	14	93	middle	9	0	7.80	8
37	sessions	14	0	12.13	12	94	month	9	0	7.80	8
38	critical	13	0	11.26	9	95	pronunciation	9	0	7.80	6
39	discusses	13	0	11.26	13	96	reflection	9	0	7.80	7

	Word Form	PAC Raw freq.	NAC Raw freq.	Log-L	Range		Word Form	PAC Raw freq.	NAC Raw freq.	Log-L	Range
40	esp	13	0	11.26	10	97	although	8	0	6.93	8
41	g	13	0	11.26	10	98	backgrounds	8	0	6.93	7
42	identity	13	0	11.26	9	99	draws	8	0	6.93	8
43	large	13	0	11.26	12	100	episodes	8	0	6.93	5
44	longitudinal	13	0	11.26	12	101	followed	8	0	6.93	8
45	pragmatic	13	0	11.26	10	102	gathered	8	0	6.93	8
46	structural	13	0	11.26	8	103	home	8	0	6.93	6
47	beliefs	12	0	10.40	10	104	illustrate	8	0	6.93	8
48	collaborative	12	0	10.40	8	105	influenced	8	0	6.93	7
49	comparison	12	0	10.40	12	106	journal	8	0	6.93	6
50	conversation	12	0	10.40	10	107	mandarin	8	0	6.93	6
51	novice	12	0	10.40	7	108	nature	8	0	6.93	8
52	value	12	0	10.40	10	109	norms	8	0	6.93	6
53	author	11	0	9.53	10	110	old	8	0	6.93	8
54	discussion	11	0	9.53	10	111	per	8	0	6.93	7
55	dutch	11	0	9.53	7	112	perspective	8	0	6.93	8
56	efficacy	11	0	9.53	6	113	swedish	8	0	6.93	7
57	interactions	11	0	9.53	10	114	via	8	0	6.93	7

Table 9 M4 keyword list of PAC in relation to zero occurrence in m3 of NAC, arranged by log-L filtered by range
(Criteria: Log likelihood ≥ 6.63, Range ≥ 6)

	Word Form	PAC Raw freq.	NAC Raw freq.	Log-L	Range		Word Form	PAC Raw freq.	NAC Raw freq.	Log-L	Range
1	analyses	27	0	23.39	26	16	interview	10	0	8.66	8
2	gains	23	0	19.92	13	17	languages	10	0	8.66	8
3	spanish	22	0	19.06	10	18	article	9	0	7.80	8
4	over	21	0	18.19	19	19	longer	9	0	7.80	8
5	forms	20	0	17.33	11	20	months	9	0	7.80	8
6	discourse	17	0	14.73	14	21	range	9	0	7.80	8
7	across	15	0	12.99	15	22	regular	9	0	7.80	7
8	feedback	14	0	12.13	11	23	we	9	0	7.80	8
9	children	13	0	11.26	8	24	acquisition	8	0	6.93	8
10	disciplinary	13	0	11.26	11	25	emerged	8	0	6.93	8
11	our	13	0	11.26	12	26	influenced	8	0	6.93	7
12	genre	12	0	10.40	6	27	institutional	8	0	6.93	8
13	writers	12	0	10.40	8	28	led	8	0	6.93	7
14	accurate	10	0	8.66	7	29	particular	8	0	6.93	8
15	evidence	10	0	8.66	9	30	posttest	8	0	6.93	7

Table 10 M5 keyword list of PAC in relation to zero occurrence in m5 of NAC, arranged by log-L filtered by range
(Criteria: Log likelihood ≥ 6.63, Range ≥ 6)

	Word Form	PAC Raw freq.	NAC Raw freq.	Log-L	Range		Word Form	PAC Raw freq.	NAC Raw freq.	Log-L	Range
1	article	36	0	31.19	35	12	memory	11	0	9.53	8
2	we	25	0	21.66	23	13	our	11	0	9.53	11
3	effects	17	0	14.73	12	14	conclude	9	0	7.80	9
4	i	15	0	12.99	9	15	discourse	9	0	7.80	8
5	model	14	0	12.13	14	16	meaning	9	0	7.80	6
6	frequency	13	0	11.26	6	17	nature	9	0	7.80	8
7	processing	12	0	10.40	9	18	changes	8	0	6.93	8
8	professional	12	0	10.40	10	19	evidence	8	0	6.93	8
9	first	11	0	9.53	10	20	interaction	8	0	6.93	7
10	genre	11	0	9.53	9	21	show	8	0	6.93	8
11	indicate	11	0	9.53	11	22	written	8	0	6.93	7

J. Comparative Keyword Lists with British National Corpus (BNC)

Table 1 Comparative keyword lists of the overall NAC VS the overall PAC VS the BNC

No	NAC VS PAC			NAC VS BNC			PAC VS BNC		
	Word	Freq.	Keyness	Word	Freq.	Keyness	Word	Freq.	Keyness
1	thai	197	411.68	english	634	4205.58	language	618	3557.46
2	english	634	185.74	students	498	3520.69	learners	313	3326.17
3	students	498	165.65	study	412	2420.44	english	522	2641.58
4	was	389	165.38	thai	197	2251.19	students	384	2060.27
5	the	2658	158.89	reading	218	1328.24	l	352	1951.52
6	strategies	151	149.75	strategies	151	1201.59	study	408	1904.25
7	respondents	85	148.46	questionnaire	114	1050.78	learning	264	1458.76
8	program	104	133.85	language	206	994.96	teachers	223	1081.45
9	questionnaire	114	112.42	learning	152	859.35	efl	90	1043.48
10	skill	71	111.43	respondents	85	732.43	findings	160	1041.96
11	thailand	57	109.88	skills	128	686.95	proficiency	92	998.01
12	most	157	108.30	findings	99	672.18	article	182	990.68
13	employees	56	107.82	program	104	668.39	results	191	751.36
14	reading	218	106.14	communication	112	656.53	writing	171	741.11
15	were	396	92.69	learners	69	631.08	research	226	718.64
16	used	175	91.88	motivation	80	630.50	corpus	86	705.81
17	skills	128	89.90	participants	86	629.21	vocabulary	94	698.43
18	subjects	64	87.03	efl	50	596.00	classroom	107	698.17
19	attitudes	82	83.11	data	140	581.44	participants	100	640.03
20	spss	38	79.41	analyzed	49	580.10	comprehension	77	637.86

