LEXICAL BUNDLES IN AUTHENTIC AND TEXTBOOK ENGLISH BUSINESS EMAILS : A CORPUS-BASED COMPARATIVE STUDY

Miss Aroonrung Siricharoen

A Thesis Submitted in Partial Fulfillment of the Requirements
for the Degree of Master of Arts Program in English as an International Language
(Interdisciplinary Program)
Graduate School
Chulalongkorn University
Academic Year 2016
Copyright of Chulalongkorn University

กลุ่มคำศัพท์ที่พบในจดหมายอิเล็กทรอนิกส์ธุรกิจภาษาอังกฤษที่ใช้จริงกับที่ปรากฏใน
ตำราเรียน: การศึกษาเชิงเปรียบเทียบตามแนวภาษาศาสตร์คลังข้อมูล

นางสาวอรุณรุ่ง ศิริเจริญ

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาศิลปศาสตรมหาบัณฑิต
สาขาวิชาภาษาอังกฤษเป็นภาษานานาชาติ (สหสาขาวิชา)
บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย
ปีการศึกษา 2559

Thesis Title	LEXICAL BUNDLES IN AUTHENTIC AND TEXTBOOK ENGLISH BUSINESS EMAILS : A CORPUS-BASED COMPARATIVE STUDY
By	Miss Aroonrung Siricharoen
Field of Study	English as an International Language
Thesis Advisor	Assistant Professor Raksangob Wijitsopon, Ph.D.

Accepted by the Graduate School, Chulalongkorn University in Partial Fulfillment of the Requirements for the Master's Degree

....Dean of the Graduate School
(Associate Professor Sunait Chutintaranond, Ph.D.)

THESIS COMMITTEE

....Chairman
(Associate Professor Wirote Aroonmanakun, Ph.D.)

....Thesis Advisor
(Assistant Professor Raksangob Wijitsopon, Ph.D.)

....External Examiner
(Assistant Professor Passapong Sripicharn, Ph.D.)

อรุณรุ่ง ศิริเจริญ : กลุ่มคำศัพท์ที่พบในจดหมายอิเล็กทรอนิกส์ธุรกิจภาษาอังกฤษที่ใช้จริง
กับที่ปรากฏในตำราเรียน: การศึกษาเชิงเปรียบเทียบตามแนวภาษาศาสตร์คลังข้อมูล
(LEXICAL BUNDLES IN AUTHENTIC AND TEXTBOOK ENGLISH
BUSINESS EMAILS : A CORPUS-BASED COMPARATIVE STUDY) อ.ที่
ปรึกษาวิทยานิพนธ์หลัก: ผศ. ดร. รักสงบ วิจิตรโสภณ, 90 หน้า.

งานวิจัยนี้มีจุดมุ่งหมายเพื่อศึกษากลุ่มคำศัพท์ (lexical bundles) ที่พบในจดหมาย
อิเล็กทรอนิกส์ธุรกิจภาษาอังกฤษที่ใช้จริงกับที่ปรากฏในตำราเรียนเพื่อดูว่ามีความเหมือนหรือ
แตกต่างระหว่างกันหรือไม่อย่างไร ข้อมูลที่ใช้ในงานวิจัยนี้ประกอบด้วยตัวอย่างจดหมาย
อิเล็กทรอนิกส์จากตำราเรียนภาษาอังกฤษธุรกิจจำนวน 77 เล่ม และตัวอย่างจดหมายอิเล็กทรอนิกส์
ที่ใช้ในองค์กรเอ็นรอน คอร์ปอเรชั่น (Enron Corporation) เป็นตัวแทนจดหมายอิเล็กทรอนิกส์
ธุรกิจภาษาอังกฤษที่ใช้จริง ผู้วิจัยใช้การจำแนกประเภททั้งโครงสร้างและการใช้งานตามแนวทาง
ของ Biber *et al.* (2004) และ Biber (2006) เป็นทฤษฎีต้นแบบ

ผลวิจัยชี้ให้เห็นว่าประเภทโครงสร้างของกลุ่มคำศัพท์ทั้งในตำราเรียนและในจดหมาย
อิเล็กทรอนิกส์ธุรกิจที่ใช้งานจริงมีความคล้ายคลึงกัน ในขณะที่ประเภทการใช้งานค่อนข้างแตกต่าง
กันอย่างมีนัยยะสำคัญ ถึงแม้ว่ากลุ่มคำศัพท์ที่ระบุในข้อมูลทั้งสองแหล่งมีศูนย์รวมอยู่ที่ *การใช้งาน
แบบพิเศษ* ซึ่งเกี่ยวข้องกับกิจกรรมที่เป็นการสื่อสาร อาทิ คำร้อง, การแสดงออกถึงความคาดหวัง,
การยื่นข้อเสนอ และอื่นๆ แต่ตำราเรียนภาษาอังกฤษธุรกิจยังนำเสนอจดหมายอิเล็กทรอนิกส์ธุรกิจ
ได้น้อยกว่าความเป็นจริง เนื่องจากขาดประเภทการใช้งานย่อยบางหมวด งานวิจัยนี้นำเสนอผลของ
งานวิจัยที่มีต่อการศึกษาและการสอนเขียนจดหมายอิเล็กทรอนิกส์ธุรกิจและคำแนะนำสำหรับ
งานวิจัยในอนาคต

สาขาวิชา	ภาษาอังกฤษเป็นภาษานานาชาติ	ลายมือชื่อนิสิต ....
ปีการศึกษา	2559	ลายมือชื่อ อ.ที่ปรึกษาหลัก ....

5887544620 : MAJOR ENGLISH AS AN INTERNATIONAL LANGUAGE
KEYWORDS: LEXICAL BUNDLES / BUSINESS ENGLISH / TEXTBOOKS / BUSINESS EMAILS

AROONRUNG SIRICHAROEN: LEXICAL BUNDLES IN AUTHENTIC AND TEXTBOOK ENGLISH BUSINESS EMAILS : A CORPUS-BASED COMPARATIVE STUDY. ADVISOR: ASST. PROF. RAKSANGOB WIJITSOPON, Ph.D., 90 pp.

This study aimed to investigate lexical bundles found in authentic English business emails and sample emails in business English textbooks to identify their similarities and differences. The data employed in this study were sample emails from 77 business English textbooks and those emails recruited from the Enron Corporation, representing authentic English business emails. The structural and functional categorizations of lexical bundles from Biber *et al.* (2004) and Biber (2006) were used as frameworks for the analysis.

The findings show that structural types of lexical bundles in textbooks and in authentic emails were quite similar while functional types were significantly different. Although lexical bundles identified in both sets of data are centered around those in the *Special functions* category, which are associated with communicative acts such as request, expectation expression, offer, etc., a broader range of functional categories were featured in authentic business emails than business English textbook emails. Some implications regarding methodology and business email instruction, and suggestions for further studies are provided.

Field of Study:	English as an	Student's Signature	....
	International Language	Advisor's Signature	....
Academic Year:	2016		

ACKNOWLEDGEMENTS

CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter 1

Introduction

This chapter aims to contextualize the present study by providing its background information on business English and details about the study, including research questions and definitions of key terms, such as lexical bundles, and lastly describing significance of the study in the area of business English as well as its pedagogical implications.

1.1 Background of the Study

Business English has widely been recognized by experts in both applied linguistic and business fields as an important area of study. In applied linguistics, it is a subtype of English for Specific Purposes or ESP, which focuses on English language that serves as a communication tool for specific genres or events (Nickerson & Planken, 2016). In the present study, the term 'business' applies specifically to corporate activities, including internal and external communication. In the business field, English has long been used as a language for business communication (Nickerson, 2005). The use of English language thus plays an important role in the business world. This has led to an increasing need for English as an international business language in the past decade, especially in countries of the Asia-Pacific region, including Thailand (Nickerson, 2010). Therefore, a large number of business and language programs offer a major in business English (St John, 1996).

Business English is taught in many universities around the world, including those in Thailand. Some Thai universities even offer BA programs in business English. They teach students not only business concepts but also provide training in language skills required in the business environment as these are seen as a contribution to students' achievement in business careers ("a university's webpage on the English for Business Communication Program," 2016). As can be seen, business English has been regarded as an important field of study at the tertiary level.

One of the central communicative genres in the business field is the email, which is regarded as a standard communication system in the business environment (Baron, 1998). This is also supported in a survey conducted in Hong Kong where business emails are the most frequently used English text in written genres in the workplace (Evans, 2010). A number of email guide books have been published to demonstrate how to write a proper email (e.g. Angell & Heslop, 1994; Bly, 1999; Booher, 2001), some of which pay attention to business-related emails (e.g. Brounstein, Bell, & Smith, 2007; Lindsell-Roberts, 2004). In the academic discipline, many studies have examined business emails since the 1990s (e.g. Baron, 1998; Gains, 1999; Gimenez, 2000, 2005, 2006). Most of them focus on exploring linguistic features in English business emails while a few studies attempt to create tasks for the purpose of business email instruction (i.e. Evans, 2012).

As part of instruction, business English textbooks are one of the essential teaching material resources (Harwood, 2010). To ensure the quality of textbooks, many publishers claim to offer what is 'real-use' in the business environment. For instance, authentic materials from *the Economist* are drawn to partly construct a series of Intelligent textbooks in business English (Trappe & Tullis, 2006). The issue of authenticity has been addressed academically as one of the central topics in ESP research on business English textbooks.

From the academic perspectives, the language represented in business English textbooks may not be able to cover certain aspects of language used in the real ESP contexts. Therefore, scholars attempted to investigate the similarities and differences between the language in real contexts and the one in textbooks. There are many frameworks adopted to deal with this issue. One of the approaches that scholars employ to address this topic is 'lexical bundles'. This is because lexical bundles can be extracted automatically without individual judgement, they can show what linguistic units of meaning are truly common in a particular discourse and how one text type differs from others. In other words, lexical bundles are salient linguistic elements that characterize and differentiate text types (Conrad & Biber, 2005). Sample studies using this approach are those that explore introductory engineering textbooks (L. Chen, 2008), business and engineering textbooks (Wood & Apple, 2014), and business English coursebooks (Sriumporn, 2015).

Studies on business English textbooks have so far looked at different genres of business correspondences, such as business meetings (e.g. Angouri, 2010; Williams, 1988), business letters (i.e. Sinturat, 2010), collaborative writing (e.g. Bremner, 2010), etc. However, to the best of my knowledge, there has not been any study on samples of English business emails provided in business English textbooks.

To bridge the gap in research on business English discipline and instruction, the present study proposes to conduct a corpus-based comparative study on authentic and English business email samples presented in business English textbooks through an analysis of lexical bundles in the two sources. Since lexical bundles are frequently found in certain registers, they are considered a useful linguistic device for language learners (Biber & Barbieri, 2007). Focusing on lexical bundles in authentic and textbook emails can therefore reveal distinctive characteristics of business emails from each source. Most important of all, lexical bundles, which are a kind of multi-word expressions, can help learners produce more natural business emails in English, as Wray (2002) argues that multi-word expressions can help learners produce the language fluently. To summarise, the present study seeks to reduce a lack of empirical studies on business emails in textbooks through a corpus linguistic approach by focusing on lexical bundles.

1.2 Research Questions

1. What structural types of lexical bundles are there in authentic business emails and those in business English textbooks?

2. What functional types of lexical bundles are there in authentic business emails and those in business English textbooks?

3. What are similarities and differences between the forms and functional types of the lexical bundles found in authentic emails and those in business English textbook samples?

1.3 Objectives of the Study

1. To examine forms of lexical bundles found in authentic business emails and those in business English textbooks.

2. To investigate functions of lexical bundles found in authentic business emails and those in business English textbooks.

3. To compare and contrast the forms and functions of the lexical bundles found in authentic emails and those in business English textbook samples.

1.4 Scope of the Study

1. The study investigates English business email samples in textbooks. It excludes any language descriptions and exercises, which might be used for teaching English business emails.

2. The reference corpus draws on the Enron corpus project of University of California-Berkeley. The emails categorized for the project are claimed to retain their originality as released by the primary source.

1.5 Definition of Terms

1. *Emails*

An email is an electronic message which can be sent via a network. It is commonly used as a communication medium for many types of interactions, including business ones. In the present study, an email refers to a message sent between colleagues or business workers within or across the organization. The message has to focus on business matters such as placing orders, employment, meeting arrangement, etc.

2. *Lexical Bundles*

The term 'lexical bundle' was first introduced in *Grammar of Spoken and Written English* (Biber et al., 1999) as "recurrent expressions, regardless of their idiomaticity and regardless of their structural status" (p.990). In other words, lexical bundles are multiword sequences that recur in a certain text, which may not necessarily express idiomatic meanings or take the form of a specific kind of phrasal or clausal unit. Other terms for lexical bundles used by different scholars are *clusters* (Hyland, 2008a; Schmitt, Grandage, & Adolphs, 2004), *multiword construction* (Wood & Apple, 2014), and *n-grams* (Stubbs, 2007a; 2007b; also used in the corpus tool AntConc). As for the present study, following Biber *et al*. (1999), the term lexical bundle is adopted.

3. *Textbooks*

      The present study utilizes the term 'textbooks' in a broad sense. Textbooks refer to any texts, either commercial textbooks or materials compiled by the lecturers of business English courses, used to teach business English courses.

1.6 Significance of the Study

      The main purpose of the study is to contribute to the field of business English as well as English Language Teaching (ELT).

      First of all, this study can provide lists of lexical bundles used in authentic business emails. This can add empirical evidence of linguistic features of emails to the field of business English.

      Secondly, the study involves a specialized corpus containing sample emails used in business English textbooks, some of which were used in Thai universities. Such a corpus has not been compiled before. This contributes empirical data of business emails that can be studied in corpus linguistics or ESP.

      Lastly, findings from the study can provide pedagogical implications for business English instruction since it compares sample English business emails in textbooks and authentic ones, which might shed some light on how English language of business email is presented to business English learners, especially those in the Thai context.

Chapter 2
Literature Review

This chapter aims to discuss concepts relevant to the present study in detail, namely (1) business English, (2) business emails and (3) lexical bundles. In each of the following sections, theoretical frameworks and previous studies related to the above concepts are discussed to provide background information for the present study.

2.1 Business English

In this section, definitions of and approaches in business English are given, followed by a discussion on previous studies on business English and its pedagogical implications.

2.1.1 Definitions and Characteristics

The term 'business English' functions as an umbrella term to refer to any written or spoken business-related interactions that are conducted in English (Nickerson & Planken, 2016). Other scholars use different terms to refer to the concept similar to business English such as English for Business Purposes (EBP) (Nickerson, 2010) and English for Business Communication (EBC) (Bhatia & Bremner, 2012). The present study utilizes the term 'business English' since it is concise and well recognized not only in the business field but also in applied linguistics.

It is noteworthy that business English is a sub-field of English for Specific Purposes (ESP) since it addresses the use of English language in a specific context, in this case business context. Business English can be subdivided into two major types (Dudley-Evans & St John, 1998) on the basis of target learners: English for General Business Purposes (EGBP) and English for Specific Business Purposes (ESBP)

EGBP mainly focuses on linguistic elements in the business context. It is similar to general English language courses in that it teaches English but differs in terms of its focus on business setting (Dudley-Evans & St John, 1998). EGBP is suitable for students who lack professional business experience. On the other hand, ESBP courses are usually taken by learners with work experience. They aim to

improve one or two specific language skills such as writing a report, giving a presentation, etc. Therefore, ESBP tends to be offered to people in workplaces, e.g. in-house training by an external educator or a private course for senior personnel.

Business English is distinctive in that it can be a combination of specific content (i.e. business-related issues) and general content (i.e. general language ability) (Ellis & Johnson, 1994). A number of studies have been conducted in order to discover salient characteristics of different text types in business English. The next section explores major trends in doing research in this area.

2.1.2 Previous Studies on Business English

Business English has been studied extensively over the past few decades (Bargiela-Chiappini & Zhang, 2012; Dudley-Evans & St John, 1998; Hutchinson & Waters, 1987; St John, 1996). The main applied linguistic approach to business English has been the focus on text types, such as business letters (i.e. Hooi & Shuib, 2014; Jenkins & Hinds, 1987; Maier, 1992; Munir, 2001; Xu, 2012), business meetings (i.e. Yamada, 1990), business negotiations (i.e. de Moraes Garcez, 1993; Lampi, 1986), annual reports (i.e. J. Flowerdew & Wan, 2010; Kohut & Segars, 1992), and fax (i.e. Louhiala-Salminen, 1999; Warwick, 1992). Although business emails are relatively new when being compared with other traditional business text types like letters, they have been increasingly examined by numerous linguistic studies (i.e. Baron, 1998; Gains, 1999; Gimenez, 2000, 2005, 2006; Warren, 2016; Yue & Wang, 2014). Research on emails will be discussed in detail in section 2.2.

From an analytical point of view, studies on business English have been based on the approaches of genre analysis, discourse analysis and pragmatic frameworks (McCarthy & Handford, 2004; Nickerson, 2010). Jenkins and Hinds (1987), for instance, work on business letters from four different nations within the genre analytical approach. Their findings revealed that the main factor contributing to cross-cultural variations lies in the orientation (i.e. reader-, writer-, and nonperson-oriented). Maier (1992), on the other hand, based her study on a pragmatic framework, politeness strategy, while examining business letters by native and non-native English speakers. It was found that although non-native English writers can write grammatically flawless letters, they may be viewed negatively due to their "inappropriate use of politeness strategies" (Maier, 1992, p. 189).

The present study opted for corpus-based analysis since this approach is considerably useful when it comes to a large volume of data to be analyzed, which in this case is business emails. In addition, business emails were drawn from both real world business contexts and business English textbooks. Increasingly, the latter source has gained attention from researchers in business English, which will be elaborated below.

2.1.3 Research into Business English Textbooks

Despite the fact that business English places emphasis on the use of English for business interactions in the real world, as can be seen from a large number of studies and courses offered in the academic (i.e. business English programs) and professional settings (i.e. in-house training), research into business English textbooks usually found discrepancies between the contents in English for business purposes textbooks and the language in real business contexts (i.e. Chan, 2009; Donna, 2000; Harwood, 2010). Williams (1988), for instance, found a significant discrepancy between authentic language use and that taught in the textbooks for teaching meeting talk. In a study by Nelson (2000), business English textbooks were examined in comparison with a corpus of language used in real-life business settings, e.g. newspapers, journals, annual reports, emails, job interviews, etc. The findings revealed that the vocabulary in the business textbooks is limitedly represented in comparison to the authentic business-related language. A decade later, Angouri's (2010) contrastive study on business meeting language taught in the textbooks and those in the real setting still points to a similar direction, revealing that language prescribed in the business English textbooks does not correlate with language in real business use.

In the Thai context, there are a number of studies focusing on business English. Two studies are particularly relevant to the present study as they concern business English textbooks. The first one is on lexical phrases in business letters by Sinturat (2010) and the second on lexical bundles in business English textbooks by Sriumporn (2015). In her contrastive study, Sinturat (2010) examined lexical phrases found in model letters in business English textbooks used in Thai universities and those extracted from an online corpus of business letters. Her findings showed that two-thirds of the top 30 frequently used lexical phrases were shared between the two

sources. On the other hand, Sriumporn (2015) paid attention to lexical bundles found in business English coursebooks used in Thailand in comparison with a corpus of business articles retrieved from two Thai English newspapers. Based on her study, a low correlation was found between the two corpora. The present study is different from Sinturat (2010) and Sriumporn (2015) in that, this study investigates lexical bundles in business emails, which, to the best of my knowledge, has not been addressed by any studies that look at lexical bundles or business English textbooks.

As can be seen, business English textbooks might not be able to capture all essential linguistic elements found in real business settings. However, it does not mean that textbooks completely lack authenticity or usefulness. The practice of comparing teaching materials with authentic usage of certain business genres, on the contrary, can reinforce the utility of textbooks since such practice can identify strengths and weaknesses of contents in those materials. As a result, teachers and textbook publishers can improve their textbooks on the basis of comparative studies. One of the essential business communication tools that require urgent attention in this area is business emails since they have been increasingly used in the business world (Baron, 1998; Evans, 2010). A comparison between business emails in textbooks and those in real-life business settings is still scarce. The following section discusses central issues about business emails and previous studies on them.

2.2 Business Emails

Business emails are profound in the corporate world (Angell & Heslop, 1994, p. 1; Baron, 1998). The use of emails for business purposes was recognized in studies on business English as early as the 1980s, as exemplified by the work of Sherblom (1988) and still called for attention in more recent studies (e.g. Evans, 2012; Gimenez, 2000). In spite of the existence of email for some decades, characteristics of emails, either in general or business-related, seem to vary. For example, salutations can range from formal, i.e. *Dear* followed by the recipient's name, (Bly, 1999), specific use, i.e. *Greetings* used when sending a message to a group of people, (Angell & Heslop, 1994), to informal, e.g. *Hi*, *Hello,* or the first name of recipient (Lindsell-Roberts, 2004). On the other hand, in the research circle, many attempts have been made to define and categorize emails as well as prescribe how emails, especially in business settings, should be written. These aspects will be discussed below.

