Lexical bundles in authentic and business English textbook emails:
A case study of informal business emails

Aroonrung Siricharoen and Raksangob Wijitsopon
Chulalongkorn University

Abstract

The present study aims to examine lexical bundles, i.e. uninterrupted lexical sequences extracted automatically by corpus software, found in authentic English business emails and those in business English textbooks. Although emails are commonly used in the business context (Evans, 2010), little research has paid attention to this text type. To fill the gap, a combined quantitative and qualitative analysis of lexical bundles is conducted on corpora of authentic business emails and textbook email samples, with a focus on the category of informal business emails. The findings show that lexical bundles identified in this type of emails are central around those in the *Special functions*, which concern communicative acts such as request, expectation expression, offer, etc. Some pedagogical implications from this set of findings are discussed.

1. Introduction

Email is a widely recognized communication tool in many settings, from academia to business corporations (Sabater, Turney, & Fleta, 2008). The latter type is the focus of the present study. Based on previous studies (i.e. Baron, 1998; Gains, 1999), business emails have played an important role in the corporate world since 1990s. A large-scale survey conducted in Hong Kong reveals that emails are the most frequently used text type in the organizations (Evans, 2010). This suggests that business emails should receive attention in business English curriculum. However, little research has been done to investigate how this important business genre is taught to students in business English class even though there have been quite a number of studies on English business emails, which center around linguistic and discoursal features of business emails, such as hedging devices (Yue & Wang, 2014), greetings and closing (Waldvogel, 2007), genre analysis (Thaweewong, 2006), etc. This seems to contrast with a general tendency of research in Business English over the past few decades, which have focused on the question of a correlation between real and textbook business English.

With regards to the issue of representativeness of language in business English textbooks, the discussion features two approaches: linguistic-oriented and register-oriented. The former seeks to discover to what extent linguistic features in the teaching materials correspond to those in real contexts. For example, Nelson (2000) investigated vocabulary in business English textbooks in comparison with a corpus of spoken and written language used by native speakers of Business English. His findings suggest that textbooks represent Business English in limited ways; for example, words denoting concrete items are emphasized but less attention has been paid to descriptions of states and qualities. In addition, Sriumporn (2015) also compares lexical bundles found in business English coursebooks at Thai universities with those extracted from a corpus of business news articles. It has been found that there is little correspondence between business teaching materials and business news.

Regarding the register-oriented approach, studies can be divided into two groups, according to modes of communication: those that explore spoken business English and the other written. Examples of the former group include Williams (1988) and Angouri (2010), both of which compare language used in business meetings with that presented in business English textbooks. Both studies, though more than 20 years apart, found that there was almost no correlation between the language used in real business meetings and that referred to in the meeting sections in textbooks. In written business English, scholars usually look at business

letters and examine linguistic elements such as hedging devices (Zhang, 2007) and lexical phrases (Sinturat, 2010) since letters are regarded as traditional types of business correspondences (Koester, 2006).

With the growing importance of emails, a comparative perspective on authentic and taught business emails is needed. The present study therefore addresses this subject, with a focus on informal business emails. The term 'informal business emails' applied in this study refers to emails concentrating on personal but work-related matters, such as congratulating on promotion or arranging appointments. The research questions addressed in the study are:

1) What are functional types of lexical bundles in authentic and textbook informal business emails?
2) What are similarities and differences between lexical bundles found in informal authentic and textbook business emails?

The present paper employs a corpus-based method to identify most common lexical bundles, multi-word sequences, in informal business emails from both samples in real use and in textbooks. The paper begins with detailed information about lexical bundles, followed by the methodology adopted in the present study, concerning corpus building and lexical bundles extraction. Findings are then reported and discussed linguistically and pedagogically before a conclusion is given.

2. Lexical bundles

Lexical bundles are sequences of words which are in the exact form and repeatedly used in a certain register (Biber et al., 1999). They are automatically extracted from a corpus, a collection of dataset compiled to maximally represent a particular language or language variety (McEnery & Wilson, 2001). A lexical bundle does not necessarily express an idiomatic meaning in the way that an idiom like *kick a bucket* does. An everyday literal expression extracted from a corpus like *please let me know* is also a lexical bundle. Because lexical bundles are linguistic elements that recur in a particular text type, they are considered "discourse building blocks" (Biber, 2009) that "fulfill communicative purposes that are particularly important for each of the register" (Conrad & Biber, 2005, p.63), whether the spoken or written register (Biber & Barbeiri, 2007). Specifically, Biber *et al*. (2004) and Biber (2006) suggest that functions of lexical bundles in communication can be divided into four categories, which will be applied as an analytical framework in this study. They are discussed in turn below.

1. **Stance expressions** are bundles that show the assessments of certainty or attitudes towards expression following the bundles. There are five sub-categories of stance expressions: Epistemic stance, Desire, Obligation/directive, Intention/prediction, and Ability/effort.