Table 2 Comparative keyword lists of move 1

No	NAC VS PAC			NAC VS BNC			PAC VS BNC		
	Words	Freq	Keyness	Words	Freq	Keyness	Words	Freq	Keyness
1	thai	13	83.34	english	71	523.70	language	126	822.86
2	thailand	13	76.21	language	40	264.99	english	94	524.38
3	english	71	63.82	reading	26	180.73	learners	38	407.10
4	important	17	46.87	learners	15	169.10	learning	51	315.66
5	has	25	38.40	thai	13	148.28	students	53	284.60
6	reading	26	36.30	students	24	147.74	research	62	276.56
7	communication	17	33.50	thailand	13	135.59	efl	18	231.48
8	aec	5	32.05	learning	20	133.97	teachers	42	228.96
9	most	18	32.01	efl	9	125.48	l	41	214.05
10	been	18	29.87	communication	17	122.27	esl	11	150.16
11	is	42	28.90	strategies	13	106.69	contexts	18	145.84
12	strategies	13	26.66	motivation	10	88.68	writing	29	137.00
13	adhd	4	25.64	adhd	4	83.30	classroom	19	133.84
14	facebook	4	25.64	facebook	4	83.30	tesol	8	130.20
15	hotel	4	25.64	teachers	14	78.93	esp	11	122.46
16	have	19	24.79	skills	13	76.68	teaching	23	108.53
17	one	18	24.42	aec	5	71.01	proficiency	10	107.65
18	product	7	23.69	important	17	62.36	reading	24	107.16
19	motivation	10	22.62	proficiency	5	58.96	tasks	18	106.95
20	skills	13	21.19	native	8	58.22	second	38	105.33

Table 3 Comparative keyword lists of move 2

No.	NAC VS PAC			NAC VS BNC			PAC VS BNC		
	Words	Freq	Keyness	Words	Freq	Keyness	Words	Freq	Keyness
1	thai	72	347.34	english	186	1414.31	study	236	1493.71
2	investigate	63	194.70	study	162	1200.67	language	212	1360.59
3	english	186	179.90	thai	72	938.20	english	197	1161.60
4	study	162	177.70	students	110	823.50	learners	85	943.86
5	aimed	43	134.60	investigate	63	630.01	article	111	802.22
6	attitudes	38	108.59	aimed	43	361.43	this	333	525.28
7	this	149	98.07	attitudes	38	287.54	students	90	475.05
8	purpose	33	90.97	this	149	263.47	efl	34	436.68
9	at	69	89.85	efl	19	259.41	l	79	424.60
10	thailand	20	88.63	learning	38	239.21	learning	73	420.64
11	students	110	78.87	language	45	233.91	examines	41	394.87
12	towards	29	77.39	reading	38	229.19	investigates	31	348.39
13	program	26	70.91	learners	21	215.20	paper	65	295.00
14	problems	28	70.62	skills	34	208.19	teachers	58	287.64
15	employees	14	60.38	purpose	33	194.55	classroom	37	267.01
16	bangkok	12	57.89	comprehension	19	193.91	corpus	28	247.23
17	school	34	55.98	thailand	20	191.90	research	64	219.72
18	skills	34	55.21	program	26	184.04	investigated	29	215.78
19	to	241	52.29	communication	28	182.16	second	73	212.51
20	satisfaction	13	47.84	research	38	159.19	reports	46	208.77

Table 4 Comparative keywords of move 3

No.	NAC VS PAC			NAC VS BNC			PAC VS BNC		
	Words	Freq	Keyness	Words	Freq	Keyness	Words	Freq	Keyness
1	thai	114	377.46	english	282	1887.21	language	325	1903.15
2	questionnaire	112	231.84	study	279	1883.61	english	320	1744.27
3	was	244	220.82	students	234	1694.22	learners	158	1710.17
4	study	279	168.71	thai	114	1382.74	study	307	1673.00
5	program	75	167.48	questionnaire	112	1214.50	students	212	1177.79
6	were	255	161.71	data	131	737.93	l	151	792.34
7	spss	38	125.82	analyzed	49	661.58	article	128	781.38
8	data	131	124.16	program	75	557.42	learning	121	648.16
9	students	234	117.60	spss	38	540.30	corpus	68	620.69
10	thailand	37	113.68	investigate	63	520.25	efl	51	617.11
11	english	282	113.15	were	255	451.89	analysis	118	542.85
12	subjects	46	107.10	participants	55	445.21	teachers	105	496.60
13	investigate	63	106.22	thammasat	25	426.15	speakers	65	461.31
14	the	1186	103.50	strategies	55	420.18	participants	65	449.73
15	used	100	96.23	research	96	411.51	proficiency	40	432.11
16	instrument	41	95.53	questionnaires	36	370.75	vocabulary	55	426.74
17	at	121	93.95	reading	67	363.59	data	103	382.65
18	bangkok	28	92.71	thailand	37	335.16	classroom	56	370.90
19	percentage	32	91.59	learning	58	313.64	native	58	364.11
20	deviation	27	89.40	using	76	312.65	examines	43	357.27