2.2.1 Definitions and Characteristics of Email

The term 'email' or 'e-mail' is an abbreviation of 'electronic mail'. These two terms can be used interchangeably. The term 'email' is adopted here since it is commonly used based on the frequency of this term as a search enquiry in the past few years (GoogleTrends, 2016). As for definitions, there are many approaches to define what an email is. Some focus on its production. As the term suggests, an email is a typed message sent via computer networks (Biber & Conrad, 2009; Merriam-Webster, 1993). Therefore, it is generally regarded as a written text. Others emphasize its technological property. For example, an email is regarded as an interactive communication tool in comparison to traditional ones, such as letters and memos. This is because the email can delimit the time and space of users in that it can be sent anytime to anywhere around the world regardless of the time and place of the email recipient. Nevertheless, the interactivity of email is not as high as conversation (Conrad & Biber, 2005) because of a lack of shared setting and time between users. It can be seen that emails are written messages but they function more like a spoken register owing to their quick and convenient response process, unlike traditional written communication tools such as letters, memos.

Perhaps, an important characteristic which distinguishes an email from traditional communication media, such as telephones, memos, and letters, lies in a combination of spoken and written registers in an email. Baron (1998), for example, used three different models to analyze language of emails. Her analysis showed that emails contain characteristics of both written and spoken languages. For example, although email is typed, email users tend to use it as if they are on telephone calls or having face-to-face conversation, as reflected by no editing and more informal style of writing. Therefore, due to their unique characteristics, business emails have been explored linguistically from different perspectives, which will be discussed in detail below.

2.2.2 Previous Studies on Business Emails

As mentioned in 2.2, emails have been a common standard of business communication in the past few decades. In pioneer research into business emails, linguistic features of authentic emails received particular attention. For example, Gains (1999) examined emails from two different domains, academia and commerce.

On the one hand, academic writers incorporated some discoursal features of conversations such as "dialogue device" (Gold, 1991 cited in Gains, 1999), e.g. *you see, well*, the use of rhetorical questions, e.g. *what do you think*, etc., into their email messages, reflecting flexible and relaxed writing styles. On the other hand, commercial emails seemed to follow standard written English, as evidenced by their use of grammar and punctuation. However, it must be noted that his data source was limited to a single company, resulting in small samples of emails. In contrast to Gains' (1999) study, Gimenez (2000) found that email possesses its own style of business writing. To illustrate such distinctions, he conducted a comparative study of business emails and business letters from the same company. Based on his examination, business emails tended to give less attention to punctuation, capitalization, and spelling. Also, he identified the hybridity of spoken and written language in business emails. For instance, some emails contained elliptical forms, such as "if interested, notify us accordingly" (Gimenez, 2000, p. 242). These grammatical features reflect the unplanned spoken register. This leads to Gimenez' (2005) subsequent conclusion that there were three factors that cause such contradictory results: the purposes of the messages, the relationship between email users, and the company's culture. In addition, Mallon and Oppenheim (2002) also found that contractions were most popularly used in personal emails and informal business emails. On the other hand, a small-scale study conducted by Danet (2002) observed that business emails still follow a template of business letter writing in some aspects like opening and closing, but some 'speech-like features' such as colloquial expressions (e.g. hello, hi) emerged in this text type . These studies illustrate how business emails were analyzed in late the 1990s up to the early 2000s.

After 2000, genre analysis (Bhatia, 1993) was often used as an analytical framework to see how business people write emails, as exemplified in Thaweewong (2006), Carrió-Pastor and Muñiz-Calderón (2013), Mehrpour and Mehrzad (2013), Moreno and Sznajder (2013) and Asztalos (2014). Also, textual analysis was conducted to discover how particular linguistic features are used in business emails as in Yu & Wang's (2014) work on hedging devices. Furthermore, Critical Discourse Analysis (CDA) was also used to investigate business emails. For example, Peterson, Hohensee, and Xia (2011) examined formality in workplace, using the Enron corpus

as their case study. They claimed that the formality of an email should not be necessarily defined by its nature being business or personal one. Two years later, a similar study was conducted by Knight, Adolphs, and Carter (2013), which revealed that emails are prone to "the more formal, written end" (p. 149) of continuum of formality in comparison to tweets, SMS, and discussion boards. Apart from that, Danielewicz-Betz (2016) investigated over 4,000 business emails in the Sales divisions from three international organizations. His findings focus on power and solidarity among managers and their subordinates exercised via linguistic elements. The findings show that some linguistic patterns are used more in certain groups. For example, the conditional sentence is more frequently used as a request form between co-workers or in external communication than by superiors to subordinates.

Although in the past few decades business emails have gained increasing attention in ESP studies, business English and World Englishes, a relatively small number of studies have looked at business emails in the pedagogical context. One of the few exceptions is Evans (2012), which made suggestions on email tasks for the business English classroom. Based on his data, comprising professional interviews, case studies, and four hundred business emails, Evans analyzed functions, structures and characteristics of business emails and found that they possess qualities similar to those of spoken language. He then proposed a simulation-based approach to teach students how to write business emails by using not only writing but all four fundamental language skills, namely writing, reading, speaking, and listening. In addition, research on errors in emails written by EFL writers were conducted: for example, in a case study of Thai EFL learners' business emails (Yoosawat & Tangkiengsirisin, 2016), it was found that the perceptions towards non-native English speakers' language errors vary, depending on several factors, for example the background of the participants (e.g. age, familiarity with non-native speakers of English) and the type of errors such as tone, verboseness, and choice of word. Nevertheless, business emails still seem to be on the margin of the pedagogical circle in applied linguistics research. Evans (2012) explained that although a number of studies focused on non-linguistic features of business emails, such as intercultural communication (Murphy & Levy, 2006), they were motivated by general interest, not by pedagogical considerations.

Therefore, it seems that while many types of business discourse have been approached, whether from a linguistic or from a pedagogical perspective, and that emails are an important type of business correspondence, work on business email instruction remains lacking. The present study aims to fill this gap by conducting a corpus-based contrastive study on business emails in business English textbooks and authentic business ones. The target textual element for the analysis is lexical bundles. The next section will discuss what lexical bundles are, why they are used in the present study to analyze business email samples, and how they can be categorized structurally and functionally. Moreover, previous studies on lexical bundles, including those in business English textbooks, will be discussed in more detail.

2.3 Lexical Bundles

Lexical bundle is one of the major descriptive tools in corpus linguistics, apart from collocation, keyword, etc. Though often seen as a function in corpus software, as they can be extracted automatically, lexical bundles can be regarded as a concrete reflection of the central theoretical concept in corpus linguistics. That is, through an investigation of corpus data, it has been found that meaning in language arises not from an occurrence of an individual word but from co-occurrences between two lexical items. In other words, units of meaning are created through different forms of co-occurrences, e.g. collocation (co-occurrence between lexical items) and colligation (co-occurrence between a lexical item and a grammatical category) (Sinclair, 2004). One of the co-occurrence patterns is "lexical bundle", a recurrent prefabricated unit consisting of a lexical and a grammatical item, e.g. *good morning* or *I don't think*. This will be discussed in detail below.

2.3.1 Definitions and Characteristics of Lexical Bundles

Lexical bundle refers to a group of words that occur in exactly the same sequence and form, and it is repeatedly used by multiple speakers or authors. This means that a lexical bundle must occur across several texts in a register (Conrad & Biber, 2005; Coxhead & Byrd, 2007) and that it is extremely common (Biber, 2009). A lexical bundle does not necessarily express an idiomatic meaning in the way that *kick the bucket* does. This is partly influenced by its frequency-based approach, which targets the multi-word sequences frequently used in a certain register. Lexical bundles are automatically extracted via corpus software, on a frequency-based parameter.

Details about the way in which lexical bundles are extracted automatically will be discussed in the Methodology chapter.

Since their introduction in 1999, lexical bundles have been found to shed light on various aspects of language use. First of all, because of frequency-based significance, the distribution of lexical bundles can differentiate text types and the language use in the target register. This leads to their second benefit, their salience in language learning. Although lexical bundles are neither idiomatic in meaning nor perceptually salient, they still function as important building blocks in a discourse (Biber, 2009; Biber & Barbieri, 2007). Because lexical bundles fulfill communicative needs of a particular register (Conrad & Biber, 2005), they are important for language learners to understand and make use of lexical bundles appropriate to a register they deal with.

The concept of lexical bundles thereby has been the focus in a great deal of language teaching and learning research (e.g. Cortes, 2006; Kazemi, Katiraei, & Rasekh, 2014). Cortes (2006), for example, taught four-word lexical bundles to university students in an intensive writing course. The findings of pre- and post-tests showed that students used a similar proportion of lexical bundles in both tests but their awareness and interest in these recurrent words were increased. Kazemi *et al.* (2014), on the other hand, found that teaching lexical bundles helped improve master's degree EFL Iranian students' writing. Given the lists of frequently used lexical bundles in the field of applied linguistics, the students were able to produce better writing tasks in their post-test compared to their pre-test. In addition, some studies offer methods to integrate lexical bundles in classroom instruction. For example, Barbieri and Eckhardt (2007) explain how to apply corpus-based findings to a form-focused model of instruction after they discovered that textbooks in their data cannot cover every angle of reported speech in real use, which is why lexical bundles can supplement the textbooks. The form-function relationship is a major characteristic of lexical bundles and will be elaborated below.

2.3.2 Classification of Lexical Bundles

To classify lexical bundles in different registers, a framework was developed by Biber *et al.* (1999). However, the framework focused only on grammatical based forms of bundles. Later in 2004, Biber, Conrad, and Cortes (2004) developed a framework for

analyzing the functions of lexical bundles, resulting in an improved framework covering both forms and functions. Therefore, there are two major approaches to lexical bundles: structural and functional. The former type concerns grammatical forms of the main word in a bundle while the latter looks at functional aspects of the bundle.

2.3.2.1 Structural Types of Lexical Bundles

Originally, a framework of structural types of lexical bundles was based on the registers in which conversation and academic prose were analyzed (Biber et al., 1999). It was improved in 2004 when Biber, Conrad, and Cortes classified structural types according to their grammatical forms into three main categories: (1) VP-based, (2) Dependent clause, and (3) NP/PP-based bundles. Each structural type is discussed below.

(1) VP-based bundles

As the name suggests, VP-based bundles incorporate verbal elements. Sample lexical bundles include a subject pronoun preceding a verb phrase (e.g. *it's going to be*) or a verb phrase as a whole bundle (e.g. *is going to be*), as well as question fragments (e.g. *what do you mean*).

(2) Dependent clause bundles

A lexical bundle incorporates a dependent clause fragment in addition to a verb phrase. For example, a main clause followed by a complementizer (e.g. *I want you to*, *I don't know if*), a WH-word starts before a dependent clause (e.g. *what I want to*), or a complementizer or subordinator starts at the beginning of the dependent clause bundles (e.g. *to be able to, if you look at*).

(3) NP/PP-based bundles

A lexical bundle in this category can incorporate either a nominal or prepositional phrase. In contrast to categories (1) and (2), which contain a clausal fragment, a lexical bundle in category (3) is phrasal. For example, a postmodifier is embedded at the end of the noun phrase within the lexical bundle (e.g. *the end of the*, *those of you who*), or a prepositional phrase embedded with modifiers (e.g. *by the end of*, *at the same time*).

2.3.2.2 Functional Types of Lexical Bundles

There are two functional frameworks for analysis of lexical bundles: Biber *et al.* (2004) and Biber (2006). The present study combines both frameworks since each

contains subcategories that can be adapted to the functional analysis of lexical bundles in business emails.

In Biber *et al.*'s (2004) work on university language, they proposed "three preliminary functions" of lexical bundles, namely: (1) Stance expressions, (2) Discourse organizers, and (3) Referential expressions. These three functions are used in the analysis of spoken and written language. The fourth category, Special Function, can be included or excluded, depending on the registers. For example, Biber (2006) employed this category to label functions of lexical bundles found in service registers.

(1) Stance expressions

Stance expressions or stance bundles can be divided into two main groups: Epistemic (degree of text producer's certainty) and Attitudinal/modality (attitude of text producer).

A. Epistemic stance bundles

Epistemic stance bundles express the certainty of the text producer towards the idea following the bundle in three main degrees: certainty, uncertainty, or probability. In other words, the bundles reflect the status of knowledge towards the information of the following proposition. They can be either personal or impersonal (Biber, 2006). For personal stance bundles, personal pronouns are given in the bundle (e.g. *I don't know if*, *you know what I*) whereas impersonal stance bundles also express certainty similar to personal one regardless of the presence of an overt speaker/writer (i.e. *are more likely to*, *the fact that the*).

B. Attitudinal/Modality stance bundles

This group of bundles conveys speaker perception towards the phenomenon or actions expressed in the following idea. There are five main subtypes: Desire, Obligation/directive, Imperatives, Intention/prediction, and Ability/effort. Each will be discussed with examples below.

- *Desire*: This subcategory demonstrates only personal expressions of stance. For example, the bundle *I don't want to* displays the self-motivated desire of the first person speaker while *do you want to* frames an inquiry concerning the other person's desire (Biber et al., 2004, p.390).

- *Obligation/directive*: This group of lexical bundles includes personal and impersonal expressions of obligations or directives. Usually, the personal

type contains a second personal pronoun (you), such as *you don't have to*, *you have to do*. The impersonal obligation/directive bundles exclude a personal pronoun, as in *it is important to*, *don't have to.*

- *Imperatives*: This group of bundles expresses a stronger degree of directiveness than obligation/directive bundle, e.g. *don't worry about it*, *just go ahead and*.

- *Intention/prediction*: This group of bundles tends to be personal expressions related to the text producer's own intention to perform an action in the future, for example, *we're going to do, are you going to*. As for the impersonal subtype, they can predict the future without text producer's volition, such as *it's going to be, not going to be*.

- *Ability/effort*: This subcategory concerns the ability and effort of the text producer, e.g. *to be able to*, *to come up with*.

(2) Discourse Organizers

Discourse organizers or discourse organizing bundles refer to those that assist in organization of discourse. This category offers frames for distinguishing lexical bundles into three groups: Topic introduction, Topic elaboration/clarification, and Conditions.

A. Topic introduction/focus

This subcategory signals the reader/audience to what is going to be discussed. The lexical bundle in this subcategory can contain first, second or no personal pronouns. For instance, *if you look at* draws the participant's attention to the proposition that follows the lexical bundle.

B. Topic elaboration/clarification

It is used when a speaker/writer wants to elaborate, clarify, or compare and contrast an issue, for example, *at the same time* is used to signal comparison and contrast of two things or events   whereas *has to do with* can be used to add new information to or clarification of the issues discussed.

C. Conditions

This function is found in lexical bundles consisting of the complementizer *if*. Its primary function is to signal conditional relations between two propositions, i.e. when

one event occurs, another event will follow. Sample bundles include *if you are a, if you do not,* and *if you have not*.

(3) Referential Expressions

Referential bundles serve as references to either physical or abstract units. They can be divided into five main types: Identification/focus, Imprecision, Specification of attributes, Time/place/text reference and Multi-functional reference. This functional type is found the most in academic prose (Biber, 2009, p.285).

A. Identification/focus

This group of bundles channels into the noun phrase following the lexical bundle. For example, *those of you who* specifies the subgroup of addressed participants.

B. Imprecision

These bundles make reference to entities in an imprecise manner. There are two specific roles of imprecision bundles: to suggest that a target reference is not strictly precise (e.g. *or something like that*) and to signal that other entities of the same kind could be found (e.g. *and stuff like that*)

C. Specification of attributes

This group of lexical bundles indicates "specific attributes of the following head noun" (Biber, 2006, p.145). They can specify quantity, as in *have a lot of*, *in a lot of*, or frame the noun phrase tangibly or intangibly, such as *the dean of the* and *in terms of the* respectively.

D. Time/place/text reference

Lexical bundles in this group include those that refer to time, place, or text within the data. For example, *the end of each, in the college of, as shown in Figure* are time, place, and text referential bundles, respectively.

E. Multi-functional reference

This function is performed through a bundle that can refer to more than one pragmatic function of referential bundles. For example, the bundle *at the end of* can function as either a place reference if it precedes a place or a time reference if a specific time is given after the bundle.

(4) Special Functions

Regarding the studies of Biber *et al*. (2004) and Biber (2006), this function has been so far identified only in the spoken register. The number of sub-categories in this

group varies from one study to another. In other words, the data-driven approach leads to some ad hoc categories, which might not be specified in every study on lexical bundles. For example, in Biber *et al.*'s (2004) study, three subtypes of Special Function are *Politeness*, *Simple inquiry* and *Reporting*. The *Politeness* category targets expressions that serve politeness strategies, as in *thank you very much*. *Simple inquiry* refers to an expression that serves to enquire, such as *what are you doing*. Last, *Reporting* focuses on the expression of the speaking reporting to someone, such as *I said to him*.

In sum, lexical bundles help fulfill communicative purposes (Conrad & Biber, 2005) and each structure or function can be found across registers or can be specific to some. This is one of the reasons why a number of studies have been conducted on lexical bundles for almost two decades, which will be discussed in detail in the following section.

2.3.3 Previous Studies on Lexical Bundles

Although lexical bundles were originally used to distinguish between academic prose and conversation in the pioneer research by Biber and his colleagues (1999), they have been applied in studies of other registers as well. This section illustrates the variety of previous studies on lexical bundles in general and those with specific focus on teaching materials, which are directly relevant to the present paper.

2.3.3.1 Previous Studies on Lexical Bundles in General

Lexical bundles are widely used as a descriptive tool in a number of text and discourse studies, including academic and professional contexts. There are two main approaches to lexical bundles for textual analysis: text producer-oriented and discourse-oriented. The text producer-oriented approach focuses on participants in a specific discourse such as non-native English language learners, native English speakers, expert writers, etc. On the other hand, the discourse-oriented approach pays attention to discourse variety such as research articles in different disciplines, conversations in the context of university, pharmaceutical discourse, etc.

A large number of text producer-oriented studies focus on L2 English learners as lexical bundles have been argued to be essential to language teaching and learning. As mentioned in 2.3.1, lexical bundles are discourse building blocks (Biber, 2009). Therefore, lexical bundles can be used as an index to determine students' English

language proficiency. In the 2000s, most of these studies are contrastive in nature, involving differentiation between native and non-native groups of participants, for example, Y.-H. Chen and Baker (2010) and Lie (2013) whose work focuses on L1 and L2 academic writing. Some studies in this line investigate not only English as a foreign language (EFL) learners, such as Chinese, but also compare to different English proficiency levels EFL learners in Huang (2015) and Ruan (2016).

There are also the text-producer oriented studies that focus on a comparison between language learners and expert writers. Cortes (2004), for example, identified lexical bundles found in students' writing in biology and history in comparison to published ones. This trend appears in recent work, including Öztürk and Köse (2016) and Pan, Reppen, and Biber (2016). Biber *et al.* (2016) introduced a new research design by comparing L1 English academic professional writers to L2 English ones. This focus illustrates how lexical bundles can be considered as a tool for an expert writing index. Sample studies investigating the lexical bundles in research articles include those by Cortes (2013), which focuses on various disciplines, by Sánchez (2014), which looks at research in biology, and by Mbodj-Diop (2016), which investigates lexical bundles in the medical discipline.

In sum, from a text-producer oriented approach, lexical bundles are used to identify the relationship between text and identities in both academic and professional writing as illustrated above. Therefore, it can be said that lexical bundles can be used to distinguish novice from expert writers as well as non-native English speakers from native ones. In order to teach English learners how to write academic essays, lexical bundles can be employed to guide students' writing to conform to the norm of academic writing as well as to conform to the discipline in which they study. The latter is discussed below.

The other approach, discourse-oriented, has a wider range and is more complex than the previous approach since any types of study on lexical bundles can fit into this approach. Nevertheless, work that adopts the discourse-oriented approach can be divided into two categories: register focus and discipline focus. The pioneer work of Biber *et al.* (1999) falls into the register focus category since they investigated different text types of English language from academic and general prose to fiction, conversation, and newspaper. In their analysis of lexical bundles, two registers

examined were conversation and academic prose. The findings revealed a large number of structural similarities and differences between the two registers. This work has inspired numerous scholars to look at lexical bundles in various registers. For example, classroom teaching and textbooks were compared based on lexical bundles extracted from the two registers (Biber et al., 2004). Other studies with similar focuses are those by Barbieri and Biber (2007), which looks at spoken and written university registers, e.g. office hours, class management talk, and by Gray and Biber (2013), which focuses on conversation and academic writing.

Within the discipline focus, scholars tend to investigate academic writing, looking at lexical bundles in different disciplines, such as science, engineering, etc. In terms of disciplinary variation, Hyland (2008b) examined four different disciplines, namely electrical engineering, biology, business studies, and applied linguistics, and found that lexical bundles can differentiate disciplines in the written domain. However, many scholars have looked at lexical bundles in only one discipline to identify the salient characteristics of a certain field. Sample studies in this trend include an analysis of lexical bundles found in the introduction section in medical research (Sadeghi et al., 2014). Apart from academic texts, lexical bundles have been adopted to look at professional discourses. For example, Jablonkai (2010) conducted a corpus-driven analysis of documents in English issued by EU institutions, such as the Commission, the Parliament, by looking at the most frequently found lexical bundles within those texts. Grabowski (2015) examined lexical bundles identified in "samples of patient information leaflets, summaries of product characteristics, clinical trial protocols and chapters from academic textbooks on pharmacology" (p. 23). This illustrates how lexical bundles have been employed in a broader area than academic discourse.

