A Corpus-based Comparative Study of Lexical Bundles in Authentic and Textbook English Business Emails

Aroonrung Siricharoen
English as an International Language Program, Graduate School, Chulalongkorn University, Thailand
aroonrungsiri@yahoo.com

Raksangob Wijitsopon
Faculty of Arts, Chulalongkorn University, Thailand
raksangob.w@chula.ac.th

Abstract

This study investigates lexical bundle types found in authentic English business emails and sample emails in business English textbooks in order to identify their similarities and differences. The data employed in the study were sample emails from 77 business English textbooks and emails sourced 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. Findings show that structural categories of lexical bundles in textbooks and those in authentic emails are generally similar while functional categories are noticeably different. Although there are more lexical bundle types in the textbook email corpus, most of them actually belong to the same categories, pointing to a limited range of functional categories of lexical bundles presented in business English textbooks. This leads to a major observation that forms of expressions tend to outweigh their functions in the way textbooks present emails for pedagogical purposes. It is therefore suggested that the functional dimension of linguistic expressions be given more attention in business English teaching.

Keywords: Lexical bundles, Emails, Business English, Textbooks, Corpus linguistics

Introduction

Lexical bundles are multi-word sequences that recur in a certain text type. The concept was introduced by Biber *et al*. (1999) as "recurrent expressions, regardless of their idiomaticity and regardless of their structural status" (p.990). In other words, lexical bundles may not necessarily express idiomatic meanings or take the form of a specific kind of phrasal or clausal unit but must occur repeatedly in a particular text type. The emphasis on the frequency and distribution patterns of lexical bundles makes it an operationalizable concept for identifying and differentiating text types (Conrad & Biber, 2005).

Because lexical bundles are linguistic elements common in a certain discourse, they have been used to explain characteristics of various text registers. For example, Biber *et al*. (1999) compared lexical bundles found in academic prose and those in conversation and found that academic writing mainly consists of nominal and prepositional phrases while conversation tends to feature verbal phrases or clauses. Biber, Conrad and Cortes (2004) extended the scope of the previous research by examining textbooks and classroom teaching and also developed a framework for both structural and functional analysis of lexical bundles.

Because lexical bundles serve as discourse building blocks essential to communicative acts (Biber, 2009), they are regarded as central to language teaching and learning, specifically as an index to determine students' English language proficiency and

facilitate their language learning. As Wray (2002) argues, multi-word expressions can help learners produce the language fluently and naturally. Lexical bundles have therefore been widely employed in applied linguistic studies on language and education, including those that focus on disciplinary variations (e.g. Cortes, 2004; Hyland, 2008a), native and non-native English writing (e.g. Chen and Baker, 2010; Leelasetakul, 2014; Lie, 2013) or written texts produced by language learners and expert writers (e.g. Öztürk and Köse, (2016) and Pan, Reppen, and Biber, (2016)).

Some scholars have employed the concept of lexical bundles in a comparison between real-life language use and language samples presented in textbooks and other teaching materials, including Chen (2008), Sriumporn (2015) and Wood and Apple (2014). This practice is relatively common in the area of English for Specific Purposes (ESP) (Nesi, 2012) since lexical bundles are characteristic of text registers and domains. They can, therefore, reveal distinctive characteristics of language in real use and the one presented in textbooks.

Textbooks for teaching business English have been investigated in the past few decades. One of the pioneer studies is the work by Williams (1988), which looked at language use in meetings by comparing that presented in textbooks and that in the real context. Her findings revealed how the two discourses, albeit from the same genre, were considerably different. For over thirty years, the comparative study of business English textbooks has been conducted on a number of text types, such as business letters (e.g. Sriumporn 2015) and collaborative writing (e.g. Bremner 2010).

However, to the best of our knowledge, there has not been any comparative study on English business emails in textbooks and those used in authentic online communication. Although a number of scholars have increasingly paid attention to emails in the business world as they have now become a standard communicative method, most of these studies tend to focus on stylistic features of business emails.

To bridge the gap in research on business discourse and instruction, the present study aims to compare authentic business emails and email samples presented in business English textbooks, focusing on structural and functional properties of lexical bundles. It is believed that an examination of lexical bundles in authentic and textbook emails can show not only important features of business email discourse in each register but also benefit the development of business English pedagogy in terms of email writing.

To this end, we first give an overview of previous research on business English textbooks and business emails, followed by methodological descriptions. Then, results from structural and functional analyses are reported and discussed before a conclusion is given on the significance and pedagogical implications of the study.