Table 5 Comparative keywords of move 4

No.	NAC VS PAC			NAC VS BNC			PAC VS BNC		
	Words	Freq	Keyness	Words	Freq	Keyness	Words	Freq	Keyness
1	thai	48	172.12	students	181	1279.06	learners	92	1002.03
2	respondents	55	169.42	english	196	1236.49	results	125	682.61
3	most	105	166.44	strategies	67	560.07	l	114	653.82
4	revealed	71	129.23	reading	88	553.22	students	115	621.25
5	skill	45	126.60	respondents	55	537.22	language	117	574.26
6	the	1058	123.80	thai	48	520.77	findings	66	474.66
7	strategies	67	110.22	revealed	71	511.75	proficiency	35	406.69
8	listening	52	80.50	results	90	503.47	learning	68	358.47
9	students	181	80.21	findings	61	477.10	english	84	328.14
10	was	138	76.28	listening	52	379.29	analysis	65	290.95
11	that	277	69.50	motivation	39	330.19	showed	60	288.24
12	had	53	68.89	skills	57	327.55	participants	37	254.00
13	skills	57	68.14	skill	45	321.88	vocabulary	31	238.46
14	agreed	21	67.53	speaking	49	308.42	comprehension	26	224.15
15	reading	88	67.50	errors	37	285.74	writing	50	215.84
16	employees	19	60.55	language	63	282.26	nouns	22	215.24
17	motivation	39	58.98	study	65	276.17	reading	49	214.72
18	errors	37	54.01	proficiency	23	267.63	linguistic	34	214.32
19	highest	18	52.27	most	105	253.92	analyses	27	195.93
20	problem	26	52.20	communication	39	224.84	teachers	46	191.73

Table 6 Comparative keywords of move 5

No.	NAC VS PAC			NAC VS BNC			PAC VS BNC		
	Words	Freq	Keyness	Words	Freq	Keyness	Words	Freq	Keyness
1	should	38	123.70	english	83	565.17	findings	73	639.49
2	thai	21	116.98	students	63	452.62	implications	69	557.32
3	be	66	91.60	study	55	334.40	language	91	524.49
4	will	22	69.79	findings	34	301.78	learners	47	516.67
5	improve	18	56.70	reading	36	241.89	l	76	478.56
6	can	34	47.68	thai	21	240.78	discussed	49	321.67
7	help	19	45.09	learning	34	233.62	learning	48	285.19
8	english	83	45.05	teachers	28	171.58	study	59	273.03
9	reading	36	42.77	efl	12	163.25	pedagogical	20	249.85
10	recommendations	9	39.96	teaching	24	150.35	results	46	223.77
11	students	63	36.30	language	29	148.73	research	53	213.91
12	increase	11	32.18	strategies	17	133.61	proficiency	18	211.72
13	skill	9	30.71	results	25	131.66	article	34	200.96
14	findings	34	29.70	improve	18	116.62	teaching	35	189.13
15	strategies	17	29.19	learners	11	107.55	concludes	20	182.87
16	staff	8	27.88	communication	16	99.20	english	45	181.47
17	cover	5	27.85	research	24	97.98	students	39	181.31
18	recommended	5	27.85	should	38	97.53	teachers	36	180.87
19	employees	6	27.81	motivation	10	79.88	suggest	31	163.99
20	future	12	26.26	facebook	4	79.76	acquisition	23	158.28

Table 7 Comparative keywords of embedded moves

	NAC VS PAC			NAC VS BNC			PAC VS BNC		
Rank	Words	Freq	Keyness	Words	Freq	Keyness	Words	Freq	Keyness
1	thai	71	343.99	english	181	1370.63	language	242	1492.96
2	investigate	59	179.68	study	156	1148.09	study	241	1413.14
3	english	181	171.94	thai	71	924.92	learners	105	1153.22
4	study	156	166.41	students	113	854.51	english	209	1150.97
5	aimed	40	123.18	investigate	59	583.73	article	113	763.00
6	attitudes	36	101.51	aimed	40	331.41	students	120	640.11
7	this	143	89.69	efl	22	306.73	l	106	576.94
8	students	113	85.96	attitudes	36	269.38	learning	90	509.87
9	at	67	85.90	this	143	245.93	efl	36	447.33
10	purpose	31	83.94	learning	38	240.06	this	349	436.91
11	towards	29	77.95	reading	37	222.05	examines	41	373.79
12	program	26	71.41	learners	20	203.49	investigates	31	332.42
13	problems	28	71.15	language	40	199.67	classroom	46	328.11
14	thailand	16	70.10	skills	32	192.86	corpus	36	317.00
15	bangkok	12	58.14	program	26	184.62	analysis	68	300.18
16	school	34	56.57	purpose	31	179.66	teachers	63	291.10
17	skills	32	50.08	comprehension	17	170.16	paper	67	274.92
18	satisfaction	13	48.10	strategies	22	167.02	linguistic	40	258.77
19	find	17	43.73	teachers	30	164.86	proficiency	23	245.91
20	employees	10	41.93	speaking	25	159.72	writing	56	245.58

K. Lexical Bundles categorized by their structures

1. NP-based

1.1 NP with pre-modifier

NAC		PAC	
a foreign language (m1)	**the present study** (m5)	**a foreign language** (m3, em)	the control group (m4)
english language learning (m1)	the research instrument (m3)	a second language (m1, m2, m3, em)	**the present study** (m2, m3, em)
the academic year (m3)	this research study (m2, m3)	second language l learners (m2, m3)	
the english language (m1)			

***Please note that **the bold types** signal that the lexical bundles are found in both NAC and PAC*

1.2 NP with post-modifier fragment

NAC		PAC	
english as a (m1)	**the findings of** (m5)	a discussion of (m5)	second language l (m2)
findings from this study (m5)	the findings of the (m5)	a number of (m1)	speakers of english (m3)
findings of this study (m5)	the findings of this (m5)	an analysis of (em)	the acquisition of (m5)
most of the (m4)	the majority of (m4)	analysis of the (m4, em)	the context of (m1)
one of the (m1)	the purpose of (m3, em)	**english as a** (m1, m2, m3, em)	**the development of** (m1, m2, m5, em)
one of the most (m1)	the purposes of this (m2, m3. em)	english as a foreign (m2, m3, em)	the effects of (m1, m2, m3, em)
purpose of this (em)	the purposes of (m2, em)	english as a second (m1, m2, m3, em)	the extent to which (m2, m3, em)
purpose of this study (m2, m3, em)	the purposes of this (m2, em)	english for specific (m1)	the implications of (m5)
purposes of this study (m2, em)	**the results of** (m4, m5)	english for specific purposes (m1)	the implications of the (m5)
results of the study (m4, m5)	the results of the (m4, m5)	implications of the (m5)	the importance of (m5)
the development of (m5)	the subjects of (m3)	implications of the findings (m5)	the nature of (m5)
the findings from this (m5)	**the use of** (m2, m5, em)	implications of the study (m5)	the relationship between (m1, m2, m3, em)
		implications of these findings (m5)	**the results of** (m3, m4, em)
		learners of english (m2, m3)	the role of (m2)
		one of the (m1)	**the use of** (m1, m2, m3, m4, m5, em)
		results of the (m4)	