As can be seen, there has rarely been studies on lexical bundles in business correspondences. The present study thus is one of the few that apply the concept of lexical bundles to a very important business genre, that is email. In particular, there are only a few studies that pay attention to lexical bundles in textbooks, which are one of the primary materials in language teaching and learning. This will be discussed below.

2.3.3.2 Previous Studies on Lexical Bundles in Textbooks

The significance of textbooks lies in its combination of "pedagogic and disciplinary discourse" (Wood & Apple, 2014, p. 3). Textbooks, especially for specific

purposes, are expected to display not only the language fundamentals in a certain discipline but also language in real, disciplinary-specific use. Despite the fact that textbooks are one of the most important registers in academia, there is little research on lexical bundles of this text type (Biber et al., 2004, p. 374).

One of the early works in this area comes from a comparative study between classroom teaching and textbooks by Biber *et al.* (2004). They also compared the results with the previous results in Biber *et al.* (1999), in which conversation and academic prose were examined. Four types of registers were thus described in Biber *et al.*'s (2004) work. What they found was that lexical bundles in classroom teaching were the greatest in both type and volume while those in textbooks seemed to be less varied. In the past decade, studies have increasingly employed lexical bundles as a tool to investigate textbooks of different academic fields in order to identify whether they are sufficiently representative of language use in specified contexts. Two major approaches have been developed to answer the aforementioned question: comparison across textbooks and comparison between textbooks and other text types.

The first approach to lexical bundles in textbooks is a contrastive analysis of textbooks. For example, L. Chen (2008) investigated lexical bundles in electrical engineering introductory textbooks and those in ESP ones. The findings pointed out a gap between the two corpora concerning a lack of some functions of lexical bundles in the ESP materials. In addition, it is suggested that lexical bundles can be used as a criterion to evaluate authenticity of teaching materials. Another sample was from Wood and Apple (2014), which is a contrastive study on ESP textbooks from two disciplines, business and engineering, and English for Academic Purpose (EAP) textbooks. In this work, the terminology 'multiword constructions' was employed with similar definition of lexical bundles. The findings also showed that lexical bundles found in ESP and EAP textbooks differ from each other.

As said above, the other approach adopted in research on lexical bundles in textbooks is a comparison between English textbooks and authentic use of English language. This practice is common in the field of ESP. Using this method, researchers can identify differences between language in teaching materials and language in the real context (Nesi, 2012). For example, lexical bundles found in biology textbooks and those in research articles were compared in terms of distribution, structural and

functional types (Sánchez, 2014). One major finding of this study showed that textbooks provided a wider range of lexical bundles than research articles in the same field. The findings of this study supported the claim that lexical bundles constitute a discourse and also that textbooks can be representative of language used in this field. Another related study was conducted in the Thai context by Sriumporn (2015). Lexical bundles in business English coursebooks were compared against those found in business articles in English newspapers published in Thailand. It was found that there was a low correlation between the two corpora. One reason offered by the researcher was the fact that the business English coursebooks lacked authentic texts as well as local specific terms.

     In conclusion, lexical bundles have been used as an analysis unit to examine characteristics of particular text types, and compare two different text types under the same genre or vice versa. While there has been a call for many text types, research on lexical bundles in business emails has still been rare and hence is needed, especially in business English textbooks.
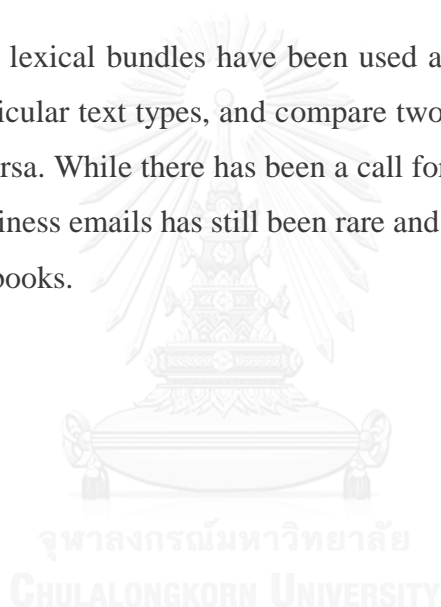
Chapter 3
Methodology

This chapter describes methodology involved in this study. It starts with a description of two corpora on which the study was based, followed by details regarding the selection and compilation of the corpora. In this respect, of particular importance are the issues of corpus size, identification of lexical bundles, and analytical frameworks.

3.1 Corpus Building

For the purpose of the present study, two corpora were compiled: textbook email corpus and authentic email corpus.

3.1.1 Textbook Email Corpus

The first corpus is Textbook Email Corpus (henceforth TEC) which contains 69,902 tokens with 751 business email samples, which were taken from 77 business English textbooks. Of these 77 textbooks, 16 were retrieved from a survey conducted to elicit the titles of business English textbooks used at Thai universities. I contacted 20 Thai universities and received 17 responses (See Appendix A for the list of universities). A total of 16 textbooks were being used in business English classes by 17 Thai universities (two of them using the same textbooks). The corpus thus contains a number of email samples that have really been used in business English classrooms. For those textbooks that were not identified as being used in classes, they are all published and commercial textbooks available in bookstores or in online versions. As TEC contains business email samples from 77 textbooks, it can be a good resource for teachers, who generally can use or rely on only a few textbooks.

There are two main criteria in selecting commercial textbooks: (1) textbooks must be related to language use in business context and published after 2000 A.D. and (2) they must contain at least one sample of business email. The procedure in extracting the email samples in business English textbooks can be divided into three stages. The first one is to scan all email samples provided within the textbook. The

second stage involves converting those images into plain text (.txt), using the Free OCR program, in order to build a corpus. The last stage is categorizing each email sample into a particular subtype of email. Those that are personal emails were not included in the corpus used in the present study.

3.1.2 Authentic Email Corpus

Authentic emails in this study are email samples from a large set of email messages from the Enron Corporation[1] (henceforth ENRON). The Enron dataset is regarded as the largest public dataset of authentic emails in the world (Leber, 2013), with more than half a million messages in total.

ENRON contains 1,061 texts with a size of 277,919 tokens. The data were based on the Enron database from 'UC Berkeley Enron Email Analysis Project' provided by University of California-Berkeley (Berkeley, n.d.). There are several benefits of using the data of UC Berkeley project instead of using the originally released and unmodified dataset. First of all, this database suits the purpose of the present study because it represents authentic use of emails. Although it comes from one company, there are various types of emails from a large group of writers. Moreover, this project was organized by UC Berkeley, which is one of the world's top 30 universities in 2016 according to many institutions, including QS Ranking, Times Higher Education, etc. Therefore, the data quality seems highly reliable. Most important of all, the data in this project were categorized and labelled to identify specific features or purposes of each email message. Consequently, UC Berkeley Enron dataset provided the descriptions of email message divided into several subcategories, such as business genre, tone, topics, etc., which allows for an effective comparative analysis of lexical bundles in different email types.

Although UC Berkeley Enron dataset seems to select mainly business-relevant emails, for example, company policy, meeting arrangements, reports, etc., other genres, e.g. news articles, advertisements, could be embedded within the message

---

[1] Enron was one of the world's biggest electricity, natural gas, and communications companies. In 2001, it was revealed that the company had committed fraud, a situation that became known as the Enron scandal. It has since become an infamous example of deliberate corporate fraud and corruption. Investigations into Enron brought to the public the release of corporate emails sent by 150 Enron employees.

body. Therefore, despite a number of advantages of using UC Berkeley Enron dataset, a preliminary analysis revealed that there were some irrelevant data based on the purpose of the present study, which focuses on business-related topics. For example, some emails were delivered along with news articles related to the corporation, resulting in a different genre embedded within business emails. In addition, every email was recorded along with codes as shown in the two boxes in Figure 3.1, resulting in repetitive information from headers of the email. Since the database provided by UC Berkeley still contains a large proportion of irrelevant data, selecting emails from UC Berkeley Enron dataset is a necessary procedure in order to ascertain that only relevant data are included in the reference corpus.
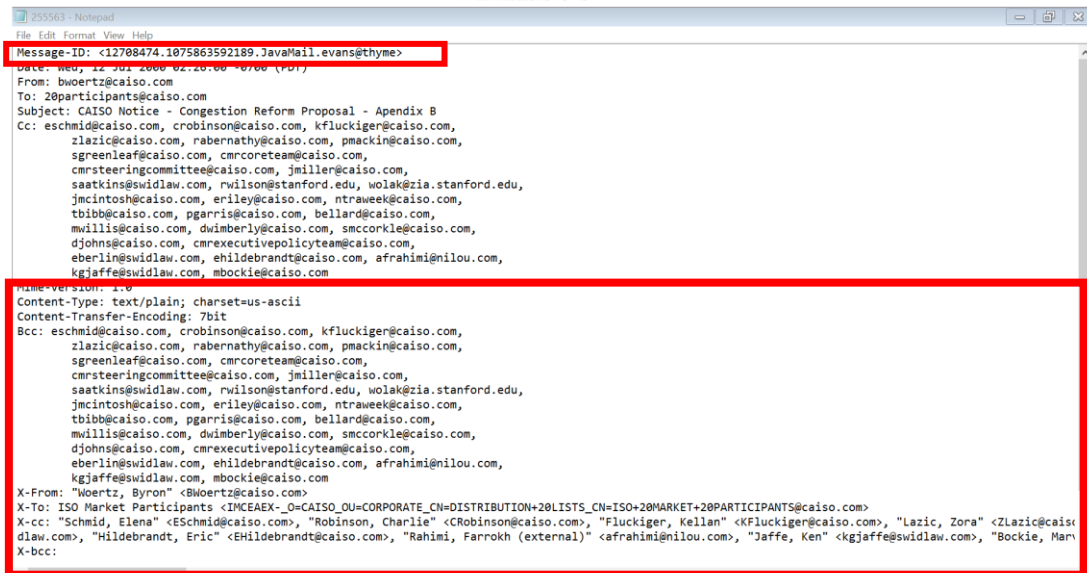
Figure 3.1 A sample of codes (in two frames) repeating the header of Enron email

There were three main criteria for recruiting emails from the UC Berkeley Enron dataset.

1) The body part of email contains at least four words and the minimum four words must not be proper nouns (i.e. London, James) or must not cross a sentence boundary (i.e. I accepted it. Thanks). This is in order to guarantee that the email message can yield a lexical bundle based on lexical bundle extraction criteria in 3.3.

2) Other irrelevant data such as news articles, codes, etc. will be eliminated during the process of building the authentic business email corpus.

3) The reply/forwarded emails embedded within the email can be included on the following conditions:

  a. The Subject line shows the term 'Re' or 'FW'. (See Figure 3.2)

  b. The most recent message (the upper one) refers to the contents of the embedded messages.
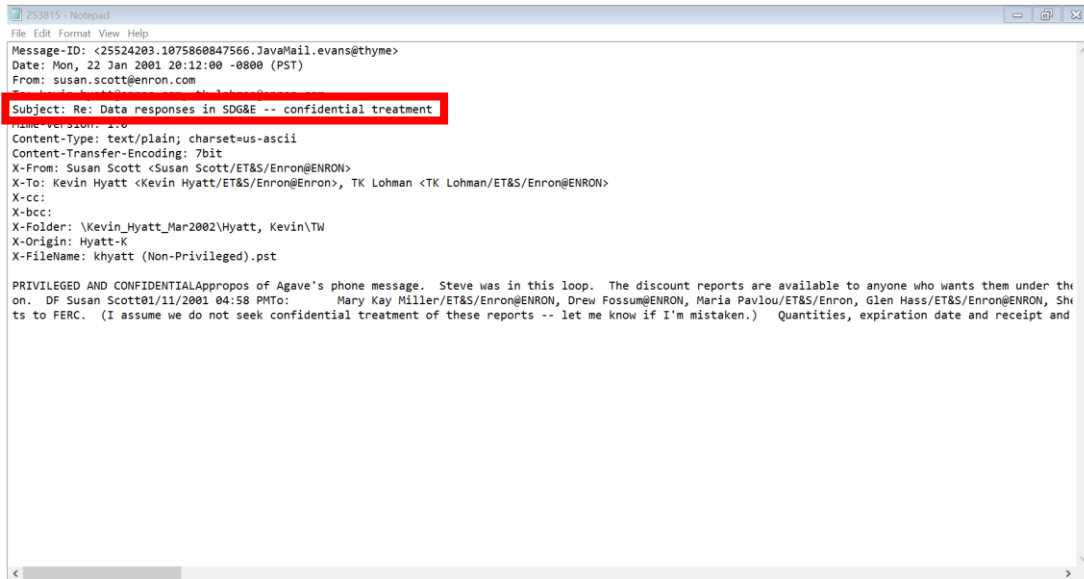
Figure 3.2 The subject line contains the word 'Re'

The emails of the two corpora then were categorized into five categories. These categories were based on the categories adopted in the UC Berkeley Enron dataset.

1. Company business, e.g. internal projects, company image, meeting minutes, etc.

2. Personal but in professional context, e.g. congratulatory email, thank-you email, etc.

3. Logistic arrangements, e.g. meeting scheduling, technical support, etc.

4. Employment arrangements, e.g. job seeking, hiring, recommendations, etc.

5. Document editing/checking

Although ENRON is three times larger than TEC in size, the number of texts (emails) in the former is only 30% bigger than the latter. In sum, both ENRON and TEC can thus be regarded as 'specialized' in that they are defined by a specific text

type (email), domain (business) and have been constructed with a specific purpose in mind (L. Flowerdew, 2008).

3.2 Software

In order to extract lexical bundles from each corpus, I used the corpus tool AntConc 3.4.4w (Anthony, 2004). This program is a freeware which is available at http://www.laurenceanthony.net/software/antconc/. It is free download and can work on computer running either Windows or Macintosh OS as well as Linux. The program was developed by Laurence Anthony at Waseda University in Japan and was first released in 2007. The software serves as a multiplatform tool for conducting research in corpus linguistics and data-driven learning.

For the present study I exploited the 'Clusters/N-Grams' tool[2] on AntConc for generating a list of lexical bundles in both corpora. I selected the 'N-grams' function because 'N-grams' can automatically generate a list of lexical bundles without prior identification of a search term. It functions by analyzing the data in the target corpus on two major criteria: (1) the length of a lexical bundle and (2) either the number of texts required to contain a lexical bundle or the frequency of a lexical bundle.

3.3 Lexical Bundles Extraction

As mentioned in 2.3, lexical bundles are multi-word sequences such as *as a result, I don't know if.* They were automatically extracted out of a corpus via AntConc. To extract the optimal number of types of lexical bundles, a threshold must be set up. It should be noted that there is no fixed criteria for extracting a lexical bundle. For the present study, several experiments have been conducted to come up with the appropriate threshold that yields the optimal number of types of lexical bundles. After several experiments, the following threshold was set up since it yielded the optimal and manageable results.

The first threshold deals with a target length of extracted lexical bundle. The present study adopts the four-word length for two main reasons. First, three-word bundles were embedded in four-word ones. For example, three-word bundles *as the result* and *the result of* were part of the four-word bundle *as the result of*. Therefore, it

---

[2] The 'Clusters/N-grams' tool consists of two major functions which are 'Clusters' and 'N-grams'. The 'Clusters' tool allows the user to search for a word or a group of words. It can be ordered by frequency, the initial or final word, the distributional range, the word length, etc.

is no use to include overlapping bundles which originate from the same longer expression. Second, a choice of bundles longer than four-word would reduce the number of bundles to be studied. This corresponds to many previous studies that show four-word lexical bundles are more commonly found than five-word bundles (Cortes, 2013). In fact, it should be noted that the selection of four-word bundle corresponds to several previous studies (e.g. Biber et al., 2004; Cortes, 2002; Hyland, 2008b).

The second threshold is distributional range. This concerns the number of texts required for a lexical bundle to appear. This criterion can help assure the representativeness of a lexical bundle since it is found across different texts, which are supposedly produced by different language users, within the same corpus. Unlike the length of word in a single bundle, the distributional range is rather arbitrary, depending on the characteristic of the corpus. For example, Biber and Barbieri (2007) recommended that a lexical bundle should occur in at least three to five different texts from corpora with the size of 50,000 to 200,000 words. On the other hand, Hyland (2008a) included the bundles which occur in at least 10% of total texts in his academic writing corpus of 120 research articles, which requires at least 12 research articles containing the target bundle.

In this study, a lexical bundle to be analysed was set to occur at least in 1% of the total email samples in each corpus. This is because after several experiments were conducted, it was found that the higher the number of texts required, the lower the number of lexical bundles that can pass. A slight change from 1% to 2% can reduce a large number of lexical bundles. For instance, 1% of the total texts in ENRON and TEC equals 11 and 8 texts respectively. However, 2% of the total texts in ENRON and TEC would equal 22 and 15 texts respectively, which turn out to yield less than 20 lexical bundles from each corpus. Therefore, the threshold of 1% of the total texts in both corpora suits the purpose of the present study. The cut-off distribution range set at 1% might seem small, compared with the number of text required in previous studies such as Biber *et al.* (2004), but it is because the corpora used in the present study are smaller and that is because they contain only one specific type of text which is business emails, not a variety of text type as found in Biber *et al.* (2004) or of discipline as identified in Hyland (2008a).

The distributional range has been applied not only to the whole corpus but also to the subcorpora as well, resulting in a different minimum distributional range requirement for each subcorpus. This is because each subcorpus contains a different number of texts. The minimum distribution range thus had to be adjusted to suit the size of each email category.

Table 3.1 The list of minimum distributional range required in each category

Category	Minimum Distributional Range	No. of texts		The minimal no. of texts	
		TEC	ENRON	TEC	ENRON
Overall	1%	751	1,061	8	11
1. Company business	1%	506	569	6	6
2) Personal but in professional context	2%	139	70	3	2
3) Logistic arrangement (meeting)	2%	73	265	2	6
4) Employment	5%	29	57	2	3

It must be noted that instead of using the distributional range, it is possible to draw on cut-off frequency as a criterion. However, the present study does not rely on the frequency because of considerable differences in the number of tokens between the two data sources. As mentioned in 3.1, the two corpora are three times different in size. As a result, the raw frequencies set in each corpora would be considerably different. For instance, if the normalized frequency is set at 20 times per million words, the minimum raw frequency of a lexical bundle would equal 71 times in ENRON and 250 times in TEC. Meanwhile, the distributional range offers a similar number of texts from the two corpora since their number of emails are 30% different.

Based on the two criteria, four-word length and distributional range, a list of lexical bundles in each corpus was generated. When examining the two lists, two categories of lexical bundles were excluded from further analysis: (1) lexical bundles that contain context-dependent words or proper nouns (e.g. *San Francisco in July,*

House Energy and Commerce) and (2) lexical bundles that go over clause or sentence boundary (e.g. *Wednesday, September would be*, *me know if you*).

In addition, overlapping lexical bundles generated from the same expressions were combined into five- or six-word lexical bundles. For example, the bundles *look forward to hearing, forward to hearing from*, and *to hearing from you* were part of a longer expression, *look forward to hearing from you*. Therefore, they were put into a six-word bundle after the results were analyzed.

3.4 Data Analysis

As shown in Chapter 1, the present study involves the analysis of forms and functions of lexical bundles. There are three main stages in this procedure. First, after lists of lexical bundles were generated, I excluded the bundles that did not meet the aforementioned criteria (See 3.3). Second, I categorized the rest of the lexical bundles on the list into structural and functional types on the basis of Biber *et al*. (2004)'s and Biber (2006)'s frameworks.

Along with this, to ensure the validity of the researcher's functional categorization of lexical bundles, two raters were asked to identify functions of the extracted lexical bundles to enhance the reliability for the functional categories. One rater was an EFL and business English professional native English teacher and the other was an advertising and marketing manager with 10 years of experience. It should be noted that the raters examined all lexical bundles identified in the two corpora. They were given a list of lexical bundles with my suggested categories. This is in order to give them an idea of the types of categories they had to deal with. In addition, they were also given concordance lines for each lexical bundle to aid their decision-making (See Appendix C for examples of validation sheets given to the raters). They then had to identify on the sheet whether they agreed with my categorization. If they did not agree on any particular lexical bundle, they were required to suggest a category. The suggestions from the raters were taken into my consideration and may be applied to adjustment of my category assignment. For example, the bundle *I look forward to* was categorized in the *Topic introduction/focus* at first. However, one rater suggested that it could be labelled as *Expectation*, which refers to the writer's expectation towards the reader's future action. On the other

hand, one rater might interpret the meaning in a different way from the researcher or from the other. This leads to 73% similarity in identification among the two raters and the researcher. Although the number might be considered not very high from a general point of view, this is because functional categorization is rather flexible and depending on an individual judgement.

Chapter 4
Findings and Discussions

This chapter presents findings obtained from structural and functional analyses conducted on the two corpora. The findings are reported and discussed in five main sections. Firstly, I report on the distribution of structural types of lexical bundles found in ENRON and TEC in 4.1 and 4.2, respectively. Later, functional types of lexical bundles from ENRON and TEC were categorized in 4.3 and 4.4, respectively, followed by a discussion of similarities and differences between lexical bundles from the two corpora.