Business English textbooks

Research into textbooks has often found discrepancies between contents in English for business purposes textbooks and English in real business contexts (e.g. Chan, 2009; Donna, 2000; Harwood, 2010). Williams (1988), for instance, found a significant discrepancy between the use of authentic language and the language taught in the textbooks for teaching English use at meetings. 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 vocabulary in business textbooks was limitedly represented. A decade later, Angouri's (2010) contrastive study on business meeting language taught in the textbooks and that in the real business meeting setting still points to a similar direction, i.e. language prescribed in the business English textbooks rarely correlated with the language in real business use. Sinturat (2010) also examined lexical phrases frequently 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.

Based on the aforementioned studies, although there seems to be discrepancies between the business English presented in textbooks and that used in real business settings, the practice of comparing teaching materials with authentic usage of different business genres can direct our attention to communicative essentials in materials design as well as enable us to identify strengths and weaknesses of contents in teaching materials.

Business emails

One of the central business genres that deserves urgent attention in the field of ESP is business emails. This is because business emails are central to the corporate world (Angell & Heslop, 1994; Baron, 1998) but little has been researched on the way email writing is taught via textbooks. In other words, a comparison between business emails in textbooks and those in real-life business settings in the ESP literature remains very scarce.

In spite of the existence of email for some decades, characteristics of emails, either in general or business-related, seem to vary. Although emails are written messages, they function more like a spoken register owing to their real-time messaging process, unlike traditional written communication tools such as letters or memos. Baron (1998), for example, used three different models to analyze the 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 were on telephone calls or having face-to-face conversation, as reflected by no editing and a more informal style of writing.

In early years of research into business emails (late 1990s up to early 2000s), linguistic features of authentic emails received particular attention. For example, Gains (1999) examined emails from two different domains, academia and commerce, and found that academic writers incorporated some discoursal features of conversations, such as 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 the use of grammatical structure and punctuation. In contrast to Gains' (1999), Gimenez (2000) found that email has its own style of business writing. Based on his study, business emails tend to give less attention to punctuation, capitalization, and spelling. Gimenez (2000) also identified the hybridity of spoken and written language in business emails. This leads to his 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, 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) are also common in emails.

After the year 2000, genre analysis (Bhatia 1993) and Critical Discourse Analysis (CDA) have often been employed as an analytical framework to investigate how business people write emails. For example, Danielewicz-Betz (2016) investigated over 4,000 business emails and found that some linguistic patterns are used more in certain groups. For instance, the conditional sentence is more frequently used as a request form between co-workers or in external communication than by superiors to subordinates. Thus, it can be said that business emails can vary among colleagues within the same company, depending on their positions.

While there have been quite a number of studies on the language of emails, as illustrated above, only a small number of studies have looked at business emails in relation to the pedagogical context. As Evans (2012 argued, although there are a number of studies on the language of business emails, they tend to be motivated by general interest, rather than by pedagogical considerations. He therefore analyzed functions, structures and characteristics of business emails and found that they possess qualities similar to those of the spoken language. Based on the findings, he proposed a simulation-based approach to business email writing lessons that integrates all four fundamental language skills, i.e. writing, reading, speaking and listening.

Based on these previous studies, it seems that while emails are an important type of business correspondence, research on business email instruction remains scarce. The present study, therefore, aims to fill this gap through an examination of lexical bundles identified in business emails in business English textbooks and authentic business ones. Here are our research questions:

1. How are forms of lexical bundles in authentic emails and those in business English textbooks similar and different?

2. How are functions of lexical bundles in authentic emails and those in business English textbooks similar and different?

Methodology

To answer the above research questions, we adopted a corpus-based method. Four major aspects were involved in the methodology of the study: corpus building, software used in the study, lexical bundle extraction, and data analysis.

Corpus Building

Two corpora were used in the study. The first one is a Textbook Email Corpus (henceforth TEC), which contains 69,902 tokens with 751 business email samples taken from 77 business English textbooks. Of these 77 textbooks, 16 were identified as being used at universities in Thailand by business English lecturers at different institutions. The other corpus is a corpus of authentic emails, drawn from the Enron database of the 'UC Berkeley Enron Email Analysis Project' provided by University of California-Berkeley (Berkeley, n.d.) (henceforth ENRON). ENRON contains 1,061 texts of email messages from the Enron Corporation[1], with a size of 277,919 tokens.

The emails in the two corpora were categorized into five categories, based on the categorization system 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

Software

In order to extract lexical bundles from each corpus, we used the corpus tool AntConc 3.4.4w (Anthony, 2004), which is a freeware made accessible online and has been used in a number of lexical bundles, e.g. Leelasethakul (2014) and Jalali and Moini (2014). The software serves as a multi-platform tool for conducting research in corpus linguistics and data-driven learning. We used the 'Clusters/N-Grams' tool[2] on AntConc for generating a list of lexical bundles in both corpora. We 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.