1.3 NP with pre- and post-modifiers

NAC		PAC	
important role in (m1)	students attitudes towards (m2, em)	a second language esl (m1) first language l (m2, m3, em)	native speakers of (m3) second language l (m1,m3, em)

1.4 Others

NAC		PAC
mean and standard deviation (m3) teaching and learning (m5)	that the students (m4)	-

2. PP-based

2.1 *of* + NP fragment

NAC		PAC	
of the students (m5) **of the study** (m2, m3, m4, m5) of this research (m2, m3, em)	**of this study** (m2, m3, m5, em) of the most important (m1)	of english as (m2) of english as a (m3) of the findings (m5)	**of the study** (m5) **of this study** (m5)

2.2 Other PPs

NAC		PAC	
as well as (m5) for further research (m5) from this study (m5) in addition the (m4)	**in order to** (m2, m3, m5, em) **in terms of** (m1, m4) **in this study** (m3)	as a foreign language (m2, m3, em) as a second (m1) as a second language (m1, m2, m3, em) **as well as** (m1, m3, m4, m5, em) for future research (m5) **in order to** (m3	in recent years (m1) in second language (m1, m2) **in terms of** (m4) **in this study** (m3) with a discussion (m5) with a discussion of (m5)

3. VP-based

3.1 VP-active

NAC		PAC	
be useful for (m5)	revealed that the (m4)	have implications for (m5)	reports on a study (m2)
find out the (m3, em)	showed that the (m4)	have investigated the (m1)	show that the (m4)
is one of the (m1)	was a questionnaire (m3)	indicate that the (m4, m5)	suggest that the (m5)
is the most (m1)	will be beneficial (m5)	reports on a (m2, m3, em)	

3.2 VP-passive

NAC		PAC
be used to (m5)	was used to (m3)	based on the (m5)
can be used (m5)		

3.3 (V) + to infinitive

NAC		PAC	
aimed to investigate (m2, m3, em)	to find out (m2, m3, em)	need to be (m5)	**to investigate the** (m2, m3, em)
aimed to investigate the (m2, em)	to improve the (m2, m3, em)		
to analyze the (m3)	to improve their (m5)		
to communicate with (m1)	**to investigate the** (m2, m3, em)		
to examine the (m2, em)	was to investigate (m2)		
to explore the (m2)			

3.4 Anticipatory *it/there* + verb

NAC	PAC	
-	it was found (m4)	there has been (m1)

4. Others

4.1 NP + V

NAC		PAC	
of the study revealed (m4)	study were to (m2, m3, em)	article concludes with a (m5)	the article concludes (m5)
of this study was (m2, m3, em)	that the students (m4)	article reports on a (m3, em)	the article concludes with (m5)
of this study were (m2, m3, em)	the data was (m3)	findings indicate that (m4, m5)	the results show (m4)
of this study will (m5)	the data were (m3)	findings suggest that (m4, m5)	the results show that (m4)
results revealed that (m4)	the findings revealed that (m4)	findings suggest that the (m5)	the results showed (m4)
study aimed to (m2, m3, em)	the results revealed (m4)	little is known about (m1)	the results showed that (m4)
study aimed to investigate (m2, m3, em)	the study revealed that (m4)	paper reports on (em)	the results suggest that (m5)
study revealed that (m4)	the study was (m3)	previous research has (m1)	this article describes (m2, em)
study aimed to (m2, m3, em)	this study aimed (m2, m3, em)	research are discussed (m5)	this article explores (m2)
study aimed to investigate (m2, m3, em)	this study was (m2, m3, em)	results are discussed (m5)	this article reports (m2, m3, em)
study revealed that (m4)	this study was to (m2, m3, em)	results show that (m4)	this article reports on (m2, m3, em)
study was a (m3)	this study were (m2, m3, em)	results show that the (m4)	this paper reports (em)
study was to (m2, m3, em)	this study were to (m2, m3, em)	results showed that (m4)	this paper reports on (m2)
study was to (m2, m3, em)	this study will (m5)	results suggest that (m4, m5)	this study examined (m2, em)
study was to investigate (m2, em)	who work in (m1)	study examined the (m2)	this study examines (m2, m3, em)
		study investigated the (m2)	this study investigated (m2, em)
		study investigates the (m2, m3, em)	this study investigates (m2, m3, em)
		that the participants (m4)	

4.2 Modifier

NAC	PAC
very important for (m1)	-

L. Lexical Bundles categorized by their functions

1. Referential

1.1. Intangible framing attributes

NAC	PAC
in terms of (m1, m4)	based on the (m5) **in terms of** (m4) the context of (m1) the extent to which (m2, em) the nature of (m5)

1.2. Tangible framing attributes

NAC	PAC
-	-

1.3 Identification/ identity specification

NAC		PAC	
is one of the (m1)* **one of the** (m1)*	one of the most (m1)*	**one of the** (m1)*	there has been (m1)

1.4. Qualifying

NAC		PAC
-	-	

1.5. Quantity specification

NAC		PAC	
is one of the (m1)*	one of the (m1)*	a number of (m1)	one of the (m1)*
most of the (m4)	one of the most (m1)*		
the majority of (m4)			