4.1 Classification of Structural Types in ENRON

Based on the criteria in section 3.3, a total of 34 lexical bundles (See Appendix B for a full list of lexical bundles) were then divided into three main groups according to Biber *et al.*'s (2004) analytical framework, i.e. VP-based, dependent clauses, and NP/PP-based bundles, as explained in 2.3.2.1. Table 4.1 below illustrates the distribution of three main structural categories in ENRON, where VP-based bundles outnumber those in dependent clauses and NP/PP-based bundles, with 47%, 26.5%, and 26.5%, respectively. Lexical bundles in each category are listed and the categories that have no lexical bundle are accompanied with examples in parentheses. Each lexical bundle is accompanied with its raw frequency in parentheses.

Table 4.1 Structural types and proportional distribution in ENRON

Basic structural type	Structural Category and relevant lexical bundle	No. of types	Percentage (No.)
VP-based	**(connector +) 1st/2nd person pronoun + VP fragment** *you have any questions(58), I look forward to(33), I don't know(18), I don't think(18), I don't have(12), we look forward to(11)*	6	47% (16)
	(connector +) 3rd person pronoun/ noun + VP fragment *it would be a (13), attached is a draft(12)*	2	

	Discourse marker + VP fragment (e.g. *I mean you know*, *you know it was*)	0	
	Verb phrase with active verb *please let me know(70)*, *thank you for your(27)*, *have any questions or(14)*, *look forward to hearing(13)*, *is a draft of(12)*, *feel free to contact(11)*, *get in touch with(11)*	7	
	Verb phrase with passive verb (e.g. *is based on the*, *can be used to*)	0	
	Yes-no question fragments (e.g. *are you going to*, *do you want to*)	0	
	Wh-question fragments *what do you think(20)*	1	
Dependent clause	**(connector +) 1st/2nd person pronoun + dependent clause fragment** *I would like to(41)*, *I'd like to(15)*, *we would like to(12)*	3	26.5% (9)
	Wh-clause fragments *let me know what(11)**	1	
	If-clause fragments *let me know if(87)**, *if you have any(73)*, *if you want to(11)*	3	
	(Verb/Adjective+) to-clause fragment *please feel free to(31)**, *don't want to(12)**	2	
	That-clause fragments (e.g. *that there is a*, *that I want to*)	0	
NP/PP-based	**Noun phrase with *of*-phrase fragment** *the end of the(21)*, *a draft of the(16)*, *a copy of the(13)*	3	26.5% (9)
	Noun phrase with other post-modifier fragment *thanks for your message(13)*	1	
	Other noun phrase expressions (e.g. *a little bit more*, *or something like that*)	0	
	Prepositional phrase expressions *as a result of(16)*, *in the process of(13)*, *to hearing from you(13)*, *in light of the(11)*	4	
	Comparative expressions *as soon as possible(17)*	1	
Total		34	100%

* A VP-based-look-like lexical bundle without a subject is classified into Dependent clause type because it precedes a dependent clause, marked by the final words 'to', 'what', 'if', etc., according to Biber *et al*. (2004)

Based on this structural categorization of lexical bundles, it is observed that the overall distribution pattern shown above is similar to that found in Biber *et al.*'s (2004) study of conversational register in that verbal elements (VP-based and Dependent clause bundles) predominate the lists. It is noteworthy that a number of features, such as frequent use of personal pronouns (i.e. $1^{st}/2^{nd}$ person pronoun + VP / + dependent clause fragments), dense use of verbal elements (i.e. VP-based and Dependent clause bundles), etc., identified in lexical bundles of ENRON's emails correspond to high incidence of those features in spoken registers as found in previous studies (e.g. Baron, 1998; Gimenez, 2000; Danet, 2002).

However, it is not entirely conversational as 26.5% of the corpus is made up of NP/PP-based bundles, the same type that was found in Biber *et al.*'s (1999) study of academic writing. Based on this observation of ENRON, it can be said that email is a business written genre that is very close to conversation. This general tendency will be seen more clearly in the discussion of each basic structural type below.

4.1.1 VP-based Bundles

VP-based bundles comprise 16 structural types out of 34 in ENRON. As shown in Table 4.1, these VP-based bundles were further divided into seven structural subcategories; however, only four of them can be identified in ENRON, which are:

(1) (connector +) $1^{st}/2^{nd}$ person pronoun + VP fragment

(2) (connector +) 3rd person pronoun/ noun + VP fragment

(3) Verb phrase with active verb

(4) Wh-question fragments

Based on the density, the two biggest types are (3) and (1), to be discussed in detail below. As for groups (2) and (4) of the VP-based bundles, they neither have high frequencies nor many types, resulting in insufficient evidence to describe patterns. Therefore, they are not discussed in detail.

The first largest type of VP-based bundles (3) contains only active verbs without a subject, i.e. *feel free to contact, get in touch with, is a draft of, look forward to hearing, have any questions or, thank you for your,* and *please let me know*. It can be observed that these lexical bundles share not only their form but also their meaning. That is, they are related to 'communicative acts' such as *feel free to contact, get in touch with, look forward to hearing*, etc. Also, one lexical bundle – *thank you*

for your - is considered as a politeness type. In other words, this group of lexical bundles highlights communicative functions and interactions.

In the second biggest group of VP-based bundles, "(connector +) 1st /2nd person pronoun + VP" contains *I*, *we* and *you*. The use of first and second person pronouns reflects "high personal involvement" (Danet, 2002, p.9) and helps "produce the warm and friendly tone" (Gartside, 1974 cited in Mallon & Oppenheim, 2002, p. 9). Apart from that, two lexical bundles, *I don't know* and *I don't think*, are considered hedging devices (O'keeffe, McCarthy, & Carter, 2007, pp. 73-74). They mostly illustrate the uncertainty of the writer towards the complement clause, as shown in examples below. It should be noted that these forms also correspond with their functions as will be discussed in detail in 4.3.1.

1) given the other demands on his schedule, but ***I don't know*** that for sure. (Enron_173245)
2) I think they should try to get in to see Lundquist, but ***I don't think*** we need to participate in this. (Enron_175359)

Furthermore, three out of six lexical bundles in this structural subcategory are in contracted form, which are *I don't have, I don't know,* and *I don't think*. Contractions are found popular in both personal business and non-business emails (Mallon & Oppenheim, 2002, p. 17). The use of contractions expresses speech-like linguistic element of business email (Danet, 2002). This can thus be inferred that business emails in ENRON are rather personal and close to spoken registers.

4.1.2 Dependent Clause Bundles

This structural category accounts for 26% of all lexical bundles in ENRON, with 9 lexical bundle types. Four out of five structural types identified in Dependent clause bundles are:

(1) (connector +) 1st/2nd person pronoun + dependent clause fragment (e.g. *we would like to*)

(2) Wh-clause fragments (*let me know what*)

(3) If-clause fragments (e.g. *if you have any*)

(4) (Verb/Adjective+) to-clause fragment (e.g. *don't want to*)

The biggest two categories are (1) and (3), with three lexical bundles in each subcategory. Because these two groups are the largest, they are discussed in detail below.

In the group of "(connector +) 1st /2nd person pronoun + dependent clause fragment", every bundle consists of *would like to* either in a full or contracted form (i.e. *I'd like to*). Interestingly, all three of these lexical bundles share the same semantic properties – desire expressions. This will be elaborated more in 4.3.1.

In the case of the "If-clause fragments" group, there are *if you have any, if you want to,* and *let me know if*. A look at the fragments suggests that these 'if' bundles can be used as one form of requests. This form of request is found in "horizontal and external email exchanges" (Danielewicz-Betz, 2016, p.213). In other words, the colleague talk (horizontal email) or outsider contact (external email) tended to employ conditional sentences to request an email reader to perform certain action as shown in the following sample sentences.

1) Let either of us know ***if you have any*** additional questions. Regards, Eric (ENRON_47986)
2) Not top-shelf enough in my view. ***If you want to*** discuss further let me know. (ENRON_174461)

According to the categorization provided in the Berkeley's Project, the first sample is part of correspondence on internal projects, which can be grouped into the "horizontal" email. Meanwhile, the second sample is in line with the former since the message is part of communication among co-workers with a friendly tone.

It is noticeable that lexical bundles in the "Dependent clause" type is very interactive, engaging the reader via the use of personal pronoun 'you', or verbs denoting feelings like 'want', 'feel', etc. Most important of all, seven out of nine lexical bundle types share the same verbs 'like', 'know', and 'want'.

4.1.3 NP/PP-based Bundles

This category also contains nine types of lexical bundles like 4.1.2, equaling 26.5% of all the lexical bundle types. There are four out of five subcategories classified in this group:

 (1) PP-based bundles, i.e. *in light of the, as a result of, in the process of,* and *to hearing from you*

(2) NP-based bundles with *of*-phrase fragment, i.e. *a copy of the, a draft of the,* and *the end of the*

(3) Other NP-based bundles, i.e. *thanks for your message*

(4) Comparative expressions, i.e. *as soon as possible*

While the previous two aforementioned basic structural types are central to verbal elements and suggest the tendency towards the spoken register, these NP/PP-based bundles show that business emails still contain features often associated with the written register. According to Biber *et al.* (1999), the prepositional phrase expressions containing 'of' are the largest proportional type in academic prose. There are three ENRON's PP-based bundles with 'of', which are *in light of the, as a result of,* and *in the process of*. Besides sharing the same form, they also have the same function, e.g. intangible framing attributes, which will be discussed further in 4.3.3.

There are four NP-based bundles out of nine lexical bundles in this structural type. It must be noted that NP-based bundles are also the second frequently used type in academic writing (Biber et al, 1999), following the prepositional phrase with 'of' demonstrated above. Two of the NP-based bundles concern document, which is part of business activities. They are *a copy of the* and *a draft of the*. In addition, the word 'thanks' in four-word bundle *thanks for your message* is also considered colloquial, which is regarded as speech-like feature (Danet, 2002, p.21).

The "NP/PP-based bundles" category shows an element of business email, which is associated with the written register, thereby preventing business emails from resembling conversational discourse. Furthermore, they also shows a high correlation between form and function.

4.2 Classification of Structural Types in TEC

The structural analysis of lexical bundles was conducted on sample business emails in TEC. As shown in Table 3.1 in section 3.3, the same minimum occurrence of 1% of the whole texts was applied to TEC as well, hence in at least 8 texts for the size of 751 sample business emails. Based on all criteria mentioned in Chapter 3, a total of 53 lexical bundles were obtained from this corpus.

Table 4.2 below shows the distribution pattern of three main structural categories with subcategories of lexical bundles found in TEC. Similar to ENRON, VP-based bundles in TEC outnumber the other two types, covering 83% of the lexical bundles in total whereas NP/PP-based bundles are found only 17%.

Table 4.2 Structural types and proportional distribution in TEC

Basic structural type	Structural Categories	No. of types	Percentage (No.)
VP-based	**(connector +) 1ˢᵗ/2ⁿᵈ person pronoun + VP fragment** *I look forward to(68), we look forward to(24), you have any questions(19), I would be grateful(18), you let me know(10), we would be grateful(9), I am interested in(8), I have attached a(8), I am attaching a(8), you could send me(8)*	10	49% (26)
	(connector +) 3ʳᵈ person pronoun/ noun + VP fragment *it was good to(14)*	1	
	Discourse marker + VP fragment (e.g. *I mean you know, you know it was*)	0	
	Verb phrase with active verb *thank you for your(91), look forward to hearing(46), please let me know(32), look forward to seeing(24), hesitate to contact me(13), not hesitate to contact(12), get back to me(11), will be able to(11), hearing from you soon(10), would like to know(10), give me a call(9), let you know that(9), look forward to receiving(9),*feel free to contact me(8)*	14	
	Verb phrase with passive verb (e.g. *is based on the, can be used to*)	0	
	Yes-no question fragments *please could you send(8)*	1	
	Wh-clause fragment (e.g. *what does that mean, how many of you*)	0	
Dependent clause	**(connector +) 1ˢᵗ/2ⁿᵈ person pronoun + dependent clause fragment** *I would like to(38), I am writing to(24), we would like to(21), I'd like to(9), you would like to(9), I'm writing to(8)*	6	34% (18)

	Wh-clause fragments *let me know what(17)*	1	
	If-clause fragments *let me know if(40), if you have any(29), would be grateful if(24), if you could send(13), if you would like(13)*	6	
	(Verb/Adjective+) to-clause fragment **please do not hesitate to(16), to let you know(14), to tell you that(13), to inform you that(12), to hear from you(9)*	5	
	That-clause fragments (e.g. *that there is a, that I want to*)	0	
NP/PP-based	**Noun phrase with *of*-phrase fragment** *the end of the(14), end of the week(10), a copy of the(8)*	3	17% (9)
	Noun phrase with other post-modifier fragment *thanks for your email(11)*	1	
	Other noun phrase expressions *dear sir or madam(22)*	1	
	Prepositional phrase expressions *by the end of(20), for your e-mail(12), in the near future(8)*	3	
	Comparative expressions *as soon as possible(30)*	1	
Total		53	100%

As can be seen in the above table, 83% of total lexical bundles in TEC consist of verbal elements, which are VP-based bundles and dependent clauses, while NP/PP-based bundles own 17%. These groups will be discussed qualitatively in turn below.

4.2.1 VP-based Bundles

Like ENRON, the largest structural type in TEC is VP-based bundles, covering 26 out of 53 lexical bundles in TEC. These bundles are further categorized into four subtypes:

(1) (connector +) $1^{st}/2^{nd}$ person pronoun + VP fragment

(2) (connector +) 3rd person pronoun/ noun + VP fragment

(3) Verb phrase with active verb

(4) Yes-no question fragments

Similar to ENRON, the two most found types are (3) and (1), respectively.

In group (3), many lexical bundles overlap with one another. For example, the three-word chunk *look forward to* is found in three four-word lexical bundles, which are *look forward to receiving, look forward to seeing,* and *look forward to hearing*. The verb *contact* is also repeatedly used in different lexical bundles, including **feel free to contact me, not hesitate to contact,* and *hesitate to contact me*. Therefore, it can be inferred that business English textbooks offer a great variability of lexical bundles surrounding the verbal phrases "contact" and "look forward to". This is probably due to the pedagogic purposes of textbooks. At the same time, this variety might be a result of the fact that corpus of the present study contains 77 textbook samples. This might be a factor that increases a variety of lexical bundles with similar meaning in different forms.

The second biggest VP-based bundle type is "$1^{st}/2^{nd}$ person pronoun with VP fragment". Similar to ENRON, all first and second person pronouns are included, "reflecting high personal involvement" (Danet, 2002, p. 9). In addition, the verbs used in this group tend to be action one. The action verbs as 'attach' and 'send' require concrete noun objects in the corpus' contexts. This corresponds to Nelson (2000), who examined business English textbooks and found that vocabulary in teaching materials tend to focus on concrete items rather than states or qualities. The following instances show the context of each lexical bundle with action verbs embedded.

1) We offer a 5% discount for orders made through our website. ***I am attaching** a* copy of our price list. (TEC_U11_02)
2) With reference to your enquiry ***I have attached a*** copy of our brochure. (TEC_ T25_41)
3) You have our declaration form No. 11765916. I would be grateful if ***you could send me*** a claim form. (TEC_ T07_36)

Besides, some differences between ENRON and TEC are observed regarding this structural subtype. First of all, no contraction was identified in this group of lexical bundles. While ENRON has three lexical bundles in contracted form, TEC does not have any contraction under this subcategory despite being larger in number. In addition, semantic properties of the lexical bundles are more varied in TEC than in ENRON. There are various groups of meanings, not only communication and politeness but also documentation such as 'attach' and 'send'.

4.2.2 Dependent Clause Bundles

Dependent clause bundles account for 34% of the structural types in TEC, with four subcategories namely:

(1) (connector +) 1st/2nd person pronoun + dependent clause fragment (e.g. *I'm writing to*)

(2) Wh-clause fragments (*let me know what*)

(3) If-clause fragments (e.g. *if you need any*)

(4) (Verb/Adjective+) to-clause fragment (e.g. *to hear from you*)

The majority of dependent clause bundles are (1), (3), and (4).

The first two biggest group of dependent clause bundles come from two subtypes: "(connector +) 1st/2nd person pronoun + dependent clause fragment" and "If-clause fragments". Each subtype contains six lexical bundles, two times bigger than ENRON. Similar to ENRON, contractions can be found only in "(connector +) 1st/2nd person pronoun + dependent clause fragment" (i.e. *I'm writing to* and *I'd like to*). It can then be observed that business English textbooks do not tend to expose contractions to learners; however, authentic emails tend to use this form as illustrated in this research project.

Furthermore, lexical bundles in "If-clause fragments" namely *if you have any, if you want to,* and *let me know if* can also be used as a request like the corresponding bundles found in ENRON. For example, the *if you could send* bundle requests the email recipient to send the writer what follows the lexical bundle as in:

1) I would be very grateful ***if you could send*** details of your programmes including times and prices. (TEC_ T08_02)
2) In the meantime, I would be grateful ***if you could send*** us an up-to-date list of your fees. (TEC_T58_04)

The third largest group is "(Verb/Adjective+) to-clause fragment". Upon an examination of the concordance lines, contexts of the lexical bundles in this group, it is found that this subcategory is used to introduce a topic at the beginning of email message. Sample lexical bundles are *to inform you that*, *to tell you that*, and *to let you know*. They are part of the topic introduction lexical bundles, which will be discussed in more details in 4.4.2. The following instances illustrate this pattern.

1) I am writing *to inform you that* a number of pieces of crockery were damaged in a recent shipment to MacKenzie Bros of Dawson, Canada. (TEC_ T07_36)

2) Dear colleagues   I'm pleased *to tell you that* our financial results this year have been very positive. (TEC_T04_07)

3) Makki, I just wanted *to let you know* that I got your order today. (TEC_ T09_18)

4.2.3 NP/PP-based Bundles

This structural type is the least found in TEC, covering merely 17% of lexical bundles in this corpus. Interestingly, upon an investigation of the context, five out of eight NP/PP-based lexical bundles in TEC involve the temporal expressions. This can be seen in *end of the week, the end of the, by the end of, in the near future,* and *as soon as possible*. It can be further divided into two main groups: "deadline" and "unspecific time" reference. The bundles that denote "deadline" usually include the word 'end' to indicate the deadlines of the required action. Sample instances are no. 1), 2) and 3). On the other hand, the "unspecific time reference" includes bundles used with different degrees of intensity as the sample no. 4) did not specify when the email users will contact the reader while no. 5) the lexical bundle itself emphasizes that the action should be accomplished soon. Again, it should be noted that their functions, "Time reference", also correlate with their forms, which will be discussed further in 4.4.3.

1) Could you get back to me by the *end of the week*? Thanks (TEC_ T51_01)

2) I'm afraid I've even had to put off meeting the CEO until *the end of the* month. (TEC_ T27_16)

3) I'll send them *by the end of* the week. (TEC_T04_06)

4) We will contact you again *in the near future*. (TEC_ T10_02)

5) Let them know that we will pay invoice #721-23 *as soon as possible*. (TEC_ T18_16)

4.3 Classification of Functional Types in ENRON

This section addresses the second research question of the study which deals with functional types of lexical bundles found in authentic business emails. It must be noted here that the functional analysis requires a study of an expanded context and

that some lexical bundles were found to have more than one primary function, depending on their contexts regarding the immediate text and the position of that bundle in the body of the email message. However, in order to calculate and compare lexical bundles from each corpus systematically, I labelled a lexical bundle with only one function using the frequencies as a criterion. That is, following the corpus linguistic perspectives, the highest frequency was interpreted as the dominant function of a lexical bundle.

Following Biber *et al.* (2004) and Biber (2006), the lexical bundles in ENRON were categorized into four main types: (1) Stance bundles, (2) Discourse organizers, (3) Referential expressions, and (4) Special functions (See 2.3.2.2). In each group, there are subcategories of lexical bundles, each of which serves different functions. However, upon data analysis some lexical bundles were found not to fit in the pre-existing categories suggested in Biber *et al*. (2004) and Biber (2006). This is probably because the previous work does not look at email discourse. Therefore, I created new categories in "Special functions" so that they capture all lexical bundles that perform functions determined by contexts of their occurrences. The newly created subcategories are *Request*, *Opening up for further communication*, *Expectation*, and *Hybrid*. Given that these categories were generated through the data under the present study, they can be considered particularly functionally specific to email discourse. The last subtype, *Hybrid*, refers to a combination of two functional types performed at the same time.

Table 4.3 displays the functional categories of 34 four-word lexical bundle types in ENRON. "Special functions" is the most frequently used functional type, with 14 lexical bundles (41%). The second and third functional types are "Stance expressions", with ten lexical bundles (29.5%), and "Referential expressions", with eight ones (23.5%). The least found functional type in ENRON is "Discourse organizers", in which only two lexical bundles (6%) were identified.