Lexical Bundle Extraction

There are 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 based on the requirements of the software AntConc. These thresholds are discussed in turn below.

The first threshold deals with the target length of extracted lexical bundles. The present study adopts the four-word length since three-word bundles were found to be general bundles and, in many cases, overlap with one another, e.g. *as the result* and *the result of*, while the bundles longer than four words would reduce the number of bundles to be studied. This choice of length corresponds to many previous studies that show four-word lexical bundles are optimal for the purpose of their studies (e.g. Biber et al., 2004; Cortes 2002, 2013; Hyland, 2008b).

The second criterion is distributional range. This concerns the number of texts required for a lexical bundle to appear. The designation of distributional range is rather arbitrary, ranging from occurrences in three to five texts (e.g. Biber & Barbieri, 2007) to 10% of the total number of texts in a corpus (e.g. Hyland, 2008a, 2008b). This is because the designation of distribution range depends heavily on the characteristics of a given corpus. In the present study, a lexical bundle to be analyzed was set to occur in at least 1% of the total email samples in each corpus. Although the cut-off distribution range set at 1% may appear small, when compared with the distributional range adopted in previous studies such as Biber et al. (2004), we considered this the best possible threshold for the present study because our corpora are smaller than those studies but specific in terms of text type and domain, i.e. business emails.

Based on the two criteria, four-word lexical bundles that occur in at least 1% of the texts in each corpus, a list of lexical bundles in each corpus was generated. When examining both 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 a 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 still included when counting if not all of the instances are part of the same single phrase. For example, *please let me know* and *let me know if* were not treated as the same lexical bundle even though they can form the five-word bundle *please let me know if*. This is because not all instances of *please let me know* are followed by the conditional *if*; some are followed by *that*, making up the phrase *please let me know that*.

Analytical Frameworks

After obtaining lists of lexical bundles from the two corpora and excluding the bundles that did not meet the aforementioned criteria, we categorized the rest of the lexical bundles into structural and functional categories on the basis of Biber et al.'s (2004) and Biber's (2006) frameworks.

Structural Categories of Lexical Bundles

(1) 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 incorporate 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 that starts a dependent clause (e.g. *what I want to*), or a complementizer or subordinator that starts at the beginning of the dependent clause bundles (e.g. *to be able to, if you look at*).

(3) NP-/PP-based bundles incorporate either a nominal or prepositional phrase. In contrast to the above two categories, which contain a clausal fragment, a lexical bundle in this category is phrasal. For example, a postmodifier 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*).

Functional Categories of Lexical Bundles

While the present study adopted the functional analytical frameworks of Biber et al. (2004) and Biber (2006), it must be noted that some functional categories were added later during the extraction and analysis of the lexical bundle, when it had been found that the pre-existing labels could not sufficiently cover the items that occur in the target email corpora. Four main functional categories of lexical bundles applied in this study are:

(1) Stance expressions, which can be divided into five subtypes:

- **Epistemic** bundles expresses the certainty of the text producer about the idea following the bundle, e.g. *I don't know if.*
- **Desire** bundles are personal expressions of desire, e.g. *do you want to.*
- **Obligation/directive** bundles are expressions of obligations or directives, e.g. *you don't have to*.
- **Intention/prediction** bundles express the text producer's intention to perform an

action in the future, e.g. *we're going to do*.

- **Ability/effort** bundles concern the ability and effort of the text producer, e.g. *to be able to*.

(2) Discourse organizers are lexical bundles that assist in organization of discourse, which include three groups:

- **Topic introduction/focus** bundles signal to the reader/audience what is going to be discussed, e.g. *if you look at*.
- **Topic elaboration/clarification** bundles are used when a speaker/writer wants to elaborate, clarify, or compare and contrast an issue, e.g. *at the same time*, when it is used to signal comparison and contrast of two things or events.
- **Condition** bundles signal conditional relations between two propositions, e.g. *if you do not.*

(3) Referential expressions serve as references to either physical or abstract units, which are divided into:

- **Identification/focus** bundles channels into the noun phrase following the lexical bundle, e.g. *those of you who* specifies the subgroup of addressed participants.
- **Imprecision** bundles make reference to entities in an imprecise manner, e.g. *or something like that*.
- **Specification of attribute** bundles indicate "specific attributes of the following

head noun" (Biber, 2006, p.145) or frame the noun phrase tangibly or intangibly, e.g. *have a lot of*, *the dean of the* and *in terms of the*.