1.6 Referential text/ group

NAC		PAC	
from this study (m5)	of this study were (m2, m3, em)	**in this study** (m3)	**of this study** (m5)
in this study (m5)	of this study will (m5)	of this study (m5)	previous research has (m1)
of the study (m2, m3, m4, m5)	**the present study** (m5)	of the findings (m5)	the control group (m4)
of this research (m2, m3, em)	this research study (m2, m3)	**of the study** (m4)	**the present study** (m2, m3, em)
of this study (m2, m3, m5, em)			
of this study was (m3, em)			

1.7 Referential time/sequence

NAC		PAC	
for further research (m5)	the academic year (m3)	for future research (m5)	in recent years (m1)

1.8. Reporting/description/interpretation

NAC		PAC	
aimed to investigate (m2, m3, em)	the research instrument (m3)	a discussion of (m5)	study investigated the (m2)
aimed to investigate the (m2, m3, em)	the result of (m4)	article concludes with a (m5)	study investigates the (m2, m3, em)
find out the (m2, em)	the result of the (m4, m5)	article reports on a (m2, m3)	suggest that the (m5)
findings from this study (m5)	the results revealed (m4)	findings indicate that (m4)	the article concludes (m5)
findings of this study (m5)	the result revealed that (m4)	findings suggest that (m4, m5)	the article concludes with (m5)
purpose of this (em)	the study was (m3)	findings suggest that the (m5)	the effects of (m2, m3, em)
purpose of this study (m2, m3, em)	the subjects of (m3)	have implications for (m5)	the implications of (m5)
purposes of this study (m2, em)	this study aimed (m2, m3, em)	have investigated the (m1)	the implications of the (m5)
results of the study (m4, m5)	this study aimed to (m2, m3, em)	implications of the (m5)	**the results of** (m3, m4, em)
results revealed that (m4)	this study was (m2, m3, em)	implications of the findings (m5)	the results show (m4)
revealed that the (m4)	this study was to (m2, m3, em)	implications of the study (m5)	the results show that (m4)
show that the (m4)	this study were (m3)	implications of these findings (m5)	the results showed that (m4)
study was a (m3)	this study were to (m2, em)	indicate that the (m4)	the results suggest that (m4)
study was to (m2, m3, em)	to analyze the (m3)	it was found (m4)	this article describes (m2, em)
study was to investigate (m2, em)	to communicate with (m1)	little is known about (m1)	this article explores (m2)
study were to (m2, m3, em)	to examine the (m2, em)	paper reports on (em)	this article reports (m2, m3, em)
the data was (m3)	to explore the (m2)	reports on a (m2, m3, em)	this article reports on (m2, m3, em)
the data were (m3)	to find out (m2, m3, em)	reports on a study (m2)	this paper reports (em)
the findings from this (m5)	to improve the (m5)	research are discussed (m5)	this paper reports on (m2)
the findings of the (m5)	to improve their (m5)	results are discussed (m5)	this study examined (m2, m3, em)
the finding of this (m5)	to investigate the (m2, m3, em)	results of the (m4)	this study examines (m2, m3, em)
the findings revealed that (m5)	was a questionnaire (m3)	results show that the (m4)	this study investigated (m2, em)
the purpose of (m3, em)	was to investigate (m2, em)	results show that the (m4)	this study investigates (m2, em)
the purpose of this (m2, m3, em)	who work in (m1)	results suggest that (m4, m5)	to investigate the (m2, m3, em)
		show that the (m4)	with a discussion (m5)
		study have investigated (m1)	with a discussion of (m5)
		study examined the (m2)	

1.9. Multifunctional

NAC		PAC	
be used to (m5)	**the use of** (m2, m5, em)	an analysis of (em)	the importance of (m5)
the development of (m5)	was used to (m3)	analysis of the (m3, m4, em)	the relationship between (m1, m2, m3, em)
		the development of (m1, m2, m5, em)	the role of (m2)
		the effect of (m1)	**the use of** (m1, m2, m3, m4, m5, em)

Notes:

- The fragment *one of the most*** are listed twice in 1.3 Identification/ identity specification and 1.5 Quantity Specification because it serves for both functions in the context.

2. Discourse/Textual Organizers

2.1. Linking

NAC		PAC	
as well as (m5)	**in order to** (m2, m3, m5, em)	**as well as** (m1, m3, m4, m5, em)	**in order to** (m3)
due to the (m1)			
in addition the (m4)			

2.2. Topic introduction: Not found

3. Stance/Interpersonal/Impersonal Expressions

3.1. Epistemic stance

NAC	PAC
-	-

3.2. Attitudinal/modality stance

NAC		PAC
be useful for (m5)	of the most important (m1)	need to be (m5)
can be used (m5)	this study will (m5)	
important role in (m1)	very important for (m1)	
is the most (m1)	will be beneficial (m5)	

4. Research topic-related

NAC		PAC	
a foreign language (m1)	teaching and learning (m5)	**a foreign language** (m3)	english for specific purposes (m1)
english as a (m1)	that the students (m4)	a second language (m1, m2, m3, em)	learners of english (m2, m3)
english language learning (m1)	the english language (m1)	a second language esl (m1)	native speakers of (m3)
mean and standard deviation (m3)	use of reading strategies (m1)	as a foreign language (m2, m3, em)	of english as (m2)
students attitudes towards (m2, em)		as a second (m1)	of english as a (m3)
		as a second language (m1, m2, m3, em)	second language l (m1, m2, m3, em)
		english as a (m1, m2, m3, em)	second language l learners (m2)
		english as a foreign (m2, m3, em)	speakers of english (m3)
		english as a second (m1, m2, m3, em)	the acquisition of (m5)

M. Raw Data of 3-word and 4-word Lexical Bundles of Each Move

(After Exclusion of Overlap Bundles)