Table 4.3 Functional types of lexical bundles in ENRON

Basic Functional Type	Structural Categories	No. of types	Percentage (No.)
STANCE EXPRESSIONS	**A. Epistemic stance** *I don't think, I don't know*	2	29.5% (10)
	B1 Desire *I would like to, I'd like to, we would like to, don't want to, if you want to*	5	
	B2 Obligation/directive *let me know what*	1	
	B3 Intention/prediction *it would be a*	1	
	B4 Ability/effort *I don't have*	1	
DISCOURSE ORGANIZERS	**A. Topic introduction/focus** *attached is a draft*	1	6% (2)
	B. Topic elaboration/clarification (e.g. *I mean you know, as well as the*)	0	
	C. Conditions *if you have any*	1	
REFERENTIAL EXPRESSIONS	**A. Identification/focus** (e.g. *and this is a, one of the things*)	0	23.5% (8)
	B. Imprecision (e.g. *or something like that, and things like that*)	0	
	C1 Quantity specification (e.g. *have a lot of, the rest of the*)	0	
	C2 Tangible framing attributes *a draft of the, a copy of the, is a draft of*	3	
	C3 Intangible framing attributes *as a result of, in the process of, in light of the*	3	
	D1 Place reference (e.g. *the United States and, of the United States*)	0	
	D2 Time reference *as soon as possible*	1	
	D3 Text reference (e.g. *shown in figure N, as shown in figure*)	0	
	D4 Multi-functional reference *the end of the*	1	
SPECIAL FUNCTIONS	**A. Politeness** *thank you for your, thanks for your message*	2	41% (14)
	B. Inquiries *what do you think*	1	

	C. Request *please let me know, get in touch with*	2	
	D. Opening up for further communication *you have any questions, please feel free to, have any questions or, feel free to contact*	4	
	E. Expectation *I look forward to, to hearing from you, look forward to hearing, we look forward to*	4	
	F. Hybrid *let me know if*	1	
Total		34	100%

4.3.1 Stance Expressions in ENRON

Stance bundles cover 10 out of 34 of the lexical bundles in ENRON, equaling 29.5%. Two main subcategories of *Stance bundles* are *Epistemic* and *Attitudinal/modality* bundles. They are used to show the email's stance or expressions. The functional analysis of ENRON shows that writers of the authentic business emails tend to express uncertainty based on *Epistemic stance* bundles. In addition, the *Desire* function covers 50% of all *Stance bundles*. Based on these two distinctive features of *Stance expressions* in ENRON, it is inferred that lexical bundles in authentic emails tend to be used to express uncertainty and personal desire.

In the *Epistemic stance* subcategory, two lexical bundles were *I don't think* and *I don't know*. As mentioned in 4.1.1, both of them show uncertainty towards the proposition following these bundles. They were mostly found in the "company business" subcategory of the corpus, followed by "logistic arrangements"[3], "employment arrangement" (only *I don't know*), "document editing/checking" (only *I don't think*), and "informal business" (only *I don't know*) subtypes. This illustrates that the *Epistemic stance* function can be found in various types of emails. As for the bundle *I don't think*, 17 out of 18 its occurrences were followed by a clause in present tense and the other by present future. This indicates that the writer is not sure that

---

[3] The "logistic arrangements" subcategory concerns with business emails dealing with planning, confirming a meeting, etc.

things are going to be what they are expected to be. Sample sentences are shown below.

　　　1) ***I don't think*** it helps us if we get lucky, get direct access back, then make a decision to return customers to the utility. (ENRON_121752)
　　　2) In light of recent changes in the California regulatory environment (eg the recent rate order) ***I don't think*** it makes sense for us to distribute the document at this time, so you can ignore the request. (ENRON_175580)

As can be seen above, the writers express their lack of certainty towards the issues under discussion. Sentence in 1) concerns an internal policy about rejecting the customers. A series of emails were sent back and forth on this issue. The writer in 1) said that he did not agree with the idea mentioned in the email sent earlier by using the lexical bundle *I don't think*. Meanwhile, the email of 2) involves making a decision ahead of time. The writer used this lexical bundle to express his disagreement in distributing the document.

　　　The second *Epistemic stance* bundle is *I don't know*. Similar to *I don't think*, this bundle expresses the writer's uncertainty on the condition that it precedes a complementizer either 'if' or 'whether' as shown in examples below.

　　　1) ***I don't know*** if Bud was in the field and that acknowedgment may not get us far.   (ENRON_234313)
　　　2) ***I don't know*** whether that's true, but the bottom line is that FERC has the information in its files and could publish it any time at wants (ENRON_253815)

In addition, it is found that *I don't know* can also show the lack of acknowledgement of the writer if it precedes a nominal phrase with *that* or a Wh-word at the initial position.

　　　1) ***I don't know*** that NERC has "given up" on Enron. (ENRON_136543)
　　　2) I apologize that ***I don't know*** which one of you is handling this issue (ENRON_173357)

　　　Next is the category of *Attitudinal/modality stance bundles*. The largest type in terms of number is *Desire*. The distinctive feature of this subtype is verb, which connotes the desire of either the writer or the reader (Biber, 2006). The lexical bundle

that takes the first place of *Desire bundle* is *I would like to*, followed by a contracted form, *I'd like to*. It should be noted that the full form, hitting 41 times, is frequently used than the contraction, hitting 15 times, nearly two times in number. Moreover, they can also perform the *Topic introduction/focus* function when they were found at the beginning of the email message as illustrated in 1) below. Both lexical bundles were mostly found in the "company business" subcorpus. The full form was present in all subcorpora of ENRON but the contracted form of *I'd like to* was not found in the "document editing/checking" subcorpus. This shows that the email users did not distinguish between the full and contraction forms of the lexical bundle since both forms occurred in various types of emails.

1) Re: Greenpeace letters to Enron Europe concerning our position on Kyoto All: **I would like to** discuss how we might respond to this letter from (ENRON_175419)
2) After you've had a chance to review the above, **I'd like to** get together and discuss this with you. (ENRON_255333)

Full and contraction forms are discussed in relation to formality in the business correspondence as mentioned in 4.1.1. To see whether the contraction form truly correlates with informal writing, I employed Danet's (2002) criterion used to examine the degree of formality in business emails through the different types of greetings, including 'Dear', 'Hi', and the lack of greeting expressions. After examining all 15 hits of the *I'd like to*, it is found that eight of them, equaling 53%, were classified as an informal writing style, which is either without greetings or with the use of colloquial expressions (e.g. (Hi +) First Name). This is partly due to the fact that 66% of *I'd like to* was identified in internal emails (e.g. in-company), which might require a less formal writing style compared to external (out-company) ones. Still, other factors such as organizational culture, personal relationship, etc. should be taken into account as suggested by Gimenez (2005).

Another bundle containing the phrase 'would like to' is *we would like to*. According to McCarthy and Handford (2004) who examined spoken business English, the pronoun 'we' can carry "a wide range of references, from very broad corporate reference to immediate group reference and to the individual using it to shelter behind corporate authority or responsibility or to protect interlocutor's face"

(p. 178). The bundle was found in three subcorpora: "company business", "logistic arrangement", and "document editing/checking". It is noticeable that these types of email usually involve a group of people, i.e. working with colleagues, arranging a meeting, editing and revising documents. To see whether the type of email has effects on the reference of 'we', an investigation of concordance lines of the bundle *we would like to* was conducted. It is found that there are four types of reference:

(1) literal meaning, which refers to more than one email writer identified in the email.

(2) corporate reference, which focuses on organizational status.

(3) group of colleague, which indicates a group of people working together.

(4) individual reference, which means a single person using the pronoun 'we'.

Sample sentences of each reference type are presented corresponding to the number of each type above.

1) propose to submit to the Judge and his legal staff tomorrow. While *we would like to* convince the Judge that the June 19 methodology is not (ENRON_124868)

2) because it holds a current power marketing certificate is an entity which *we would like to* move to being a direct subsidiary of Enron Corp. (ENRON_162662)

3) presentations. We are still planning trips to California and New Mexico that *we would like to* combine and would like very much to meet and (ENRON_54928)

4) Subject: HIGH PRIORITY - PR's PRC Meeting It's that time again... *We would like to* schedule PR's PRC meeting prior to Steve Kean' (ENRON_175307)

In no. 1), the use of 'we' is literal meaning since the email closes with two names of the writers. This function appeared only once in the "company business" email. At the same time, both no. 2), which focuses on corporate, and no. 3), which indicates a group of colleague reference, have been used in "business company", "logistic arrangement", and "document editing/checking" types of email. The last function represented in no. 4) referring to an individual was identified only in "logistic arrangement" type. This is because the email concerns arranging a meeting schedule. In no. 4), although the writer started off using the plural form 'we', the body of the message indicates that it is the writer who organized the meeting. Therefore,

she opted for the pronoun 'we' in order to soften the tone by not making it too individual.

Furthermore, apart from being a *Desire* bundle, the conditional lexical bundle *if you want to* can serve a hypothetical bundle as shown in the example below. It is not meant to elicit the desire of the reader but only hypothesize a situation, which may or may not be desirable. The following excerpt is an email discussing different types of benefits depending on the policy on stock options.

1) I am concerned that there may be some odd incentives in the structure -- eg are severed employees allowed more time to exercise than employees who just leave voluntarily (ie *if you want to* have a full three years to exercise options are you better off becoming a performance issue and getting severed than you would be if you simply left the company?

In this excerpt, the writer suggested the email recipient get 'severed' in order to exercise "a full three years" option rather than "simply left the company". The verb 'want' embedded in the four-word lexical bundle starting with the conditional 'if' does not emphasize the reader's own desire but conveys a hypothetical meaning. Such meaning is generated by the conditional 'if' itself as textual evidence.

The bundle *let me know what* is assigned to the *Obligation/directive* function as shown in 1) since most of its occurrence indicate; however, on some occasions, it can be served as the *Offer* function to offer help to the email reader as displayed in 2).

1) See draft release below. I have made some suggestions. *Let me know what* you think by Friday at 10:00am. Peggy (ENRON_174187)
2) Subject: Re: 2001 budget I do support the change. *Let me know what* I can do to help. (ENRON_174244.txt)

A single bundle *it would be a* is classified into the *Intention/prediction* function. Similar to the bundle *if you want to*, the meaning that comes with the modal verb 'would', which is embedded within the lexical bundle, is 'hypothetical meaning' (Leech, 1989), expressing tentativeness. The following sample sentences display how this bundle was used to make suggestions as in no. 1) and 2) in authentic business emails in ENRON.

1) We thought *it would be a* good idea to add this topic to the brand conference (perhaps as part of an existing segment).(ENRON_173931)

2) Subject: Legal Analysis on AB 1890 I think *it would be a* good idea to have Ron Carroll in Watkiss' shop (ENRON_174503)

The last subcategory of *Stance expressions* is *Ability/effort*. It must be noted that the property of possession is considered part of this category for the present study since the original meaning as 'ability/effort' was not found in the present study. As a result, the bundle *I don't have* is classified into this subtype since 75% of its use shows that this bundle is followed by a nominal phrase, denoting possession of something. The example is given in no. 1) below. Nevertheless, if this bundle precedes the word 'to', forming the modal verb 'have to', it would perform the *Obligation/directive* function due to the fact that such modal verb obliges the writer to do something, which is used in a negative form in this case. The sample of the latter function is provided in no. 2).

1) As a significant contributor to the PAC and Enron's other adventures, *I don't have* the room.   Personally, I also have difficulty making campaign (ENRON_191480)

2) I can also fly Continental through Rome or London, as long as *I don't have* to change airports. Vince (ENRON_54576)

4.3.2 Discourse Organizers in ENRON

Discourse Organizing bundles are the least frequently used functional type in ENRON, with only two lexical bundles, amounting to 6% of the lexical bundles in ENRON. One was categorized into the *Topic introduction/focus* function and the other into the *Conditions* function.

In the *Topic introduction/focus* subtype, the lexical bundle *attached is a draft* is always found at the beginning of the email message as exemplified below. The function of this word is to indicate the reader from the very beginning that a draft of something is attached to the email. Therefore, it is not surprising to find that 67% of such bundle was found in the "document editing/checking" subcorpus since this email type is centered around documentation. In addition, this bundle characterizes the email discourse since the word 'attached' is specifically used in email to refer to

attachment while the term 'enclosed' is commonly used in letters ("Business letter format How to write a business letter," 2012).

       1) Subject: Letter to residential customers ***Attached is a draft*** of the letter we'd like to send to (ENRON_175169)
       2) Subject: Letter for Lay Signature John/Steve, ***Attached is a draft*** letter to Party Secretary Zhou Yong Kang of Sichuan (ENRON_174392)

       The second subcategory of *Discourse organizing bundles* is *Conditions*. The identified bundle in this function is *if you have any*. Although the person pronoun 'you' is embedded within this bundle, it is followed by words used to introduce various communicative acts such as 'questions', 'comments', etc., illustrated in examples below. In addition, 93 % of this bundle was found at the end of the email message before closing. Consequently, it serves to introduce statements of opening up for further communications politely.

       1) slides in advance to the video teleconference locations. Please let me know ***if you have any*** questions regarding this report. Thanks, Courtney (ENRON_175704)
       2) ***If you have any*** comments or questions, please let me know by 10 a.m. Thursday. Thank you. (ENRON_ 173364)
       3) We will call you at (605) 497-4045 unless otherwise instructed. ***If you have any*** questions, please feel free to contact me at 713/853-5290. Best (ENRON_52993)

4.3.3 Referential Expressions in ENRON

Referential expressions contain eight lexical bundle types out of 34 in total, amounting to 23.5% of the lexical bundles in ENRON. They were further categorized into four subtypes: *Tangible* and *Intangible framing attributes*, *Time,* and *Multi-functional references*. The two largest types are *Tangible* and *Intangible framing attributes*, each of which has three lexical bundles. The other two subtypes have only one lexical bundle in each of them.

       With regards to *Tangible framing attributes*, all three of them, i.e. *a draft of the, a copy of the,* and *is a draft of*, are associated with documentation. It is observed that the three-word bundle 'a draft of' is embedded within two four-word bundles. Also, these three *Tangible framing attributes* lexical bundles share the same frame "*a + Noun + of the/a*", which illustrates the way in which the form "*a + Noun + of the/a*"

is tied to the function of referring to tangible framing attributes. Samples of context in which they are found are put below.

       1) Attached please find **_a draft of the_** financial schedules prepared by Enron Corp. (ENRON_173938)
       2) Kevin 213-926-2626   PS - May I have **_a copy of the_** position paper that was circulated in the Beverly (ENRON_7932)
       3) Steve This **_is a draft of_** the contact list I promised you. (ENRON_176651)

       The *Intangible framing attributes* function includes three lexical bundles – *as a result of***,** *in the process of,* and *in light of the*. The lexical bundle *as a result of* in fact gives the cause of the proposition preceding or following the bundle as exemplified in 1) and 2). It is put before nominal phrases "your making inquiries on our behalf" and "the all the new hardware and software I purchased". It should be noted that this bundle, *as a result of*, was also identified in Biber's (2006) investigation on textbooks in business field. It might thus be inferred that the lexical bundle in authentic email samples in the present study has a common expression found in the business field. Meanwhile, the bundles *in the process of* and *in light of the* precede a gerund describing an action and a noun, respectively, as shown in 3) and 4), respectively.

       1) work adheres to this principle. We regret any difficulties that have arisen **_as a result of_** your making inquiries on our behalf. (ENRON_174320)
       2) **_As a result of_** the all the new hardware and software I purchased, my office looks like a war zone with an odd mix of PUC service copies, computer documentation, and diskettes laying all around. (ENRON_255511)
       3) Subject: Market Stack   Vince,   Henwood is **_in the process of_** developing a new product that, for lack of a (ENRON_54559)
       4) times about the relative importance of OPIC, ExIm and other funding organizations **_in light of the_** change in emphasis in our business. As a result (ENRON_175772)

       The last two subcategories of *Referential expressions* are *Time* and *Multi-functional references*. The former contains the bundle *as soon as possible*. This bundle can be considered making a request since the bundle implies an urgent request by intensifying the degree of 'soon'. On the other hand, *Multi-functional references* include *the end of the* that can serve either *Time* or *Text references*, as shown in 2) and

3), respectively. In fact, Biber (2006) also classified this bundle under the same subcategory within the business textbook register.

1) What I need from you, **as soon as possible**, is a timeline (what the Administration said when -- doesn't have to be down to the day) (ENRON_176661)

2) yesterday would require this to be done on the Internet after **the end of the** trading day. (ENRON_62815)

3) With respect to the alternative responses shown near **the end of the** document, leave both alternatives in. (ENRON_173885)

4.3.4 Special Functions in ENRON

The majority of all functional types in ENRON lies in *Special functions*, with 41% or 14 lexical bundle types. Most of them perform pragmatic functions. They are further divided into six subtypes:

(1) Politeness, i.e. *thank you for your* and *thanks for your message*

(2) Inquiries, i.e. *what do you think*

(3) Request, i.e. *please let me know* and *get in touch with*

(4) Opening up for further communication, i.e. *you have any questions, please feel free to, have any questions or, feel free to contact*

(5) Expectation, i.e. *I look forward to, to hearing from you, look forward to hearing, we look forward to*

(6) Hybrid, i.e. *let me know if*

As can be seen, the two largest functional subtypes based on the number of lexical bundle type are (4) and (5), with four lexical bundles in each of them. The rest contain less than two types of lexical bundles for each function.

Opening up for further communication includes mainly lexical bundles that are found towards the end of the email message. This is influenced by the fact that this function serves as an offer for further communication which usually comes after the main messages have already been delivered. There are two pairs of overlapping bundles. The first pair is *you have any questions* and *have any questions or* and the second is *please feel free to* sharing three-word bundle with *feel free to contact*. Nevertheless, all four bundles contain words relevant to communicative acts, e.g. 'questions' and 'contact'.

Although the bundle *please feel free to* contains the word 'please', this bundle does not perform the *Request* function by nature. In fact, the whole bundle connotes encouragement or politeness pragmatically by inviting the reader to contact the person specified in the context. A sample sentence is given below.

1) If you have any questions*, please feel free to* contact Terry Bosien (x35230). Thank you again for our contributions to this important process. (ENRON_176687)

Similar to the previous subcategory, the four *Expectation* lexical bundles overlap with each other. The first group involves *I look forward to, look forward to hearing,* and *we look forward to*, which share the same three-word bundle *look forward to*. Next is the pair of *to hearing from you* and *look forward to hearing*, in which two-word bundle *to hearing* is embedded within. The main purpose of this function is to express the writer's expectation towards the email recipient. This type is usually found at the end of the email message as shown below.

1) I would like to help you succeed with this challenge. *I look forward to* our discussions.   Kevin 213-926-2626   (ENRON_7972)
2) ENA to explore the possibilities for the city of San Diego? I *look forward to hearing* from you. Best regards, Michael (ENRON_174298)
3) or fax at (905) 680-1568 or on the net at dlpits@yahoo.com. *We look forward to* hearing from you. Thank-you, Dave/Lorna Clark Allied Motorsports (ENRON_175204)

Next, the *Politeness* function has the bundles *thank you for your* and *thanks for your message*. In terms of position, the former can be found in any part of the message body, from the beginning to the end, while the latter is repeatedly found at the beginning of the email message. Moreover, the lexical bundle *thanks for your message* appears like a colloquial expression, supported by the frequent use of first name of the email recipient preceding this bundle. It might be said that this bundle serves the *Topic introduction/focus* function as it refers to the previous message sent by the email recipient earlier before introducing a new topic, plus it initiates the body message. Such tendencies are displayed in the following sample sentences.

1) to fine-tune my thinking / preparation for Thursday's meeting with Janet?
Thank you for your all of your help. I am pleased and appreciative (ENRON_9066)

2) Subject: RE: Frank, ***Thanks for your message***. I shall forward your comments to Greg. I agree (ENRON_162855)

The *Request* function also contains two lexical bundle types: *please let me know* and *get in touch with*. As for the bundle *please let me know*, the word 'please' is supposed to depict a sense of soft tone of the phrase. However, some of them turn out to reinforce the request. This pragmatic use of the word 'please' as an emphasizer is in line with Danielewicz-Betz's work (2016), which suggests that the word 'please' might not soften the tone of the message but strengthen it instead. This is illustrated in sample 1), in which the prepositional phrase "by 10 a.m." follows the bundle to specify the deadline of such action preceding the bundle. Meanwhile, the bundle *get in touch with* is categorized into *Request* since its extended view shows the word 'please' was repeatedly used along with this lexical bundle as shown in 2).

1) If you have any comments or questions, ***please let me know*** by 10 a.m. Thursday. Thank you. kd (ENRON_173364)

2) John (Woody) Wodraska, managing director of Azurix is interested in attending. Please ***get in touch with*** him. I haven't heard from Stan yet.   Susan (ENRON_173992)

The *Inquires* function is one of subcategories in the framework of Biber (2006). It concerns a question as suggested by the label. In ENRON, the lexical bundle classified into this functional type is *what do you think*. Based on the concordance lines examination, this bundle can be used in three ways:

(1) Standing alone (i.e. *what do you think*?)

(2) Preceding a prepositional clause. (e.g. *of* and *about*)

(3) Preceding a nominal clause (e.g. a covert complementizer 'that')

These patterns are arranged according to their frequencies, from the highest down to the lowest, which is (1), (2), and (3). As for (1), six out of 20 hits or 30% of this lexical bundle were used to refer to the embedded emails (forwarded emails) without any other notes accompanying the lexical bundle as displayed in 1) below. This highlights one characteristic of email discourse that allows the email users to

write as if they are talking by embedding the message they want to ask for the opinion from the email recipient. According to Biber's (2006) findings, this lexical bundle was identified in classroom and conversation registers, which are part of spoken language. At the same time, Gold (1991, cited in Gains, 1999) identified this expression as 'rhetorical question' used in conversation discourse. Therefore, the usage of *what do you think* in ENRON's email samples shows the tendency towards spoken language in line with Biber's (2006) and Gold (1991, cited in Gains, 1999) studies. Samples in 2) and 3) correspond to ways (2) and (3) above, respectively.