- **Time/place/text reference** bundles refer to time, place, or text within the data, e.g. *the end of each, in the college of, as shown in Figure*

- **Multi-functional reference** bundles can refer to more than one pragmatic function of referential bundles, e.g. *at the end of* can function either as a place reference if it precedes a place or a time reference if a specific time is given after the bundle.

(4) Special function bundles have so far been identified only in the spoken register (cf. Biber et al., 2004 and Biber, 2006). The sub-categories in this group vary from one study to another, depending on the data. In other words, the data-driven approach leads to some ad hoc categories, which might not be identified in all studies of lexical bundles. In the present study, this functional group of lexical bundles covers the remaining lexical bundle functions that cannot be categorized according to the aforementioned categories. There are six sub-categories in the *Special function* group:

- **Politeness** bundles are expressions that serve politeness strategies *thank you for your* and *thanks for your message*.

- **Inquiry** bundles serve to enquire, e.g. *what do you think.*

- **Request** bundles are used to make a request, e.g. *please let me know.*

- **Opening up for further communication** bundles serve as an offer for further communication, e.g. *you have any questions.*

- **Expectation** bundles express the speaker/ writer's expectation, e.g. *I look forward to.*

- **Hybrid** bundles can perform multiple functions, e.g. *let me know if* can serve as a directive on condition, as in "*Let me know if* there is a problem with this request", or it can offer help to the email reader, as in "Please *let me know if* you need any help such as adding months". Also, it can perform the *Opening up for further communication* function, as in "Please *let me know if* you have any questions or comments".

It should be pointed out that the first two sub-categories are also found in Biber et al. (2004) while the remaining four were newly created to specifically explain the data in this particular study, i.e. business emails. In creating new sub-categories, we relied not only on the meanings of lexical bundles themselves but also their textual positions and types of emails. For example, the *Opening up for further communication* group is coined to include lexical bundles that serve as an offer for further communication which usually comes after the main messages have already been delivered.

Along with the above analytical framework, to ensure the validity of the researcher's functional categorization of lexical bundles, two raters (one was a native English speaker, who has been teaching business English at college for no less than five years, and the other was an advertising and marketing manager at a company) were asked to examine all lexical bundles identified in the two corpora and rate if they agree or disagree with our categorization of lexical bundles' functions. This leads to 73% similarity in the functional identification between the two raters and the researchers. Although the number might be considered not very high from a general point of view, this is because functional categorization is rather flexible and heavily context-dependent. It should be noted that the raters also provided some helpful suggestions on the category labels and definitions, which were taken into account when developing new sub-categories of lexical bundles in the *Special function* group.

It is also noteworthy that the fuzzy nature of functional categories of lexical bundles has an implication for the way lexical bundles were counted and calculated in the present study. In order to calculate and compare lexical bundles from each corpus systematically, we

labelled a lexical bundle with only one function, using frequency information as a criterion. Thereby, when a lexical bundle has different uses, its most frequent function was interpreted as the dominant function of that lexical bundle.

Results and Discussion

In this section, only numbers and percentages of types of lexical bundles in each category are reported, like Wei and Lei (2011), even though it is also possible to compare lexical bundles by token, i.e. by looking at frequency of an individual lexical bundle. This is because, based on our pilot study, discrepancy in the frequency of each lexical bundle seems to mainly reflect different contents of emails in the two corpora. To illustrate, common lexical bundles *I don't think* and *I don't know* in ENRON, which occur 18 times (0.65 per 10,000 words), do not occur at all in TEC; on the other hand, the bundle *I would be grateful*, which occurs 18 times (2.58 per 10,000 words) in TEC, is not found in ENRON. The reason is that ENRON contains internal correspondences that involve corporate discussion on certain topics while TEC does not obviously show this context. Such distinction in individual frequency occurs with a number of lexical bundles in the two corpora (see Appendices A-B). The present study, therefore, focuses on the degree of variety of lexical bundle types in the two corpora by percentage, which more clearly sheds light on general characteristics of emails in textbooks and in authentic uses.

Based on the threshold above, a total of 34 lexical bundle types were derived from ENRON and 53 from TEC, which means that TEC contains more varied lexical bundles than ENRON. This is probably a result of the pedagogic purposes of textbooks which aim to present an array of expressions to learners. These lexical bundles were then put in different categories as spelled out above. In this section, the comparative findings are reported in terms of the structural and functional categories.