Move 1 Clusters

Min freq ≥ 3, min range ≥ 3 #Total No. of 3-Gram Types/Tokens: 12/53 #Total No. of 4-Gram Types/Tokens: 4/20				Min freq ≥ 4, min range ≥ 4 #Total No. of 3-Gram Types/Tokens: 19/104 #Total No. of 4-Gram Types/Tokens: 5/24			
			NAC				PAC
No.	freq	range	clusters	No.	freq	range	clusters
1	16	13	**one of the***	1	10	9	**english as a**
2	7	7	is one of the*	2	10	10	second language l
3	6	6	one of the most*	3	9	9	as a second
4	5	5	in terms of	4	8	8	a second language
5	4	3	**english as a**	5	8	8	english as a second
6	4	4	of the most important	6	6	6	as a second language
7	4	4	the english language	7	6	5	as well as
8	3	3	a foreign language	8	6	6	**one of the***
9	3	3	due to the	9	6	5	there has been
10	3	3	english language learning	10	5	5	previous research has
11	3	3	important role in	11	5	5	the context of
12	3	3	is the most	12	5	5	the use of
13	3	3	to communicate with	13	4	4	a number of
14	3	3	use of reading strategies	14	4	4	a second language esl
15	3	3	very important for	15	4	4	english for specific purposes
16	3	3	who work in	16	4	4	have investigated the
				17	4	4	in recent years
				18	4	4	in second language
				19	4	4	little is known about
				20	4	4	studies have investigated
				21	4	4	the development of
				22	4	4	the effects of
				23	4	4	the relationship between

***Notes**: 1. The LBs printed in bold occur both in the NAC and the PAC.*

2. The LBs having an asterisk are functionally classified into two functions.

Move 2 Clusters

Min freq ≥10, min range ≥10 #Total No. of 3-Gram Types/Tokens: 23/378 #Total No. of 4-Gram Types/Tokens: 9/152				Min freq ≥ 10, min range ≥10 #Total No. of 3-Gram Types: 25/391 #Total No. of 4-Gram Tokens: 9/122			
			NAC				PAC
No.	freq	range	clusters	No.	freq	range	clusters
1	39	39	**to investigate the**	1	33	33	second language l
2	37	37	of this study	2	30	28	english as a
3	28	28	study was to	3	27	27	this article reports
4	24	24	this study was	4	25	25	reports on a
5	23	23	study aimed to	5	25	25	the effects of
6	23	23	the purpose of this	6	22	22	a second language
7	20	20	this study aimed	7	21	21	this article reports on
8	18	18	of this study was	8	17	17	as a second language
9	18	18	purpose of this study	9	17	17	the present study
10	18	18	this study aimed to	10	16	16	study investigates the
11	18	18	this study was to	11	15	15	english as a second
12	17	17	aimed to investigate	12	15	15	this study examines
13	16	16	study were to	13	15	15	**to investigate the**
14	15	15	the purposes of	14	14	14	as a foreign language
15	15	15	to find out	15	14	14	this study investigates
16	14	14	study aimed to investigate	16	13	13	english as a foreign
17	13	13	in order to	17	13	13	this study investigated
18	13	13	of this research	18	12	12	study examined the
19	13	13	the purposes of this	19	12	11	the development of
20	13	13	was to investigate	20	12	11	the extent to which
21	12	12	aimed to investigate the	21	12	12	this study examined
22	12	12	of this study were	22	11	11	in second language
23	12	12	purposes of this study	23	11	10	the relationship between
24	12	12	students attitudes towards	24	11	10	the use of
25	12	12	study was to investigate	25	10	10	first language l
26	12	11	the use of	26	10	10	learners of english
27	12	12	this study were to	27	10	10	of english as
28	11	11	find out the	28	10	10	reports on a study
29	10	10	of the study	29	10	10	second language l learners
30	10	10	this research study	30	10	10	study investigated the
31	10	10	to examine the	31	10	10	the role of
32	10	10	to explore the	32	10	10	this article describes
				33	10	10	this article explores
				34	10	10	this paper reports on

Move 3 Clusters

Min freq ≥15, min range ≥15 #Total No. of 3-Gram Types/Tokens: 26/585 #Total No. of 4-Gram Types/Tokens: 7/138				Min freq ≥ 14, min range ≥14 #Total No. of 3-Gram Types/Tokens: 24/498 #Total No. of 4-Gram Types/Tokens: 6/135			
			NAC				PAC
No.	freq	range	clusters	No.	freq	range	clusters
1	58	36	this study was	1	53	49	english as a
2	38	15	to analyze the	2	34	34	a second language
3	37	17	the study was	3	34	34	second language l
4	32	17	this study aimed to	4	29	29	as a second language
5	28	15	was a questionnaire	5	27	26	the effects of
6	28	54	of this study	6	27	27	this article reports
7	27	25	of this study were	7	26	26	english as a second
8	25	17	the subjects of	8	25	25	a foreign language
9	23	17	the research instrument	9	25	25	reports on a
10	23	21	of the study	10	24	24	as a foreign language
11	22	18	this study was to	11	21	21	english as a foreign
12	21	27	study was to	12	21	21	this article reports on
13	20	37	to investigate the	13	20	17	the use of
14	19	18	purpose of this study	14	19	19	the present study
15	19	15	study were to	15	18	15	analysis of the
16	19	16	the data was	16	18	18	as well as
17	19	16	aimed to investigate	17	16	16	**in this study**
18	18	15	this research study	18	16	16	learners of english
19	18	15	of this research	19	16	14	native speakers of
20	18	19	this study aimed	20	16	15	of english as a
21	18	16	was used to	21	16	16	study investigates the
22	18	18	mean and standard deviation	22	15	15	first language l
23	17	23	the purpose of this	23	15	15	speakers of english
24	17	26	**in this study**	24	15	14	the relationship between
25	17	21	the data were	25	15	14	the results of
26	16	19	**in order to**	26	15	15	this study examines
27	16	28	the purpose of	27	15	15	to investigate the
28	16	15	study was a	28	14	14	article reports on a
29	16	16	the academic year	29	14	14	**in order to**
30	15	32	this study were	30	14	14	this study investigates
31	15	19	of this study was				
32	15	22	study aimed to				
33	15	18	to find out				