1) To: nels.olson@kornferry.com Subject: RE: EC Changes and Related Matters ***What do you think***? ---------------------- Forwarded by Steven J Kean/HOU/EES (ENRON_173850)
2) Subject: Pension Reform Bill ***What do you think*** of us sending out a U.S. Domestic employee (ENRON_175373)
3) Subject: Re: Legal Analysis on AB 1890   ***What do you think*** he'll add? Jeff Dasovich Sent by: Jeff Dasovich (ENRON_174503)

Last, the *Hybrid* function contains only one bundle – *let me know if*. This is because the bundle serves as a directive upon conditional function. It can be directive on the condition identified in the proposition following the bundle as shown in 1) below. Besides, some concordance lines reveal that this bundle can offer help to the email reader as exemplified in 2). Also, in 3) it could perform the *Opening up for further communication* function. This illustrates one form can perform many functions, depending on its co-text.

1) The only card I received is from Michael Geffroy. ***Let me know if*** there is a problem with this request. Thanks, Rick (ENRON_19961)
2) Please copy him on future communications. Robert, or Rick: Please ***let me know if*** you need any help such as adding months, (ENRON_138105)
3) Please l***et me know if*** you have any questions or comments.   Thanks (ENRON_60590)

To conclude, ENRON's lexical bundles display a variety of functions. They reflect one characteristic of business emails in authentic use associated with an amalgam of spoken and written languages. The former is reflected by the conversationalised feature such as *Inquiries* in 4.3.4, in which the bundle *what do you*

think was used as if the writers are speaking to the reader face to face. Still, some written language features can be perceived in ENRON's lexical bundles. For example, the *Referential expressions* function in 4.3.3 capture the lexical bundles found in business textbooks in the study of Biber (2006). Apart from linguistic expressions, the position of lexical bundles can determine their functions as shown in subcategories of *Special functions* such as *Expectation, Opening up for further communication,* and *Politeness*. Most important of all, authentic business emails show that one form may perform many different functions. This particular feature of lexical bundles in business emails was rarely identified in textbook email samples, which will be demonstrated and discussed in detail in the next section.

4.4 Classification of Functional Types in TEC

After analyzing the functional categories in ENRON, the same process was performed with TEC. All 53 lexical bundles were analyzed, also under the frameworks of Biber *et al*. (2004) and Biber (2006), with newly created subcategories in *Special functions*. Table 4.4 below shows the list of all 53 lexical bundles classified into functional types. It is found that *Special functions* is the most frequently found functional type, with 26 lexical bundles (49%), followed by *Discourse organizers*, with 13 lexical bundles (25%). The least frequently used functional types are *Stance expressions* and *Referential expressions*, with seven lexical bundles (13%) in each type.

Given that the third research question spelled out in Chapter 1 addresses similarities and differences between functional types of the lexical bundles found in authentic emails and those in business English textbook samples and that the results from ENRON have been shown in 4.3, I will report the results in TEC and discuss the common and different points between the two corpora in each subcategory throughout the functional analysis of TEC in this section.

Table 4.4 Functional types of lexical bundles in TEC

Basic Functional Type	Structural Categories	No. of types	Percentage (No.)
STANCE EXPRESSIONS	**A. Epistemic stance** (e.g. *I don't know if, I think it was*)	0	13% (7)
	B1 Desire *I would like to, we would like to, if you would like, would like to know, you would like to, I'd like to*	6	
	B2 Obligation/directive (e.g. *I want you to, you have to be*)	0	
	B3 Intention/prediction (e.g. *I was going to, are we going to*)	0	
	B4 Ability/effort *will be able to*	1	
DISCOURSE ORGANIZERS	**A. Topic introduction/focus** *I am writing to, I would be grateful, to let you know, to tell you that, to inform you that, let you know that, I'm writing to, I am interested in, I have attached a, I am attaching a*	10	25% (13)
	B. Topic elaboration/clarification (e.g. *I mean you know, as well as the*)	0	
	C. Conditions *if you have any, would be grateful if, if you need any*	3	
REFERENTIAL EXPRESSIONS	**A. Identification/focus** (e.g. *and this is a, one of the things*)	0	13% (7)
	B. Imprecision (e.g. *or something like that, and things like that*)	0	
	C1 Quantity specification (e.g. *have a lot of, the rest of the*)	0	
	C2 Tangible framing attributes *a copy of the, for you e-mail*	2	
	C3 Intangible framing attributes (e.g. *the nature of the, as a result of*)	0	
	D1 Place reference (e.g. *the United States and, of the United States*)	0	
	D2 Time reference *as soon as possible, by the end of, the end of the, end of the week, in the near future*	5	
	D3 Text reference (e.g. *shown in figure N, as shown in figure*)	0	

|	D4 Multi-functional reference (e.g. *the end of the, at the end of*)	0	
SPECIAL FUNCTIONS	A. Politeness *thank you for your, dear sir or madam, it was good to, thanks for your email, we would be grateful*	5	49% (26)
	B. Inquiries (e.g. *what are you doing*)	0	
	C. Request *please let me know, let me know what, if you could send, get back to me, you let me know, give me a call, please could you send, you could send me*	8	
	D. Opening up for further communication *you have any questions,*please do not hesitate to, hesitate to contact me, not hesitate to contact, *feel free to contact me*	5	
	E. Expectation *I look forward to, look forward to hearing, look forward to seeing, we look forward to, hearing from you soon, to hear from you, look forward to receiving*	7	
	F. Hybrid *let me know if*	1	
Total		53	100%
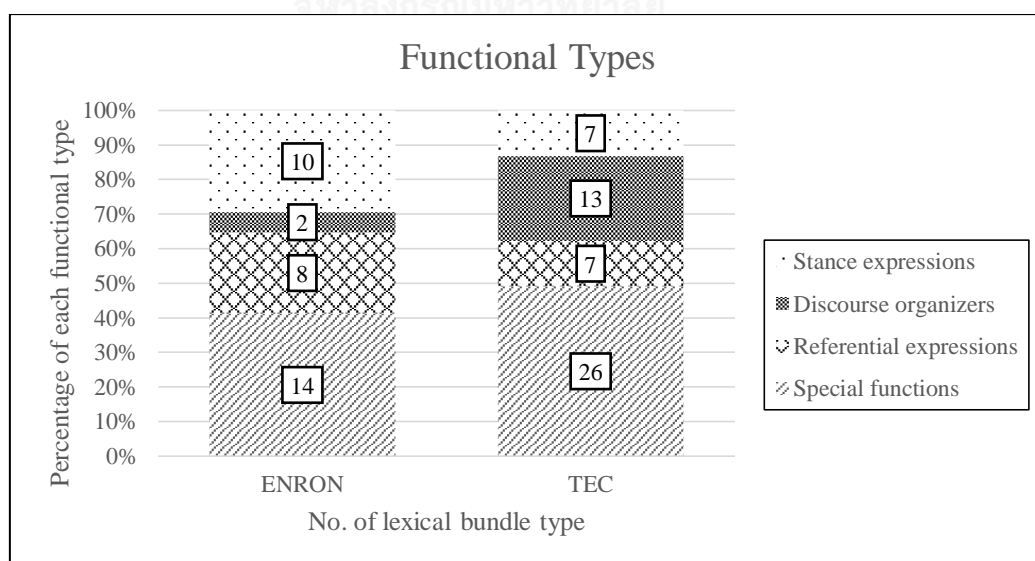
Figure 4.1 Number of functional types in each basic functional category in ENRON and TEC

In terms of functional distribution pattern, one similarity between lexical bundle types in TEC and ENRON is that the *Special functions* category predominates the list, with 41% and 49% of the total lexical bundle types in ENRON and TEC, respectively, as can be seen in Figure 4.1 above. However, the difference lies in the fact that the *Discourse organizers* function is the second largest type in TEC but the least one in ENRON while *Stance expressions* and *Referential expressions* are the least found in TEC but the second and third type in ENRON, respectively.

4.4.1 Stance Expressions in TEC

TEC contains eight lexical bundle types in *Stance expressions*, in which only two subcategories are identified. Unlike ENRON, no *Epistemic stance* bundle is found in TEC. However, the *Desire* function covers 86% of all *Stance bundles* in TEC, which has higher proportion than the one in ENRON (50%). It can therefore be inferred that textbooks concentrate on the *Desire* function while other functions receive less attention.

The *Desire* function in TEC includes six lexical bundles: *I would like to, we would like to, if you would like, would like to know, you would like to,* and *I'd like to*. It can be observed that every *Desire* bundle contains the phrase "would like to". Given that "would like" means 'want', it is noteworthy that in TEC the verb 'want' does not pass the threshold set-up. But in ENRON, this verb is found in the bundle *if you want to* and only "would like to" occurs in TEC.

The other difference is that the pronoun 'you' is part of the verbal phrase 'would like to' in TEC, while 'I' and 'we' are the only subjects in ENRON. In terms of frequency, the full form of *I would like to*, occurring 38 times, is three times more frequently found in TEC than the contracted form of *I'd like to*, which is two times less frequently found than the full form in ENRON.

Upon an investigation on functions, there are some similarities and differences of *Desire* bundles between TEC and ENRON. The similarities are discussed here first. As for the bundle *if you would like*, it denotes the same semantic property of the bundle *if you want to* in ENRON, which is to elicit the desire of the reader. Sample sentence is given in 1) below. In addition, the bundle *I would like to* can sometimes serve the *Topic introduction/focus* function as exemplified in 2). This corresponds with the function of the same bundle in ENRON in 4.3.1.

1) would be pleased to visit you in Copenhagen to discuss any of our products. ***If you would like*** any further information, please do not hesitate to contact me. Yours (TEC_T09_06)

2) Madam I am writing to ask for some information about your hotel. First of all, ***I would like to*** know if you have a single room for the night of (TEC_T59_04)

However, two main differences are identified in two lexical bundles. The first bundle is *I'd like to*. Although the bundle is less popular than the full form, *I would like to*, in both corpora, the usage of contracted form turns out to be surprisingly different from each other. First of all, regarding email recipients, *I'd like to* in ENRON is 66% used in internal emails as mentioned in 4.3.1 but 50% of this bundle, equaling four out of eight occurrences, is clearly identified as external ones in TEC (It must be noted that two samples of *I'd like to* in TEC provide insufficient information, e.g. a lack of email users' identity, to determine if it is external or internal). Second, the subcorpora contained the bundle *I'd like to* are only "company business" and "informal" in TEC whereas five out of six email types contain such bundle in ENRON. This might be due to the fact that the contracted form is less found than the full one in TEC. Last, the pragmatic function of this bundle can reflect the degree of intimacy. While the contracted form is preferred in internal emails in ENRON, where the relationship between the emails users is assumedly positive, three out of four external emails in TEC were written by the customers who are dissatisfied with the product/service of the company as these emails were labelled as complaint type. This might be inferred that the contracted form in TEC can be used to intensify the degree of dissatisfaction. Although this shade of meaning is not observed in ENRON, it might be due to the fact that the majority of emails containing the contracted form are internal, which tends to be associated with rapport rather than aggression. Sample sentences of external and internal emails are displayed in 1) and 2), respectively.

1) As result we have had to shut down production on three occasions. ***I'd like to*** meet you to discuss how things can improve in the future. (TEC_T25_48)

2) don't know about you but I just want to say well done and ***I'd like to*** congratulate every one of you by inviting you to a dinner (TEC_T51_12)

The second lexical bundle is *we would like to*. As shown in 4.3.1, there are four main types of reference of pronoun 'we' in ENRON. However, upon a concordance line examination, it is found that in TEC only two types are used: 'corporate' and 'individual' references. This is partly influenced by the fact that all emails containing this bundle in TEC are external while only internal emails are found using this bundle in ENRON. Also, the subcorpora revealed that different types of email contain this bundle: ENRON has "logistic arrangement" and "document editing/checking" while TEC has "informal" and "employment". Only the "company business" subcorpus contains the bundle *we would like to* in both corpora. Owing to the fact that both corpora shared only one subcorpus, the reference of pronoun 'we' in TEC can be either 'corporate', as in 1), or 'individual', as in 2) below, since the picture of group of colleague is not obvious as much as those in ENRON.

1) Taiwan, but these do not have a Belgian Standards Institute stamp of approval and *we would like to* complete our tests before putting them on the market. (TEC_T07_09)
2) are in the very early stages of planning for a book titled Business Advantage. *We would like to* use information from the Havaianas website. I have attached a very (TEC_T32_08)

The second and last subcategory of *Stance expressions* is *Ability/effort*. The bundle *will be able to* was identified. It is used to express the ability of the subject, as in 1). However, it can serve the *Intention/prediction* function as exemplified in 2).

1) must follow the login instructions posted on the intranet. The Help Desk at Vector *will be able to* assist employees with the procedure. Phone Number Recognition Data will a (TEC_T30_35)
2) I have been out of town on vacation. I will return on May 21 and *will be able to* attend the orientation. Please note that all necessary documents are (TEC_U15_04)

4.4.2 Discourse Organizers in TEC

Discourse organizers comprise 13 lexical bundle types or 25% of all lexical bundle types in TEC. As shown in Table 4.4, *Discourse organizers* were further grouped into two functional types:

(1) Topic introduction/focus, i.e. *I am writing to, I would be grateful, to let you know, to tell you that, to inform you that, let you know that, I'm writing to, I am interested in, I have attached a,* and *I am attaching a*

(2) Conditions, i.e. *if you have any, would be grateful if,* and *if you need any*

In comparison with ENRON, TEC provides a wider range of forms regarding the *Topic introduction/focus* function. While ENRON has only one bundle in this type, i.e. *attached is a draft*, TEC contains 10 lexical bundle types, which cover 77% of all *Discourse organizing bundles* in TEC. Also, they can be further divided into two kinds: 'complete' and 'part of' *Topic introduction/focus*. The 'complete bundle' refers to a group of lexical bundles that are found at the initial position of the clause/sentence as shown in 1) below. Six out of ten *Topic introducing* lexical bundles are the complete type: *I am writing to, I would be grateful, I'm writing to, I am interested in, I have attached a,* and *I am attaching a*. On the other hand, the other four lexical bundles are part of a larger expression used for introducing a new topic, namely *to let you know, to tell you that, to inform you that,* and *let you know that,* three of which are in the 'to-clause' form as exemplified in 2) below.

1) Dear Ms Tourelle ***I am writing to*** confirm the reservation, made by phone this morning, for two single room at the Hotel Sorbonne(TEC_T25_21)
2) Dear colleagues   I'm pleased ***to tell you that*** our financial results this year have been very positive. (TEC_T04_07)

Despite the fact that TEC contains a variety forms of *Discourse organizing bundles*, some similarities and differences between TEC and ENRON can be found as illustrated below.

Similar to ENRON in 4.3.2, the meaning associated with documentation was also identified here in *I have attached a* and *I am attaching a*. Nevertheless, the bundle in ENRON lacks the pronoun 'I'. This perhaps show that textbooks tend to provide simple structure by saying someone doing something, i.e. subject + verb + object, rather than giving complex one like *attached is a draft* in ENRON, i.e. inverted passive verb + subject.

Furthermore, to discover the pattern for the selection on form between the full bundle *I am writing to* and the contracted one *I'm writing to*, following Danet (2002), I examined the greetings (e.g. 'Dear', 'Hi' and the absence of greeting) to determine the degree of formality. Upon an investigation, it is found that 75% of the full form is used in formal writing style as supported by the greeting with either "Dear Sir or Madam", which will be discussed in detail later in 4.4.4, or "Dear" followed by the name of the email recipient. In contrast, *I'm writing to* that occurs 8 times in TEC shows a similar proportion between the lack of greetings (4 hits) and formal greetings (3 hits). One occurrence is considered informal one since the writer wrote "Hi there David", which is a colloquial expression (Danet, 2002). Similar to ENRON, the contracted form is preferred in an informal writing style. The following sample sentences display the popular greetings of each form.

(1) Dear Sir/Madam *I am writing to* enquire about your entertainment programmes at the resort. We are a small package tour operator and (TEC_T08_02)

(2) To... Customer Service Department From... Cindy Chu Date: 24 November *I'm writing to* complain about the mobile phone I bought from you a month ago. (TEC_ T54_04)

It must be noted that four occurrences of none greetings type in the bundle *I'm writing*, three of which were derived from the same textbook (T54). Therefore, such tendency might arise from the idiosyncrasy of that particular textbook.

Last, the *Conditions* function contains three lexical bundles, namely *if you have any, would be grateful if,* and *if you need any*. The overlapping lexical bundle between the two corpora is *if you have any*. Both corpora have this bundle in every email type (except "document editing/checking" in TEC). In line with 4.3.2, this bundle tends to be found towards the end of the message and can thus be viewed as a politeness strategy by opening up for further discussions. Moreover, the contact details were provided following the lexical bundles as same with samples in ENRON.

(1) We are open 24 hours\x97and of course, we deliver. *If you have any* questions on our offer, feel free to contact me at Go Organic, Wilson branch: 716-235- 9897. (TEC_T30_28)

(2) but we'll keep it open for you until the end of the week. Do call me *if you have any* queries. Best wishes Laura (TEC_T58_06)

The next bundle is *would be grateful if*. It is frequently used as a part of a request exemplified below. The proposition following this bundle contains the modal verb 'could' in 20 out of 24 occurrences. Therefore, it appears that the writer politely asked the email recipient to perform an action upon their request.

(1) I *would be grateful if* you would send ten more players as soon as possible. (TEC_U15_18)

(2) We *would be grateful if* you could include some samples. (TEC_T25_34)

4.4.3 Referential Expressions in TEC

The other least found functional type is TEC is *Referential expressions*, with only seven lexical bundle types, like *Stance expressions*, covering 13% of all lexical bundles in TEC. They are further categorized into two types:

1) Tangible framing attributes, i.e. *a copy of the* and *for you e-mail*

2) Time reference, i.e. *as soon as possible, by the end of, the end of the, end of the week,* and *in the near future*

As can be seen, the largest type of referential bundles is *Time reference*, accounting for five of them. Their functions correlate with their forms as discussed in 4.2.3. Unlike ENRON, TEC offers more various types of *Time reference* bundles with two main groups: "deadline" and "unspecific time" reference.

The second function of *Referential expressions* is *Tangible framing bundles*. The two of them are *a copy of the* and *for you e-mail*. The former is overlapping with one lexical bundle in ENRON in 4.3.3. However, the bundle *for your e-mail* is identified only in TEC. It is part of a longer expression 'thank you for your e-mail'. Therefore, it is considered as politeness strategy. Besides, it can be considered *Topic introduction/focus* bundle since this expression is identified at the beginning of the message. Therefore, it functions similar to *thanks for your message* of ENRON as mentioned in 4.3.4. The examples of each lexical bundle are provided below.

(1) Please can you email me *a copy of the* new customer address list? I also need the quarterly sales report — (TEC_T59_11)

(2)Dear Steve Thank you *for your e-mail*. Please see my responses in between your requests below. Best wishes (TEC_T09_07)

4.4.4 Special Functions in TEC

The *Special functions* category accounts for 49% of all lexical bundles in TEC, with 26 out of 53 lexical bundle types. They are further divided into five subcategories:

(1) Politeness, i.e. *thank you for your, dear sir or madam, it was good to, thanks for your email,* and *we would be grateful*

(2) Request, i.e. *please let me know, let me know what, if you could send, get back to me, you let me know, give me a call, please could you send,* and *you could send me*

(3) Opening up for further communication, i.e. *you have any questions,*please do not hesitate to, hesitate to contact me, not hesitate to contact,* and **feel free to contact me*

(4) Expectation, i.e. *I look forward to, look forward to hearing, look forward to seeing, we look forward to, hearing from you soon, to hear from you,* and *look forward to receiving*

(5) Hybrid, i.e. *let me know if*

Based on the density, the largest group of *Special functions* lexical bundle is *Request*, with eight out of 26 lexical bundle types. The second one is *Expectation*, which contains seven bundle types. The less found types are *Politeness, Opening up for further communication* and *Hybrid*, respectively. This distribution pattern of TEC's *Special functions* is similar with the one in ENRON in that Hybrid is the least found type. However, a number of differences are that the *Inquiries* function is not identified in TEC.

The *Request* function contains eight lexical bundle types. Among these bundles, only two of them explicitly have the word 'please'. However, the concordance lines of the other bundles show that 'please' is found in the extended context. Moreover, two main patterns can be observed: First, the three-word bundle *let me know* is embedded in three different lexical bundles, i.e. *please let me know*, *let me know what*, and *you let me know*. Second, the phrase 'you could send' or 'could

you send' is identified in another three bundles. Therefore, this illustrates how textbook sample emails tend to vary in terms of the forms of lexical bundles, resulting in a higher number of lexical bundle type under the same subcategory. Sample sentences are shown below.