Structural Categories

As can be seen from Table 4.1 below, the distribution patterns of structural categories in ENRON and TEC are in the same direction in that VP-based bundles are the most frequent group. This overall distribution pattern 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 in the list. Also, in both corpora there are frequent uses of personal pronouns and verb phrases in the active voice, which are associated with the features of conversational register, e.g. *you have any questions* and *I look forward to*. These in turn point to the observation made in previous studies that emails are close to spoken registers (e.g. Baron, 1998; Gimenez, 2000; Danet, 2002).

Although the two corpora share similar tendencies regarding the predominance of the *VP-based* and *Dependent clause* lexical bundles, a closer look at individual lexical bundles in each category reveals that both corpora are structurally different in some aspects. First, ENRON contains the use of contractions in both *VP-based* and *Dependent clause* categories, e.g. *I don't have* and *don't want to*, whereas TEC, despite being larger in the number of lexical bundle types, does not contain any contraction in the *VP-based* but only in the *Dependent clause* category, e.g. *I'm writing to* (see Appendix B). The more frequent occurrences of contracted forms in ENRON corresponds to what Mallon and Oppenheim (2002) have found in that contractions are characteristic of emails, whether personal, business, or non-business emails. It can thus be observed that business English textbooks do not tend to feature contraction even though it is another characteristic feature of emails.

Furthermore, it can be seen that ENRON contains a similar proportion of *Dependent clause* bundles, 26.5%, to that of the *NP-/PP-based* category. According to Biber et al. (1999), the prepositional phrase expressions containing *of* are the largest proportional type in academic prose and hence it is interesting that this group of *'of'* lexical bundles is also found

in the ENRON email corpus, namely *in light of the, as a result of,* and *in the process of*. The occurrence of these *NP-/PP-based* bundles in the authentic email corpus suggests that business emails also contain phrases often associated with such highly formal writing style as that of academic writing, thereby demonstrating the relatively hybrid nature of business emails as well as distancing the genre from conversational discourse.

On the other hand, the *NP-/PP-based* bundles are particularly low in TEC, constituting merely 17% of the four-word lexical bundles in the corpus. Moreover, unlike those found in the ENRON, the above three *'of'* lexical bundles were not found at all on the list of lexical bundles in TEC and its *NP-/PP-based* category is mainly made up of temporal expressions, i.e. *end of the week, the end of the, by the end of, in the near future,* and *as soon as possible*. Based on these structural categories and their distribution patterns, it seems that email language use in textbooks is structurally closer to the spoken language than that used in real corporate emails.

Table 1: Structural categories and their distribution in ENRON and TEC

Category	Structural sub-category	No. of types in ENRON	Per cent	No. of types in TEC	Per cent
VP-based	**(connector +) 1st/2nd person pronoun + VP fragment**, e.g. *you have any questions*	6	47% (16)	10	49% (26)
	(connector +) 3rd person pronoun/ noun + VP fragment, e.g. *it would be a*	2		1	
	Discourse marker + VP fragment, e.g. *I mean you know*	0		0	
	Verb phrase with active verb, e.g. *please let me know*	7		14	
	Verb phrase with passive verb, e.g. *is based on the*	0		0	
	Yes-no question fragments, e.g. *are you going to*	0		1	
	Wh-question fragments, e.g. *what do you think*	1		0	
Dependent clause	**(connector +) 1st/2nd person pronoun + dependent clause fragment**, e.g. *I would like to*	3	26.5% (9)	6	34% (18)
	Wh-clause fragments, e.g. *let me know what*	1		1	
	If-clause fragments, e.g. *if you have any*	3		6	
	(Verb/Adjective+) to-clause fragment, e.g. *please feel free to*	2		5	
	That-clause fragments, e.g. *that there is a*	0		0	

Category	Structural sub-category	No. of types in ENRON	Per cent	No. of types in TEC	Per cent
NP/PP-based	**Noun phrase with *of*-phrase fragment**, e.g. *the end of the*	3	26.5% (9)	3	17% (9)
	Noun phrase with other post-modifier fragment, e.g. *thanks for your message*	1		1	
	Other noun phrase expressions, e.g. *a little bit more*	0		1	
	Prepositional phrase expressions, e.g. *as a result of*	4		3	
	Comparative expressions, e.g. *as soon as possible*	1		1	
Total		34	100%	53	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)

Functional Categories

The functional analysis reveals one obvious similarity between lexical bundle categories in TEC and ENRON; that is, as can be seen in Table 2 below, the *Special functions* category predominates in both lists, with 41% and 49% of the total lexical bundle types in ENRON and TEC, respectively. This reflects that business emails, whether in authentic use or in textbooks, consist of formulaic expressions used pragmatically in communicative acts, such as *please let me know*, *I look forward to* and *thank you for your*, etc. This points to the highly interactive nature of email discourse.