Move 4 Clusters

Min freq ≥10, min range ≥10 #Total No. of 3-Gram Types: 13/233 #Total No. of 4-Gram Tokens: 4/66				Min freq ≥ 8, min range ≥ 8 #Total No. of 3-Gram Types: 16/178 #Total No. of 4-Gram Tokens: 3/2848			
			NAC				PAC
No.	freq	range	clusters	No.	Freq	range	clusters
1	30	30	of the study	1	19	19	results show that
2	29	28	the results of	2	16	16	results showed that
3	24	23	revealed that the	3	15	15	as well as
4	21	16	**in terms of**	4	13	12	the use of
5	21	21	the results of the	5	12	12	show that the
6	18	18	results of the study	6	12	12	the results of
7	18	18	study revealed that	7	11	11	findings indicate that
8	18	18	the findings revealed that	8	10	10	**in terms of**
9	16	15	showed that the	9	10	10	the results show
10	14	14	in addition the	10	10	10	the results showed that
11	14	13	that the students	11	9	9	analysis of the
12	14	14	the study revealed that	12	9	9	findings suggest that
13	13	13	most of the	13	9	9	results of the
14	13	13	of the study revealed	14	9	9	results show that the
15	13	13	results revealed that	15	9	8	results suggest that
16	13	11	the majority of	16	9	9	the results show that
17	10	10	the results revealed	17	8	8	indicate that the
				18	8	8	it was found
				19	8	8	the control group

Move 5 Clusters

Min freq ≥ 4, min range ≥4 #Total No. of 3-Gram Types/Tokens: 19/116 #Total No. of 4-Gram Types/Tokens: 8/38				Min freq ≥4, min range ≥ 4 #Total No. of 3-Gram Types/Tokens: 24/48 #Total No. of 4-Gram Types/Tokens: 8/201			
			NAC				PAC
No.	freq	range	clusters	No.	freq	range	clusters
1	11	11	**of the study**	1	15	15	the article concludes
2	11	11	**of this study**	2	14	14	**as well as**
3	10	10	the results of	3	14	14	implications of the
4	7	7	**as well as**	4	12	12	findings suggest that
5	7	7	can be used	5	12	12	**of the study**
6	7	7	from this study	6	12	12	the importance of
7	7	7	in order to	7	11	11	results suggest that
8	7	7	the results of the	8	11	11	the implications of
9	7	7	to improve their	9	8	8	a discussion of
10	6	6	this study will	10	8	8	based on the
11	5	5	be useful for	11	8	8	implications of the study
12	5	5	findings of this study	12	8	8	suggest that the
13	5	5	for further research	13	8	8	the article concludes with
14	5	5	results of the study	14	8	8	**the development of**
15	5	5	the findings of this	15	7	7	the implications of the
16	5	5	to improve the	16	6	6	have implications for
17	4	4	be used to	17	6	5	need to be
18	4	4	findings from this study	18	6	6	of the findings
19	4	4	of the students	19	6	6	**of this study**
20	4	4	of this study will	20	6	6	results are discussed
21	4	4	teaching and learning	21	6	6	the acquisition of
22	4	4	**the development of**	22	6	6	**the use of**
23	4	4	the findings from this	23	6	6	with a discussion
24	4	4	the findings of the	24	5	5	article concludes with a
25	4	4	the present study	25	5	5	findings suggest that the
26	4	4	**the use of**	26	5	5	for future research
27	4	4	will be beneficial	27	5	5	implications of the findings
				28	5	5	implications of these findings
				29	5	5	research are discussed
				30	5	5	the nature of
				31	5	5	the results suggest that
				32	5	5	with a discussion of

N. Sample Teaching Material for Abstract Writing

Analysis of Research Abstracts

🕮 **An abstract** is a brief comprehensive summary of your research paper/study. The main components of an abstract comprise five parts, namely, Introduction, Purpose, Methods, Results/Findings, and Conclusions. Abstracts can act like a promotional discourse in that they serve as a 'hook' to encourage readers to read the whole research/article. Because the space for abstracts is often limited, ranging from 150 to 300 words, it should be succinctly written by highlighting the major points of your research.

Task 1

Directions: Read the following two research abstracts from the fields of Education and Applied Linguistics, and then discuss with your peers how they are different in terms of components, brevity and others. Please provide at least three differences.

Abstract 1:

(1) This study reports the investigated vocabulary learning strategies used by M.3* students at Satit Suansunandha School during the second semester of the academic year 2010. (2) The subjects participating in this study were 80 students from two classrooms. (3) They were asked to respond to a questionnaire relating to vocabulary learning strategies. (4) Twenty-five item questionnaires were designed based on Schmitt's taxonomy of vocabulary learning strategies, including five strategies: determination strategies, social strategies, memory strategies, cognitive strategies, and metacognitive strategies. (5) The data was analyzed by SPSS and described through descriptive statistics, including frequency, mean and standard deviations. (6) The findings revealed that all of the subjects used each of the five strategies: determination, social, memory, cognitive and metacognitive. (7) In addition, the study showed that high-achieving students preferred to use determination strategies. (8) This information will be useful for English language teachers at Satit Suansunandha School by helping them to develop effective vocabulary learning instructions and suggesting successful vocabulary learning strategies that students can adopt.

N_CR_2010_14 (159 words)

***Note**: M.3 stands for Mattayom 3, which is equivalent to secondary school grade 9.*

Abstract 2:

(1) While research on portfolio assessment (PA) has focused largely on the summative aspects of writing assessment, not much has been done to find out its formative potential. (2) Drawing upon student questionnaires and student and teacher interview data, this paper aims to explore the formative functions of PA and, specifically, how the formative potential of PA can be better utilized in the EFL writing classroom. (3) The findings of the study indicate that although students responded positively to the formative aspects of PA, they still preferred summative grading and tended to believe that grades were the best way to inform their current standards of writing. (4) The paper ends with a discussion on how the formative aspects of PA could be strengthened and how PA can integrate teaching and assessment to benefit students' learning of writing.

P_EL_2010_005 (133 words)

📖 According to Hyland's (2000, p. 67) framework of abstract analysis, there are five main communicative purposes (moves) generally found in research abstracts, as shown in the table below.