1) Please **get back to me** if this would be possible. Yours sincerely, Regine Aritady (TEC_T15_02)
2) If you are ever in London, please don't forget to **give me a call**. Perhaps we can meet for lunch. Regards Mike Michael Kennedy (TEC_U13_26)
3) I've put together the following itinerary. Can **you let me know** what you think about it? (TEC_T25_19)

The second most frequently found subtypes of *Special functions* is *Expectation*, with seven lexical bundle types. Similar to ENRON in 4.3.4, this function displays a variety of lexical bundle forms centered around the verbal phrase 'look forward to' and usually found towards the end of the email message. However, only one lexical bundle does not result from that verbal phrase - *to hear from you*. Although it looks similar to other bundles in the *Topic introduction/focus* function, it in fact follows the proposition like 'hope to' and 'wait to' as exemplified in 1) and 2) below. Nevertheless, some sample sentences also reveal that *to hear from you* is part of a politeness bundle *it was good to*, which will be discussed next.

1) at one time, please send us your current catalogue and price list. We hope **to hear from you** soon. Peter Crane Chief Buyer F. Lynch & Co. Ltd Nesson House,(TEC_T07_08)
2) available on Saturdays and Sundays due to my restaurant job. I'll wait **to hear from you** prior to the orientation and training next week. Thank you for (TEC_U03_04)

The *Politeness* function has five lexical bundle types, namely *thank you for your, dear sir or madam, it was good to, thanks for your email,* and *we would be grateful*. Unlike ENRON, not only the phrase 'thank you/thanks' serves the *Politeness* function but also other lexical bundles lacking such mark can be classified into this subtype. Those lexical bundles are *dear sir or madam, it was good to,* and *we would be grateful*. This is due to the context which shows that theses bundle types can serve the writer's purpose to be polite via greetings, small talks, and performing a request.

The function of greeting is derived from the bundle *dear sir or madam*. It must be noted that this lexical bundle cannot be found in ENRON at all. The second bundle, *it was good to*, locates within the small talks preceding the main message because it comes right after the greetings as shown in 1) and 2) below. Last, the bundle *we would be grateful* preludes a request as in 3), 4) and 5) which precedes different propositions, namely 'if'-clause, prepositional phrase, and 'to' fragment, respectively.

1) Hi Aditya **It was good to** see to you yesterday. Hope you well. Could you sending me (TEC_T03_04)
2) Dear Peter. **It was good to** meet you at the conference in Brussels last month. As promised, (TEC_T16_07)
3) send me a current price list, together with information on delivery times and costs. **We would be grateful** if you could include some samples. Best regards Paulo Sambucco (TEC_T25_34)
4) need a firm of lawyers in France to represent us on some legal matters. **We would be grateful** for some information about the legal services that your firm offers. (TEC_U05_11)
5) account, amounting to €4,500 May we please remind you that this amount is still outstanding. **We would be grateful** to receive a bank transfer in full settlement without further delay. (TEC_U05_16)

The *Opening up for further communication* function is the subcategory of *Special functions*, with five lexical bundle types: *you have any questions, *please do not hesitate to, hesitate to contact me, not hesitate to contact,* and **feel free to contact me*. One overlapping bundle with ENRON is *you have any questions*. However, the verb 'hesitate' is found only in TEC. They are found in three lexical bundle types: **please do not hesitate to, hesitate to contact me,* and *not hesitate to contact*. The variation results from two types of negative form, 'do not' and 'don't', preceding the bundle. In addition, two five-word bundles are identified only in this function throughout TEC. Therefore, it can be inferred that the two expressions are so fixed that they can sustain the form longer than four-word length.

Last, the *Hybrid* function also has one lexical bundle *let me know if* which is also identified in ENRON. Furthermore, they perform not only a directive upon condition but also the *Opening up for further communication* function as well as offering help to the email recipient.

1) weaker. Do you want me to get together a revised forecast for July-Sep? *Let me know if* there's anything else I can do to help with the (TEC_T04_06)

2) us for lunch at the track and for an evening dinner in San Paulo. *Let me know if* you can attend. Look forward to hearing from you. Best regards (TEC_T10_07)

3) received your completed questionnaire. Could you please return this as soon as possible. Please *let me know if* you have any queries. Best regards Sara Giles (TEC_T25_09)

4.5 The Comparison of Structural and Functional Types of Lexical Bundles in ENRON and TEC

The above four sections report on structural and functional analyses of lexical bundles in both ENRON and TEC. In this section, I focus on similarities and differences between structural and functional types in ENRON and TEC, respectively. Further details are discussed in turn below.

The similarities shared between ENRON and TEC can be identified in both structural and functional types of lexical bundles. Firstly, the distribution pattern of structural categories in both corpora is in the same direction, in which VP-based, Dependent clause, and NP/PP-based bundles were identified as the most to the least frequently found types of lexical bundles, respectively. Second, despite the fact that the two corpora are different in size, there are 15 identical lexical bundles found in each corpus. This suggests that some fundamental lexical bundles can be found in business emails whether in textbooks or in real use. Moreover, some of the rest of the bundles in the two corpora also share the three-word or two-word bundles; for example, a three-word bundle *thanks for your* is found in ENRON's *thanks for your message* and TEC's *thanks for your email.* Upon functional analyses in the two corpora, the findings show that the major function in both corpora is the *Special functions* category, which contains the highest number of lexical bundle types in each corpus. This reflects that business emails in the authentic use and textbooks are centered around formulaic expressions used pragmatically in communicative acts.

Overall, based on the structural and functional analyses above, a particular distinctive characteristic of emails can be observed. That is, business emails are stylistically close to the spoken language, as suggested by some structural and functional patterns of lexical bundles. Based on the structural analysis, the formal

properties of lexical bundles are marked by the predominance of verbal elements, as well as the frequent use of personal pronouns and the use of contraction forms. These characteristics are similar to those found in previous studies of spoken registers such as Biber *et al.*'s (2004) study of conversational register, and those of business emails, e.g. Baron (1998), Gimenez (2000), Danet (2002). Moreover, the majority of functional types belong to the *Special functions* category, which are primarily part of communicative acts, e.g. request, inquiries. These lexical bundles are also found in a number of texts including classroom teaching and conversation in Biber *et al.*'s (2004) study.

Apart from the lexical bundles themselves, the positions of lexical bundles found in the text and types of email also play an essential role in explaining the functions of lexical bundles in both corpora. For example, the *Topic introduction/focus* bundles tend to be found at the beginning of the message as displayed in 4.4.2. At the same time, the functions of pronoun 'we' in the bundle *we would like to* can vary, depending on the type of email as illustrated in 4.3.1 and 4.4.1.

Nevertheless, some distinctions are found between ENRON and TEC. In terms of number of types, TEC contains a larger variety of lexical bundles than ENRON, with 53 and 34 types, respectively. This might be a result of the pedagogic purposes of textbooks which aim to present a wide range of expressions to learners. Upon structural analyses of ENRON and TEC, it was discovered that although the two corpora share similar tendencies regarding the density of lexical bundle types in "VP-based" and "Dependent clause" bundles, they are different in detail in three main aspects. First of all, ENRON displays denser use of spoken features than TEC. Although TEC has 85% of distributional pattern of verbal elements of lexical bundles higher than ENRON, which has only 73.5%, further analysis reveals that ENRON is yet more inclined towards spoken register, as evidenced by the use of contractions in both "VP-based" and "Dependent clause" types whereas TEC contains contracted forms in "Dependent clause" type only. Second, structural types of lexical bundles in TEC are more varied than those in ENRON. However, a wider range of forms does not entail an even distribution of the lexical bundles across functional categories, which will be discussed further below. Last, in each corpus one subcategory is missing. This is the case of the "Yes-no question fragments" subtype which is found

only in TEC and that of the "Wh-clause fragments" subtype which can be found only in ENRON.

With regards to functional types, ENRON and TEC differ significantly. First of all, it is found that in TEC some functional categories are overwhelmed with various lexical bundles types whereas others contain very few or even no bundles. The corpus offers fewer functional subcategories than ENRON in every basic functional type, except *Discourse organizers*. This reflects an insufficient presentation of functional types in business English textbook samples in the present study though the corpus consists of 77 textbook samples. On the other hand, one reason behind the overuse of certain functional subtypes might be attributed to the pedagogic purposes of textbooks. For example, the *Topic introduction/focus* function contains 10 lexical bundles types in TEC while only one type is identified in ENRON. This could be influenced by the fact that the email samples in TEC consist of internal and external communication, resulting in a requirement for longer expressions in introducing a topic, compared with internal emails in ENRON corpus. This leads to another observation that lexical bundles in textbook email samples tend to reflect a more formal writing style in comparison with those found in authentic ones as shown in the preference on full form in *I would like to* and *I am writing to* bundles, and two lexical bundle which are used in traditional business letters, i.e. *dear sir or madam and I am writing to ("Business letter format How to write a business letter,"* 2012*)*. Thirdly, although there are 15 overlapping lexical bundles between ENRON and TEC, some might not share the same functions. Some functions in TEC can be interpreted in different ways from the same bundle identified in ENRON. This is because further analyses reveal that some pragmatic functions of certain lexical bundles in TEC have limitedly been represented. For example, the references of the pronoun 'we' in *we would like to* are less variant in TEC than those in ENRON.

In conclusion, while previous research on business English textbooks tends to highlight that textbooks insufficiently represent the language in real business use (e.g. Williams, 1988; Nelson, 2000; Angouri, 2010), the present study has found that, as far as email is concerned, textbooks manage to capture an essential characteristic of emails, i.e. it is highly spoken and conversational, and that some formulaic expressions are in fact overemphasized and surrounded only particular categories such

as *Discourse organizers*, which contain a variety of lexical bundles. At the same time, when a functional perspective is applied, the present study also yields results that correspond to those previous studies in that some functions often used in authentic emails are under underrepresented such as *Stance bundles*.

Chapter 5
Conclusion

This chapter presents an overview of results of the study. I then offer some pedagogical implications and suggestions for further research on business emails.

5.1 Summary of the Study

The present study examined lexical bundles in sample business English emails and those in authentic ones, employing the frameworks of Biber *et al*. (2004) and Biber (2006) which divide lexical bundles into two main categories: structural and functional. In order to identify the characteristics of lexical bundles in business English textbook emails and authentic business emails, I have compiled two data sets using the corpus-based method. After obtaining the lists of lexical bundles from the two corpora on the basis of criteria displayed in section 3.3, I analyzed them in terms of their structures and their functions. These findings fulfilled the first two objectives of the present study, which are to examine structural and functional types of lexical bundles found in authentic business emails and those in business English textbooks. As for the third objective, I compared and contrasted structural and functional types of lexical bundles from each corpus. Details of answers for each research question are provided below.

5.1.1 Answers to Research Question 1

The first research question of this study is "What structural types of lexical bundles are there in authentic business emails and those in business English textbooks?" The findings show that the two data sets share a number of structural types of lexical bundles, i.e. VP-based, Dependent clause, and NP/PP-based bundles. In addition, the verbal elements, i.e. VP-based and Dependent clause bundles, are predominant in ENRON and TEC. Also, in both corpora there are frequent uses of contractions and personal pronouns, which are associated with the features of conversational register identified in previous studies (e.g. Baron, 1998; Danet, 2002; Gimenez, 2000). Based on these findings, the business email is a written text type

with linguistic elements close to the spoken register. In other words, it is a hybridity of spoken and written registers.

5.1.2 Answers to Research Question 2

The second research question addresses what functional types of lexical bundles there are in authentic business emails and those in business English textbooks. It is found that the majority of functional categories lie in the *Special functions* category, which focus on 'communicative acts' such as requesting, inquiring, opening up for further communication, etc. This again points to the fact that the spoken register tends to dominate email discourse. However, despite the fact that ENRON contains fewer lexical bundle types, the corpus cover a wider range of functional subcategories of lexical bundles than those in TEC. At the same time, different forms of lexical bundles with similar meanings, such as *to let you know*, *to tell you that* and *to inform you that*, are overemphasized in TEC, resulting in a less variety of functional types.

5.1.3 Answers to Research Question 3

The third research question is on similarities and differences between the forms and functional types of the lexical bundles found in authentic emails and those in business English textbook samples. The similarity between ENRON and TEC is that most of their lexical bundles are set towards the same direction – i.e. sharing a hybridity of spoken and written registers regarding forms, being highly communicative with respect to the predominant functions of lexical bundles, and illustrating that forms and functions are considerably correlated.

However, some further investigations also reveal different emphases in each corpus. First of all, as for structural types of lexical bundles, business emails in ENRON are less formal compared to those in TEC on the basis of occurrences of contractions; there are more cases of contraction forms in lexical bundles in ENRON than in TEC. Second, TEC contains some lexical bundles which are considered formal writings such as *I am writing to* and *dear sir or madam*. Such bundles are not found in authentic email samples in the present data. Furthermore, TEC contains more various types of lexical bundles, some of which denote similar meanings, such as *I have attached a* and *I am attaching a*. Finally, there is an unbalanced proportion in the textbook email corpus. While TEC contains more variant lexical bundles, most of

them belong to only a few functional types, e.g. *Topic introduction/focus, Request, Expectation*, and some types that can be found in ENRON are missing in TEC. It can be said that although textbooks seem to capture essential characteristics of email discourse as mentioned above, some differences between textbook business email samples and authentic business emails can still be found.

5.2 Implications of the Study

The two corpora of the present study can contribute to the instruction of business emails, especially the corpus of textbook email samples which was unprecedentedly compiled. Furthermore, the frameworks I adopted from Biber *et al.* (2004) and Biber (2006) can be applied in future research into business English since it is specifically related to business texts, while Biber's aforementioned frameworks focus on expressions frequently found in academic discourse. More importantly, the present study have some contributions to three main areas: the study of lexical bundles, the discourse of business emails, and pedagogical implications.

First of all, based on the findings of the present study, it can be seen that lexical bundles can be employed to investigate other written genres besides those in academic disciplines. Previous research on lexical bundles often looked at genres from the academic circle, for example, academic prose (Biber et al., 1999), research articles (e.g. Hyland, 2008a; Hyland, 2008b), and learner writing (e.g.Y.-H. Chen & Baker, 2010; Lie, 2013), engineering textbooks (L. Chen, 2008), biology textbooks and research articles (Sánchez, 2014). However, research into business English has rarely been based on lexical bundles, with a few exceptions (e.g. Sinturat, 2010; Sriumporn, 2015). The present study has shown how lexical bundles can be used to categorize and make comparisons between the same text type (business email) but in different sources (corporate and textbooks).

Second, the present study has shown some light on the nature of business email discourse. Firstly, although the data in this study is entirely business email, the identified lexical bundles have different patterns of distributions and usages. Some of these are very much tied to subtypes of business email. This can be illustrated in the reference of "we" in the *Desire* bundle *we would like to*. As shown in Section 4.3.1 and 4.4.1, the reference of we can range from an individual to a corporate reference, depending on many factors such as the types of email, e.g. "logistic arrangement",

"informal business email", etc., and the relationship between email users, i.e. external or internal communication. Secondly, although business emails are written texts used in the apparently formal context, i.e. for work, the discourse of email does not always appear to be in a formal style, which is often associated with business writing (Gains, 1999). This is shown through a range of lexical bundles found in the study, including *I'd like to, I don't think, thanks for your message*, etc. On the other hand, a look at three lexical bundles in the contracted form reveals that 53% of *I'd like to* in ENRON, 55% of *I'd like to* in TEC and 62.5% of *I'm writing to* of TEC were identified in informal writing styled emails. This can thus be inferred that the contraction form might be one of characteristics of informal writing styles. Other factors, such as company culture, relationship between email users and topics in discussion, should be taken into accounts (Gimenez, 2005). Consequently, we cannot pin down that an email is an obviously being informal or formal communication method since it depends on various factors, including the email writer, the email recipient and the context.

Last, the present study clearly has pedagogical implications. First of all, the study lends support to lexis-based teaching approaches, e.g. lexical approach (Lewis, 2000) and genre-based approach (Dirgeyasa, 2016), since it focuses on employing lexical bundles in business email writing. As they are frequent recurrent recognizable in business communication, they would be useful to students to learn. Also, they are linguistic expressions that come with both form and function and thereby help students perform effectively or naturally in their business email writing. On a general plane, it is recommended that such a combined structural – functional perspective be applied to the design and development of textbook and teaching materials in business English. This is in order to create a well-balanced representation of linguistic units to be taught in class. For example, based on the findings in the present study, some functional subcategories of lexical bundles should be supplemented in business English textbooks, such as the *Epistemic Stance expressions* (i.e. *I don't think* and *I don't know*) and *Intangible framing attributes* (e.g. *as a result of, in the light of*) functions, since they were not found in textbook email samples. In addition, using a corpus of business emails can help teachers identify recurrent formulaic expressions

since corpus software can extract not only lexical bundles but also other linguistic expressions, including collocations, keywords, etc.

5.3 Suggestions for Further Study

Here are some suggestions for further studies.

1. Given that the corpora contain different types and that findings show that types have impacts on lexical bundles, future studies of business emails can be conducted on lexical bundles in different specific types of business emails such as request, complaint, making orders, etc.

2. The present study focuses on the comparison between authentic and textbook email samples. Further research could be done on other factors, including writers, e.g. native vs. non-Native English writers, and corporate cultural differences, e.g. private sections vs. public organizations or academic vs. business corporations. This might shed some new light on how business email discourse is conducted from different perspectives.
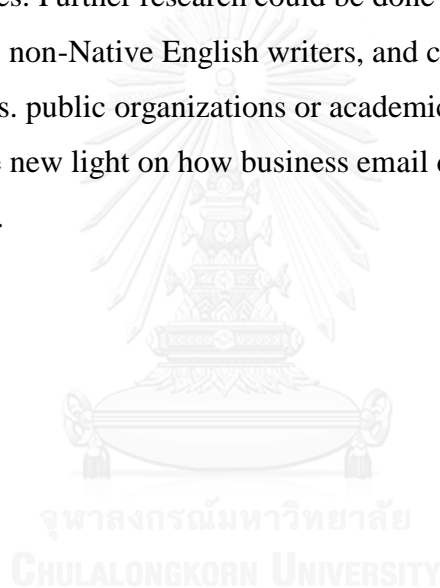
REFERENCES

Angell, D. F., & Heslop, B. (1994). *Elements of e-mail style: Communicate effectively via electronic mail*: Addison-Wesley Longman Publishing Co., Inc.

Angouri, J. (2010). Using textbook and real life data to teach turn taking in business meetings. In N. Harwood (Ed.), *Handbook of Teaching Materials.* (pp. 373-395). New York: Cambridge University Press.

Anthony, L. (2004). AntConc: A learner and classroom friendly, multi-platform corpus analysis toolkit. *proceedings of IWLeL*, 7-13.

Asztalos, R. (2014). The Move Structure of Confirmation Emails: a Genre Based Investigation of a Sub-Genre of Business Emails. *Practice and Theory in Systems of Education,, 9*(3), 201-212.

Barbieri, F., & Eckhardt, S. E. B. (2007). Applying corpus-based findings to form-focused instruction: The case of reported speech. *Language Teaching Research, 11*(3), 319-346.

Bargiela-Chiappini, F., & Zhang, Z. (2012). Business English *The Handbook of English for Specific Purposes* (pp. 193-211): John Wiley & Sons, Ltd.

Baron, N. S. (1998). Letters by phone or speech by other means: the linguistics of email. *Language & Communication, 18*(2), 133-170. doi:http://dx.doi.org/10.1016/S0271-5309(98)00005-6

Berkeley. (n.d.). Enron Email Analysis Project.   Retrieved from http://bailando.sims.berkeley.edu/enron_email.html

Bhatia, V. K. (1993). *Analysing genre: Language use in professional settings*. London: Longman.

Bhatia, V. K., & Bremner, S. (2012). English for business communication. *Language Teaching, 45*(4), 410-445. doi:10.1017/S0261444812000171

Biber, D. (2006). *University language: A corpus-based study of spoken and written registers* (Vol. 23): John Benjamins Publishing.

Biber, D. (2009). A corpus-driven approach to formulaic language in English: Multi-word patterns in speech and writing. *International Journal of Corpus Linguistics, 14*(3), 275-311. doi:10.1075/ijcl.14.3.08bib

Biber, D., & Barbieri, F. (2007). Lexical bundles in university spoken and written registers. *English for specific purposes, 26*(3), 263-286.

Biber, D., & Conrad, S. (2009). *Register, genre, and style*: Cambridge University Press.

Biber, D., Conrad, S., & Cortes, V. (2004). If you look at..: Lexical Bundles in University Teaching and Textbooks. *Applied Linguistics, 25*(3), 371-405.

Biber, D., Johansson, S., Leech, G., Conrad, S., Finegan, E., & Quirk, R. (1999). *Longman grammar of spoken and written English* (Vol. 2): MIT Press.

Bly, R. W. (1999). *The Encyclopedia of Business Letters, Fax Memos, and E-mail*: Career Press, Incorporated.

Booher, D. (2001). *E-writing: 21st-century tools for effective communication*: Simon and Schuster.

Bremner, S. (2010). Collaborative writing: Bridging the gap between the textbook and the workplace. *English for specific purposes, 29*(2), 121-132. doi:10.1016/j.esp.2009.11.001

Brounstein, M., Bell, A. H., & Smith, D. M. (2007). *Business Communication: Communicate Effectively in Any Business Environment*. The United States of America: John Wiley & Sons.