Notwithstanding this similarity, TEC and ENRON are quantitatively different in the proportions of other functional groups. While TEC contains more lexical bundle types than ENRON, which may be attributed to the pedagogical purpose of the TEC as observed above, the larger number does not entail an even distribution of lexical bundles across functional categories. As demonstrated in Figure 1 below, the *Discourse organizers* category is the second largest in TEC but least found in ENRON while *Stance expressions* and *Referential expressions* share the place at the bottom in TEC but are the second and third largest groups in ENRON.

Table 2: Functional categories of lexical bundles in ENRON and TEC

Functional Category	Functional Sub-category	No. of types	Percent (No.)	No. of types	Percent (No.)
Stance expression	**Epistemic stance**, e.g. *I don't think*	2	29.5% (10)	0	13% (7)
	Desire, e.g. *I would like to*	5		6	

Functional Category	Functional Sub-category	No. of types	Percent (No.)	No. of types	Percent (No.)
	Obligation/directive, e.g. *let me know what*	1		0	
	Intention/prediction, e.g. *it would be a*	1		0	
	Ability/effort, e.g. *I don't have*	1		1	
Discourse organizer	**Topic introduction/focus**, e.g. *attached is a draft*	1	6% (2)	10	25% (13)
	Topic elaboration/clarification, e.g. *I mean you know*	0		0	
	Conditions, e.g. *if you have any*	1		3	
Referential expression	**Identification/focus**,e.g. *and this is a*	0	23.5% (8)	0	13% (7)
	Imprecision, e.g. *or something like that*	0		0	
	Attribute Specification: Quantity specification, e.g. *have a lot of*	0		0	
	Attribute Specification: Tangible framing attributes, e.g. *a draft of the*	3		2	
	Attribute Specification: Intangible framing attributes, e.g. *as a result of*	3		0	
	Place reference (e.g. *the United States and, of the United States*)	0		0	
	Time reference, e.g. *as soon as possible*	1		5	
	Text reference, e.g. *shown in figure N*	0		0	
	Multi-functional reference, e.g. *the end of the*	1		0	
Special function	**Politeness**, e.g. *thank you for your*	2	41% (14)	5	49% (26)
	Inquiry, e.g. *what do you think*	1		0	
	Request, e.g. *please let me know*	2		8	
	Opening up for further communication, e.g. *you have any questions*	4		5	
	Expectation, e.g. *I look forward to*	4		7	
	Hybrid, e.g. *let me know if*	1		1	

Functional Category	Functional Sub-category	No. of types	Percent (No.)	No. of types	Percent (No.)
Total		34	100%	53	100%

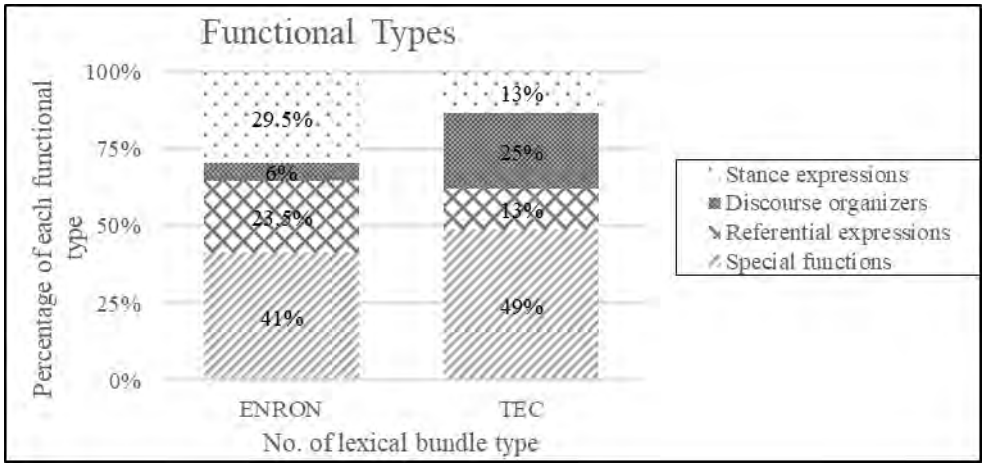

Figure 1: Proportion of functional category in ENRON and TEC

This quantitative difference points to an important qualitative difference between authentic and taught email samples. First, in terms of the *Discourse organizer* category, which features remarkably in TEC but least in ENRON, the most obvious case is the lexical bundles in the *Topic introduction/focus* sub-category. TEC contains 10 lexical bundle types, which cover 77% of all *Discourse organizing* bundles in TEC. Of these 10 *Topic introduction* lexical bundles, six are those that are found at the initial position of the clause or sentence: *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*. This correlates with the findings from the structural analysis above that verb phrases in the active voice are characteristic of TEC. 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*.