Move 1: Introduction	Establishes context of the paper and motivates the research.
Move 2: Purpose	Indicates purpose, thesis or hypothesis, and outlines the intention behind the paper.
Move 3: Methods	Provides information on design, procedures, assumptions, approach, data, etc.
Move 4: Product	States main findings or results, the argument, or what was accomplished.
Move 5: Conclusion	Interprets or extends results beyond the scope of the paper, draws inferences, and points to applications or wider applications.

Task 2

Directions: Read the two abstracts in Task 1 again and then divide the abstracts according to their communicative purposes (moves). If no information of a particular move is found, please put 'no' in the blank. Each sentence in the abstracts is numbered for your convenience so that you can fill in the blanks with sentence numbers (S1, S2, or S3).

Move	Abstract 1	Abstract 2
1- Intro		
2- Purpose		
3- Methods		
4- Product		
5- Conclusion		

Task 3

Directions: Rank the moves in terms of their importance and decide which moves can be optional. Compare and discuss your answers with your peers.

Task 4

Directions: Below is the revised version of Abstract 1. Notice and list what features have been omitted.

(1) This study reports on the investigated vocabulary learning strategies (VLS) used by secondary school students. (2) A total of 80 students from two classrooms were asked to respond to a twenty-five-item questionnaire relating to VLS. (3) The questionnaire was designed based on Schmitt's taxonomy of VLS, which includes five strategies: determination strategies, social strategies, memory strategies, cognitive strategies, and metacognitive strategies. (4) The data was analyzed using descriptive statistics, namely frequency, means and standard deviations. (5) The findings revealed that all of the respondents employed each of the five strategies. (6) In addition, the study showed that high-achieving students used determination strategies more frequently. (7) The findings suggest how VLS can be used to improve vocabulary teaching and learning, and which VLS was found to be more effective among these students. *(125 words)*

Suggested Answers:

Task 1

1. Abstract 1 is more detail-oriented. Some of these details can be omitted, such as the specific name of the place where the data was collected and the time of the data collection.
2. Abstract 2 is more concise than Abstract 1, with more complicated sentence structures.
3. Abstract 1 begins with the purposes of the study, while Abstract 2 begins with information on the literature review (introduction).
4. The term 'portfolio assessment' in Abstract 2 was shortened to 'PA'. However, the term 'SPSS' was found in Abstract 1 and the term 'EFL' in Abstract 2 without providing what they stand for, probably because they are more common abbreviations in the fields.

Task 2

Move	Abstract 1	Abstract 2
1- Intro	No	S1
2- Purpose	S1	S2 (the second half)
3- Methods	S2, S3, S4, S5	S2 (the first half)
4- Product	S6, S7	S3
5- Conclusion	S8	S4

******Note for teachers****: Embedded moves can be introduced using Sentence 2 as an example.*

Task 3

The answers may vary according to students' opinions. However, the findings from different research studies can be introduced for the categorization of moves into optional, conventional, and obligatory. For example, Siriganjanavong (2019) found that move 1 was optional, moves 2 and 3 were obligatory, and move 5 was conventional in both novice and professional abstract corpora. While move 4 was found to be obligatory in novice abstract corpus, it was found to be conventional in professional abstract corpus.

Please note that the criteria for categorizing these moves into three statuses are as follows: (1) optional (<60% in the data); (2) conventional (60-97% in the data); and, (3) obligatory (more than 97%).

Task 4

1. A proper noun representing the name of the school and the time of data collection were omitted.

2. The word 'M.3', an abbreviation for Mattayom 3 in the Thai educational system, was replaced by secondary school to make it more understandable to a wider audience.

3. Some sentences were combined and rephrased to make them more concise.

O. Guidelines and Templates for Abstract Writing

🕮 Guidelines for Abstract Writing

- Check with your institution/author guidelines regarding the abstract length and strictly follow the guidelines.
- Briefly say (1) what you and others know about the topic and why it is important; (2) what motivates your research or what the aims of your study are; (3) how it was done; (4) what was found; and, (5) what the findings mean and what the things to consider are.
- Check whether citations are permitted or common in the targeted journal by reading the author guidelines and studying previous publications.
- Avoid using acronyms/abbreviations because they are interpreted differently in different fields. For example, CA might be used to refer to 'Communication Apprehension', 'Conversation Analysis' or 'Cost Analysis'. DA might be referred to 'Discourse Analysis' or 'Drug Abuse'. If necessary, spell out the full term before using any abbreviation.
- Avoid using technical language in order to make your research accessible to wider audiences. For example, 'the study used **move analysis** for data analysis.' might be replaced by 'the study analyzed the data according to its **communicative purposes**.'
- Think of 4-8 keywords or key phrases to be included at the end of your abstract. These are known as 'indexing terms' (Swales and Feak, 2009), which will help your study more searchable and retrievable through electronic systems.

Below are model templates that novice writers can use as an aid to write their abstract in the fields of Education and Applied Linguistics. The templates are not intended to be fixed, so more information can be added as and if necessary.

Template 1

While there has been extensive research into **(one specific aspect of research)**, studies on **(another aspect of research that is still underexplored)** remain scarce. This *(paper/research/study)* aims to *(investigate/explore/report on)* **(your research topic)**. Data from *(number/educational level of participants)* was collected through *(classroom observation/ semi-structured interview/video-recording)* and *(questionnaire/multiple-choice test)*. A qualitative analysis *(showed/suggested)* that **(your findings)**, while a quantitative analysis *(confirmed/revealed) that* **(your findings)**. The paper *(concludes/suggests)* that **(your conclusion/suggestions/ implications)**.

Template 2

Previous research into **(a wider aspect of research** e.g., second language acquisition/foreign language learning**)** focused on the effects of **(an independent variable)** on **(a dependent variable)**. However, little is known about **(the research gap)**. The present study thus (*explores/examines/investigates*) **(your research interest**). Through an analysis of (e.g., *classroom observations/interviews/reading comprehension scores*), the findings (*reveal/show*) that **(your findings).** The study suggests with the implications that **(the specific method/input/treatment)** (*can be/might be/should be*) **(brief details of your implications).**