Business letter format How to write a business letter. (2012). Retrieved from http://www.macmillandictionaryblog.com/business-letter-format

Carrió-Pastor, M. L., & Muñiz-Calderón, R. (2013). Variation of English business e-mails in Asian countries. *Iberica, 26*, 55-76.

Chan, C. S. C. (2009). Forging a link between research and pedagogy: A holistic framework for evaluating business English materials. *English for specific purposes, 28*(2), 125-136. doi:http://dx.doi.org/10.1016/j.esp.2008.12.001

Chen, L. (2008). *An investigation of lexical bundles in electrical engineering introductory textbooks and ESP textbooks.* (Master of Arts), Carleton University

Chen, Y.-H., & Baker, P. (2010). Lexical Bundles in L1 And L2 Academic Writing. *Language Learning & Technology, 14*(2).

Conrad, S., & Biber, D. (2005). *The frequency and use of lexical bundles in conversation and academic prose.*

Cortes, V. (2002). Lexical bundles in freshman composition. In R. Reppen, S. M. Fitzmaurice, & D. Biber (Eds.), *Using corpora to explore linguistic variation* (Vol. 9, pp. 131-145). John Benjamins Publishing.

Cortes, V. (2004). Lexical bundles in published and student disciplinary writing: Examples from history and biology. *English for specific purposes, 23*(4), 397-423.

Cortes, V. (2006). Teaching lexical bundles in the disciplines: An example from a writing intensive history class. *Linguistics and Education, 17*(4), 391-406. doi:http://dx.doi.org/10.1016/j.linged.2007.02.001

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*, 33-43. doi:10.1016/j.jeap.2012.11.002

Coxhead, A., & Byrd, P. (2007). Preparing writing teachers to teach the vocabulary and grammar of academic prose. *Journal of Second Language Writing, 16*(3), 129-147. doi:http://dx.doi.org/10.1016/j.jslw.2007.07.002

Danet, B. (2002). The language of email. *European Union Summer School*, 1-27.

Danielewicz-Betz, A. (2016). Empirical Data Analysis: The Email Corpus *Communicating in Digital Age Corporations* (pp. 139-276): Springer.

de Moraes Garcez, P. (1993). Point-making styles in cross-cultural business negotiation: A microethnographic study. *English for specific purposes, 12*(2), 103-120.

Dirgeyasa, I. W. (2016). Genre-Based Approach: What and How to Teach and to Learn Writing. *English Language Teaching, 9*(9), 45.

Donna, S. (2000). *Teach Business English*: Cambridge University Press.

Dudley-Evans, T., & St John, M. J. (1998). *Developments in English for specific purposes: A multi-disciplinary approach*: Cambridge university press.

Ellis, M., & Johnson, C. (1994). *Teaching Business English: An Introduction to Business English for Language Teachers, Trainers, and Course Organizers*: Oxford University Press.

Evans, S. (2010). Business as usual: The use of English in the professional world in Hong Kong. *English for specific purposes, 29*(3), 153-167.

Evans, S. (2012). Designing email tasks for the Business English classroom: Implications from a study of Hong Kong's key industries. *English for specific purposes, 31*(3), 202-212. doi:http://dx.doi.org/10.1016/j.esp.2012.03.001

Flowerdew, J., & Wan, A. (2010). The linguistic and the contextual in applied genre analysis: The case of the company audit report. *English for specific purposes, 29*(2), 78-93.

Flowerdew, L. (2008). *Corpus-based analyses of the problem-solution pattern: a phraseological approach* (Vol. 29): John Benjamins Publishing.

Gains, J. (1999). Electronic Mail—A New Style of Communication or Just a New Medium?: An Investigation into the Text Features of E-mail. *English for specific purposes, 18*(1), 81-101. doi:http://dx.doi.org/10.1016/S0889-4906(97)00051-3

Gimenez, J. C. (2000). Business e-mail communication: some emerging tendencies in register. *English for specific purposes, 19*(3), 237-251. doi:http://dx.doi.org/10.1016/S0889-4906(98)00030-1

Gimenez, J. C. (2005). The language of business email: An opportunity to bridge theory and practice. *Scripta Manent, Journal of the Slovene Association of LSP teachers, 1*(1), 13-23.

Gimenez, J. C. (2006). Embedded business emails: Meeting new demands in international business communication. *English for specific purposes, 25*(2), 154-172. doi:http://dx.doi.org/10.1016/j.esp.2005.04.005

GoogleTrends (Producer). (2016, 5 December 2016). Comparing the term email and e-mail as a search enquiry. Retrieved from https://www.google.com/trends/explore?date=all&q=email,e-mail

Grabowski, Ł. (2015). Keywords and lexical bundles within English pharmaceutical discourse: A corpus-driven description. *English for specific purposes, 38*, 23-33. doi:http://dx.doi.org/10.1016/j.esp.2014.10.004

Gray, B., & Biber, D. (2013). Lexical frames in academic prose and conversation. *International Journal of Corpus Linguistics, 18*(1), 109-135. doi:10.1075/ijcl.18.1.08gra

Harwood, N. (2010). *English language teaching materials : theory and practice*: Cambridge University Press.

Hooi, C. M., & Shuib, M. (2014). An analysis of hedging devices in complaint business letters. *GEMA; Online Journal of Language Studies, 14*(3), 123-142.

Huang, K. (2015). More does not mean better: Frequency and accuracy analysis of lexical bundles in Chinese EFL learners' essay writing. *System, 53*, 13-23. doi:http://dx.doi.org/10.1016/j.system.2015.06.011

Hutchinson, T., & Waters, A. (1987). *English for specific purposes*: Cambridge University Press.

Hyland, K. (2008a). Academic clusters: Text patterning in published and postgraduate writing. *International Journal of Applied Linguistics, 18*(1), 41-62.

Hyland, K. (2008b). As can be seen: Lexical bundles and disciplinary variation. *English for specific purposes, 27*(1), 4-21.

Jablonkai, R. (2010). English in the context of European integration: A corpus-driven analysis of lexical bundles in English EU documents. *English for specific purposes, 29*(4), 253-267. doi:https://doi.org/10.1016/j.esp.2010.04.006

Jenkins, S., & Hinds, J. (1987). Business letter writing: English, French, and Japanese. *TESOL Quarterly, 21*(2), 327-349.

Kazemi, M., Katiraei, S., & Rasekh, A. E. (2014). The Impact of Teaching Lexical Bundles on Improving Iranian EFL Students' Writing Skill. *Procedia - Social and Behavioral Sciences, 98*, 864-869. doi:http://dx.doi.org/10.1016/j.sbspro.2014.03.493

Knight, D., Adolphs, S., & Carter, R. (2013). Formality in Digital Discourse: A Study of Hedging in CANELC. In J. Romero-Trillo (Ed.), *Yearbook of Corpus Linguistics and Pragmatics 2013: New Domains and Methodologies* (pp. 131-152). Dordrecht: Springer Netherlands.

Kohut, G. F., & Segars, A. H. (1992). The president's letter to stockholders: An examination of corporate communication strategy. *Journal of Business Communication, 29*(1), 7-21.

Lampi, M. (1986). *Linguistic components of strategy in business negotiations* (Vol. 85): Helsinki School of Economics.

Leber, J. (2013). The immortal life of the Enron e-mails. Retrieved from https://www.technologyreview.com

Leech, G. N. (1989). *Meaning and the English verb* (2 ed.). Harlow: Longman.

Lewis, M. (2000). Language in the lexical approach. *Teaching collocation: Further developments in the lexical approach*, 126-154.

Lie, J. A. (2013). *"The fact that the majority seems to be..." - A corpus-driven investigation of lexical bundles in native and non-native academic English.* (Master Degree), UNIVERSITY OF OSLO.

Lindsell-Roberts, S. (2004). *Strategic business letters and e-mail*: Houghton Mifflin Harcourt.

Louhiala-Salminen, L. (1999). From business correspondence to message exchange: What is there left. In M. Hewings & C. Nickerson (Eds.), *Business English: research into practice* (pp. 100-114): Longman.

Maier, P. (1992). Politeness strategies in business letters by native and non-native English speakers. *English for specific purposes, 11*(3), 189-205.

Mallon, R., & Oppenheim, C. (2002). Style used in electronic mail. *Aslib Proceedings, 54*(1), 8-22. doi:doi:10.1108/00012530210697482

Mbodj-Diop, N. B. (2016). *Lexical bundles in medical research articles: Structures and functions.* (Master of Arts), Michigan State University

McCarthy, M., & Handford, M. (2004). Invisible to us: A preliminary corpus-based study of spoken business English. *Discourse in the professions: Perspectives from corpus linguistics, 24*, 167-202.

Mehrpour, S., & Mehrzad, M. (2013). A comparative genre analysis of English business e-mails written by Iranians and native English speakers. *Theory and Practice in Language Studies, 3*(12), 2250-2261. doi:10.4304/tpls.3.12.2250-2261

Merriam-Webster, I. (1993). *Merriam-Webster's guide to business correspondence*: Merriam Webster.

Moreno, R. G., & Sznajder, H. S. (2013). Business communication across three European cultures: A contrastive analysis of British, Spanish and Polish email writing. *Ibérica: Revista de la Asociación Europea de Lenguas para Fines Específicos (AELFE)*(26), 77-98.

Munir, S. (2001). Politeness Strategies in Business Letters by Malaysian Writers. *A Journal on English for Specific Purposes, 7*(1), 39-53.

Murphy, M., & Levy, M. (2006). Politeness in intercultural email communication: Australian and Korean perspectives. *Journal of Intercultural Communication,* (12). Retrieved from https://www.immi.se/intercultural/nr12/murphy.htm

Nelson, M. B. (2000). *A Corpus-Based Study of the Lexis of Business English and Business English Teaching Materials.* University of Manchester.

Nesi, H. (2012). ESP and Corpus Studies. In B. Paltridge & S. Starfield (Eds.), *The Handbook of English for Specific Purposes* (pp. 407-426): John Wiley & Sons, Ltd.

Nickerson, C. (2005). English as a lingua franca in international business contexts. *English for specific purposes, 24*(4), 367-380. doi:http://dx.doi.org/10.1016/j.esp.2005.02.001

Nickerson, C. (2010). The Englishes of business. In A. Kirkpatrick (Ed.), *The Routledge Handbook of World Englishes* (pp. 506-519): Routledge.

Nickerson, C., & Planken, B. (2016). *Introducing Business English*: Routledge.

O'keeffe, A., McCarthy, M., & Carter, R. (2007). *From corpus to classroom: Language use and language teaching*: Cambridge University Press.

Öztürk, Y., & Köse, G. D. (2016). Turkish and Native English Academic Writers' Use of Lexical Bundles. *Anadili Türkçe ve Ġngilizce Olan Yazarların Sözcük Öbeği Kullanımı., 12*(1), 149-165.

Pan, F., Reppen, R., & Biber, D. (2016). Comparing patterns of L1 versus L2 English academic professionals: Lexical bundles in Telecommunications research journals. *Journal of English for Academic Purposes, 21*, 60-71. doi:http://dx.doi.org/10.1016/j.jeap.2015.11.003

Peterson, K., Hohensee, M., & Xia, F. (2011). *Email formality in the workplace: A case study on the Enron corpus.* Paper presented at the Proceedings of the Workshop on Languages in Social Media.

Ruan, Z. (2016). Lexical Bundles in Chinese Undergraduate Academic Writing at an English Medium University. *RELC Journal*.

Sadeghi, K., Modirkhameneh, S., Alavinia, P., Khonbi, Z. A., Jalali, Z. S., & Moini, M. R. (2014, 2014/05/06). *Structure of Lexical Bundles in Introduction Section of Medical Research Articles.* Paper presented at the Procedia - Social and Behavioral Sciences.

Sánchez, P. (2014). Lexical bundles in Biology Differences between textbooks and research articles. *Revista Española de Lingüística Aplicada, 27*(2), 492-513. doi:10.1075/resla.27.2.12san

Schmitt, N., Grandage, S., & Adolphs, S. (2004). Are corpus-derived recurrent clusters psycholinguistically valid. In N. Schmitt (Ed.), *Formulaic sequences: Acquisition, processing and use* (pp. 127-151): John Benjamins Publishing.

Sherblom, J. (1988). Direction, function, and signature in electronic mail. *Journal of Business Communication, 25*(4), 39-54.

Sinclair, J. (2004). *Trust the text: Language, corpus and discourse*: Routledge.

Sinturat, T. (2010). *An Analysis Of Lexical Phrases In Business Letters: An Online Business Letter Corpus And Textbooks.* Srinakharinwirot University.

Sriumporn, C. M. (2015). An Analysis of Lexical Bundles in Business English coursebooks by Developing a Thai BELF Corpus From The Nation and Bangkok Post Newspapers. Retrieved from

St John, J. M. (1996). Business is booming: Business English in the 1990s. *English for specific purposes, 15*(1), 3-18. doi:http://dx.doi.org/10.1016/0889-4906(95)00023-2

Stubbs, M. (2007a). An example of frequent English phraseology: Distribution, structures and functions. In R. Facchinetti (Ed.), *Corpus Linguistics 25 years on* (pp. 89-105). Amsterdam: Radopi.

Stubbs, M. (2007b). Quantitative data on multi-word sequences in English: The case of word 'world'. In M. Hoey (Ed.), *Text, Discourse and Corpora: Theory and Analysis* London: Continuum.

Thaweewong, P. (2006). *Genre analysis of English business e-mail correspondence in internal communication between Thais and Germans in profit and non-profit organizations.* (Ph.D. Non-fiction), 2006., Chulalongkorn University. Retrieved from http://search.ebscohost.com/login.aspx?direct=true&db=cat05085a&AN=chu.b1788042&site=eds-live&authtype=ip,uid Available from EBSCOhost cat05085a database.

Trappe, T., & Tullis, G. (2006). *Intelligent Business Coursebook*: Pearson.

a university's webpage on the English for Business Communication Program. (2016). Retrieved from http://dpuic.dpu.ac.th/programs/bachelor/bachelor-of-arts-in-english-for-business-communication-ibec/

Warren, M. (2016). Signalling intertextuality in business emails. *English for specific purposes, 42*, 26-37. doi:http://dx.doi.org/10.1016/j.esp.2015.11.001

Warwick, B. (1992). *Business correspondence by fax: An investigation into frequency of use, discourse structure and linguistic features.* (Master Degree), University of Aston, UK.

Williams, M. (1988). Language taught for meetings and language used in meetings: Is there anything in common? *Applied Linguistics, 9*(1), 45-58.

Wood, D., & Apple, R. (2014). Multiword constructions in first year business and engineering university textbooks and EAP textbooks. *Journal of English for Academic Purposes, 15*, 1-13.

Wray, A. (2002). *Formulaic Language and the Lexicon*. Cambridge: Cambridge University Press.

Xu, B. (2012). An Analysis of English Business Letters from the Perspective of Interpersonal Function. *English Language Teaching, 5*(7), 75-80.

Yamada, H. (1990). Topic management and turn distribution in business meetings: American versus Japanese strategies. *Text-Interdisciplinary Journal for the Study of Discourse, 10*(3), 271-295.

Yue, S., & Wang, X. (2014). Hedges Used in Business Emails: A Corpus Study on the Language Strategy of International Business Communication Online. *Higher Education Studies, 4*(6), 49-57.

APPENDICES

Appendix A

List of Thai Universities which provide textbook samples used in their business English classes

1. Assumption University
2. Bangkok University
3. Burapha University
4. Chulalongkorn University
5. Dhurakij Pundit University
6. Kasetsart University
7. King Mongkut's Institute of Technology Ladkrabang
8. Phranakhon Rajabhat University
9. Rajamangala University of Technology Phra Nakhon
10. Rajamangala University of Technology Thanyaburi
11. Silpakorn University
12. Srinakharinwirot University
13. Suan Dusit Rajabhat University
14. Suan Sunandha Rajabhat University
15. South-east Asia University
16. Thammasat University
17. University of the Thai Chamber of Commerce

Appendix B

Lists of Lexical Bundles in ENRON and TEC

ENRON	TEC
a copy of the	**a copy of the**
a draft of the	**as soon as possible**
as a result of	by the end of
as soon as possible	dear sir or madam
attached is a draft	end of the week
don't want to	* feel free to contact me
feel free to contact	for your e mail
get in touch with	get back to me
have any questions or	give me a call
I don't have	hearing from you soon
I don't know	hesitate to contact me
I don't think	I am attaching a
I look forward to	I am interested in
I would like to	I am writing to
I'd like to	I have attached a
if you have any	**I look forward to**
if you want to	I would be grateful
in light of the	**I would like to**
in the process of	**I'd like to**
is a draft of	I'm writing to
it would be a	if you could send
let me know if	**if you have any**
let me know what	if you need any
look forward to hearing	if you would like
please feel free to	in the near future
please let me know	it was good to
thanks for your message	**let me know if**
thank you for your	**let me know what**
the end of the	let you know that
to hearing from you	**look forward to hearing**
we look forward to	look forward to receiving
we would like to	look forward to seeing
what do you think	not hesitate to contact
you have any questions	* please do not hesitate to
	please let me know
	please could you send
	thanks for your email
	thank you for your
	the end of the
	to hear from you
	to inform you that
	to let you know
	to tell you that

ENRON	TEC
	you could send me
	you have any questions
	you would like to
	you let me know
	we look forward to
	we would be grateful
	we would like to
	will be able to
	would like to know
	would be grateful if

The lexical bundles **in bold** refer to those that occur in both corpora

Appendix C

Examples of sheets given to the raters for lexical bundle functional analysis

Example 1: The rating template for raters' functional analysis

No.	Lexical bundles	Functional Type	Agree	Not sure	Disagree	Suggested type
1.	feel free to contact	D. Permission				
2.	get in touch with	C. Request				
3.	if you want to	1 desire				
4.	in light of the	3. Intangible				
5.	let me know what	3 imperative / E. offer				
6.	we look forward to	F. Expectation				
7.	attached is a draft	A. Topic Introduction/focus				
8.	don't want to	1 desire				
9.	I don't have	5 ability/ effort				
10.	is a draft of	2. Tangible				
11.	we would like to	1 desire / 4 intention/ prediction				
12.	a copy of the	2. Tangible				
13.	in the process of	3. Intangible				
14.	it would be a	4 intention/ prediction				
15.	look forward to hearing	F. Expectation				
16.	thanks for your message	A. Politeness				

Example 2: The lists of lexical bundles' concordance lines

Lexical bundle	No.	Concordance line	File no.
we would like to	1	the Judge and his legal staff tomorrow. While **we would like to** convince the Judge that the	124868.txt
	2	current power marketing certificate is an entity which **we would like to** move to being a direct	162662.txt
	4	is deregulated and structured pretty much the way **we would like to** see it done in both	174266.txt
	5	cases whose accidents happened before Enron bought GSP. **We would like to** bring them back to do	230724.txt
	6	planning trips to California and New Mexico that **we would like to** combine and would like very	54928.txt
	7	is divided into 4 sessions, each with two papers. **We would like to** invite you to be a	119477.txt
	8	in Houston on Friday 1/14/00 from 8:00 AM until Noon. **We would like to** have dinner at a Houston	173338.txt
	9	's PRC Meeting. It's that time again... **We would like to** schedule PR's PRC meeting	175307.txt
	10	April 19th at the Gulf South (Koch) Building. We would like to present a speaker who can	175713.txt
	11	their feedback, it would be helpful. Once done, **we would like to** schedule a call for next	127099.txt
	12	a call for next Thursday at 3:30 pm, since **we would like to** do it when Intellibridge is	127099.txt
a copy of the	1	adership and shareholdervalue. Kevin 213-926-2626 PS - May I have **a copy of the** position paper that was circulated	7932.txt
	2	U.S. Ch.11. We have still not seen **a copy of the** petition presented to the Court,	130149.txt
	3	administratively. We are in the process of obtaining **a copy of the** letter, which has not yet	136023.txt
	4	for an RTO meeting and expect to obtain **a copy of the** draft electricity language implementing the	136378.txt
	5	seeking input in this manner. I am attaching **a copy of the** Memorandum of Understanding between the	173956.txt
	6	, I did know about the conference. Ken has **a copy of the** study and made reference to	175198.txt
	7	Ken on this, but I did send you **a copy of the** report a couple of weeks	175198.txt
	8	can do an assignment agreement, once I receive **a copy of the** current agreement. It also depends	230395.txt
	9	is confidential under Section 19.3.2 of the CalPX tariff. **A copy of the** request is attached below. The	234796.txt
	10	on humanitarian grounds. I shall fax to you **a copy of the** decision. Thanks. Vince << File: Krishna'	54536.txt
	13	recent conversation with my brother, George, I attach **a copy of the** short version?of my resume.?	175142.txt

VITA

Aroonrung Siricharoen is an MA student in the English as an International Language Program of Graduate School at Chulalongkorn University. She holds a Bachelor of Arts degree from Chulalongkorn University. She has published an article on linguistic landscape in a special issue of MANUSYA Journal of Humanities in 2016 and gave a presentation, part of her thesis, in the Doing Research in Applied Linguistics 3 / 19th English in South-East Asia Conference 2017.