On the other hand, ENRON has only one bundle in this sub-category, i.e. *attached is a draft*, which takes the passive voice form and does not contain the first-person pronoun 'I'. This perhaps shows that textbooks tend to highlight the simple syntactic structure of "Subject (mostly personal pronouns) + verb + object", rather than a complex one like the "inverted passive verb + subject" as in *attached is a draft* in ENRON, though it is likely to occur in authentic emails as suggested by ENRON.

The *Stance expressions* category also points to a conspicuous discrepancy between TEC and ENRON as it is one of the smallest functional categories in TEC but the second largest in ENRON. With only seven lexical bundle types in this functional category in TEC, six are in the same one category, i.e. *Desire* expression, while ENRON sees various *Stance* lexical bundles in all sub-categories. It can therefore be said that textbooks concentrate solely on the *Desire* function while other stance bundles, which are likely to occur in real use as suggested by their frequencies in ENRON, are missing.

The other small functional category in TEC but predominant in ENRON is the *Referential expressions*. ENRON features lexical bundles that refer to tangible and intangible framing attributes, each of which consists of three lexical bundle types: *a draft of the, a copy of the,* and *is a draft of* in the former group and *as a result of, in the process of,* and *in light of the* in the latter. Based on this, it can be seen that authentic emails feature references to documentation and explanation of the business, as illustrated below:

Tangible Attribute Lexical Nundles

1) Attached please find ***a draft of the*** financial schedules prepared by Enron Corp. (ENRON_173938)

2) 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)

Intangible Framing Attribute Lexical Bundles

1) We regret any difficulties that have arisen ***as a result of*** your making inquiries on our behalf. (ENRON_174320)

2) Subject: Market Stack Vince, Henwood is ***in the process of*** developing a new product that, for lack of a (ENRON_54559)

3) ExIm and other funding organizations ***in light of the*** change in emphasis in our business. (ENRON_175772)

By contrast, TEC email samples exhibit textbooks' emphasis on reference to time as the largest type of referential bundle in TEC is *Time reference*, accounting for five of the seven types: *as soon as possible, by the end of, the end of the, end of the week,* and *in the near future*. This suggests a tendency in which business textbooks highlight time reference in taught emails.

Overall, it can be seen that although ENRON contains fewer lexical bundle types than TEC (34 vs. 53), it displays a wider range of functional sub-categories of lexical bundles than those in TEC. This variety reflects that the business email contains an amalgam of spoken and written languages, as reflected by distributions of lexical bundles in various functional categories as *Special function*, *Stance* and *Referential* expressions as discussed above.

It may be argued that the wider range of functional categories in ENRON is attributed to more varied uses of emails in real life than in teaching, but it cannot be denied that an abundance of lexical bundle types in TEC comes from a large number of lexical bundles that actually express similar meanings, such as *to let you know*, *to tell you that* and *to inform you that*, resulting in a large variety of forms but small range of functional categories. In other words, in TEC some functional categories are particularly overwhelmed with various lexical bundle types while others contain very few or even no bundles. In fact, TEC exhibits fewer functional subcategories than ENRON in every functional category, except *Discourse organizers*, which can be taken to indicate a strong emphasis on discourse organizing expressions in English instruction, an issue that has also been observed in previous studies of multi-word expressions in the EFL context (e.g. Chen and Baker, 2010; Wijitsopon, 2019).

The findings here may be seen as textbooks' tendency to over-present forms and under-present functions of linguistic units in business English even though the textbook corpus here consists of email samples taken from as many as 77 textbooks. Of course, this is more or less due to the differing purposes of the texts in the two corpora, one being mainly pedagogical and the other professionally communicative. Also, the fact that the email samples in TEC consist of both internal and external communication may lead to the need to introduce longer and more formal expressions in contextualizing a topic, compared with dominantly internal emails in the ENRON corpus. Nevertheless, this discrepancy in turn can be taken to hint at the fundamental importance of non-verbal contextual factors in business correspondence, such as internal/ external communication and degree of formality, which should be explicitly integrated with the teaching of verbal expressions in business English materials.

Conclusion

While previous research on business English textbooks tends to highlight that textbooks insufficiently represent the language in real business use (e.g. Angouri, 2010; Nelson, 2000; Williams, 1988), the present study has found that, on the quantitative plane, textbooks manage to capture an essential characteristic of email, i.e. it is highly spoken and interactive. However, further qualitative investigation also reveals different emphases in each corpus. Business emails in ENRON are less formal than those in TEC, both on the structural and functional basis. Also, although TEC contains more varied types of lexical bundles, in many cases they denote similar meanings or perform similar functions. When a functional perspective is applied, it is found the results correspond to previous studies in that some functions often used in authentic emails are under-represented.

However, because the present study only looks at types of lexical bundles, the similarities and differences listed here mainly address the macro level of email characteristics in both corpora. Further studies that consider tokens of lexical bundles can shed light on how real-life and taught emails correspond to and differ from each other at the level of individual expressions. This might enable us to develop a detailed list of lexical bundles that should be taught when teaching business emails.

Nevertheless, the present study has some pedagogical implications. First of all, the study lends support to lexis-based teaching approaches, e.g. the lexical approach (Lewis, 2000) and genre-based approach (Dirgeyasa, 2016), since it focuses on employing lexical bundles in business email writing. As they are recognizable as frequently occurring in business communication, they would be useful for students to learn. Also, they are linguistic expressions that can be studied in terms of both form and function and thereby can help students perform effectively or naturally in their business email writing. On a general plane, it is recommended that 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 for communicative purposes.

Endnotes

[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.

[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.

About the Authors

***Aroonrung Siricharoen*:** an MA student in the English as an International Language (EIL) Program of Graduate School at Chulalongkorn University, Bangkok, Thailand. Her research interests lie in business English and corpus linguistics.

***Raksangob Wijitsopon*:** an Associate Professor at Department of English, Faculty of Arts, Chulalongkorn University, Bangkok, Thailand. Her research interests include corpus linguistics, stylistics and discourse analysis.

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.

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

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

Biber, D. (2006). *University language: A corpus-based study of spoken and written registers*. Amsterdam: 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: 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.

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

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), 230-249.

Conrad, S., & Biber, D. (2005). The frequency and use of lexical bundles in conversation and academic prose. *Lexicographica: International Annual for Lexicography*, *20*, 56-71.

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

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

Danielewicz-Betz, A. (2016). *Communicating in digital age corporations*. Munich: Springer.

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.

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

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

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

Jalali, Z. and Moini, M. (2014). Structure of lexical bundles in introduction section of medical research articles. *Procedia: Social and Behavioral Sciences*, *98*(6), 719-726.

Leelasetakul, M. (2014). Exploring lexical bundles in Thai EFL learners' writings. *Thoughts*, 101-125.

Lewis, M. (2000). *Teaching collocation: Further developments in the lexical approach*. Hove: LTP.

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.

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

Nelson, M. B. (2000). *A corpus-based study of the lexis of business English and business English teaching materials.* Manchester: University of Manchester Press.

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

Öztürk, Y., & Köse, G. D. (2016). Turkish and native English academic writers' use of lexical bundles. *Journal of Language and Linguistic Studies, 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

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: https://www.academia.edu/1775265/AN_ANALYSIS_OF_LEXICAL_BUNDLES_IN_BUSINESS_ENGLISH_COURSEBOOKS

Wei, Y. & Lei, L. (2011). Lexical bundles in the academic writing of advanced Chinese EFL learners. *RELC*, *42*(2), 155–166.

Wijitsopon, R. (2019) Key multi-word expressions in Thai learner English argumentative essays. *Asian EFL Journal, 23*(6.1*)*, 115-141.

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.

Appendices

Appendix A
Structural categories of lexical bundles in ENRON

Category	Sub-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	

Category	Sub-category and relevant lexical bundle	No. of types	Percentage (No.)
	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%

Appendix B

Structural categories of lexical bundles in TEC

Category	Sub-category	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	

Category	Sub-category	No. of types	Percentage (No.)
	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	

Category	Sub-category	No. of types	Percentage (No.)
	Comparative expressions *as soon as possible*$_{(30)}$	1	
Total		53	100%

Appendix C
Functional categories of lexical bundles in ENRON

Category	Sub-category and relevant lexical bundle	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	

Category	Sub-category and relevant lexical bundle	No. of types	Percentage (No.)
	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%

Appendix D
Functional categories of lexical bundles in TEC

Category	Sub-category and relevant lexical bundle	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	

Category	Sub-category and relevant lexical bundle	No. of types	Percentage (No.)
	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 your 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	

Category	Sub-category and relevant lexical bundle	No. of types	Percentage (No.)
	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%