- **Epistemic stance** refers to lexical bundles that express the writer's certainty or uncertainty**,** e.g. *I'm not sure, I don't think so*
- **Desire** includes lexical bundles that show the wishes and desires, or inquire about the other person's desire, e.g**.** *I just wanted to, if you want to*;
- **Obligation/directive** refers to lexical bundles that show obligations or direct the reader to do certain things that the writer wants to have accomplished, e.g. *you should know that, you have to be*;
- **Intention/prediction** includes lexical bundles that express the writer's intention to perform some future actions, e.g. *we're going to, I was going to*;

- **Ability/effort** involves lexical bundles concerning ability, e.g. *to be able to, to come up with*. In addition, the present study extends the scope of Ability/effort to include possession.

2. Discourse organizers are used to construct texts. They can introduce a topic, elaborate or clarify a topic, or used as a conditional sentence.

- **Topic introduction/focus** refers to lexical bundles normally used to start a new topic and frequently found at the beginning of a clause, e.g. *I'm sorry to, if you look at*;
- **Topic elaboration/clarification** focuses on lexical bundles that are used as a conjunction or start with a conjunction, e.g. *on the other hand, was going to say*;
- **Conditions** concern lexical bundles that contain a complementizer 'if', e.g. *if you need anything, if you have time*

3. Referential expressions refer to either physical or abstract units preceding or following the lexical bundles.

- **Identification/focus bundles** identify a subgroup of something, e.g. *as one of the*;
- **Tangible** refers to lexical bundles that describe a unit of concrete thing, e.g. *in the size of, in the form of*;
- **Intangible** includes lexical bundles which describe abstract units, e.g. *in nature of the, in the case of*;
- **Time reference** is lexical bundles that denote temporal units or duration, e.g. *at the same time, at the time of*

4. Special functions are the fourth category in Biber *et al.* (2004) and Biber (2006). This function incorporates lexical bundles that deal with pragmatic aspects of a communicative instance, such as politeness and enquiry expressions, which are not included in the previous three main categories. However, upon our preliminary functional analysis of the lexical bundles derived from an extraction, we found that there were other relevant pragmatic functions that could be put forward under this "special functions" category. They include: "request", "offer", "opening up for further communication", "expectation", and " hybrid function".

- **Politeness** includes lexical bundles that usually contain the term 'thanks/thank you', e.g. *thank you for your, thanks for your help*
- **Request** refers to lexical bundles asking the reader in a polite way to perform an action. They normally include the word 'please', e.g. *please let me know*
- **Opening up for further communication** focuses on lexical bundles that give opportunity for the reader to contact the writer in the future, e.g. *let me know if*
- **Offer** concerns lexical bundles that offer help, suggestion, or opinion in the following proposition, e.g. *let me know if*
- **Expectation** includes lexical bundles that express the writer's expectation, especially the phrase 'look forward to', e.g. *I look forward to, look forward to hearing*
- **Hybrid function** refers to lexical bundles that combine two functional types such as directive upon condition, e.g. *let me know if*

It must be noted that quite a few lexical bundles can serve more than one function, depending on their contexts. For example, *I would like to* can be seen as Desire and Topic introduction. This property of lexical bundles has an implication for the ways in which lexical bundles were interpreted and calculated in the present study as will be discussed below.

Apart from being essential to the communicative purposes of discourse, a number of studies have shown that lexical bundles are important linguistic units that deserve to be addressed in the ESL/ EFL pedagogical context. For example, teaching lexical bundles was

found to help improve EFL Iranian students' writing skills (Kazemi, Katiraei, & Rasekh, 2014). This is also a case in business English. Sinturat (2010), for instance, has shown that lexical bundles should be highlighted in business English teaching materials. Also, Hyland (2008) argues that lexical bundles can provide "an understanding of the features of the discourse they [students] will encounter in their particular courses" (p. 20), including the business field. As a result, lexical bundles should be highlighted for students who learn English for specific purposes.

3. Methodology

There are two sets of data involved in this study. The first one is a textbook email corpus (henceforth TEC) and the second one is a corpus of authentic business emails from the Enron corporation (henceforth ENRON). TEC contains a total of 751 email samples, taken from 77 business English textbooks that are either available for sale in bookstores or have really been used in university business English classes in Thailand. On the other hand, ENRON contains 1,061 emails recruited from a project organized by University of California-Berkeley (Berkeley). All emails in the Berkeley project have been selected and categorized according to their contents into six groups, such as employment emails, meeting arrangement emails, etc. This categorization system was observed when we differentiated textbook emails in TEC into sub-corpora in order to maximize the qualitative comparability of email samples in the two corpora. For the purpose of the present study, we focus on informal business emails, e.g. congratulations emails, colleague-talk. This sub-corpus of TEC contains 139 email samples with a size of 12,299 tokens. The corresponding sub-corpus in ENRON consists of 70 emails with a size of 18,184 tokens. Obviously, these two corpora are relatively small in the light of general corpus linguistic research. This is largely because we are dealing with a very specific type of discourse in specific domains, i.e. informal business emails in a corporation and textbooks. The figures to be reported as findings are hence relatively small and the percentages should be interpreted cautiously. Nevertheless, quantitative information in the present study can be of use in suggesting routes for further studies.

To extract lexical bundles, concordance software called AntConc was employed (Anthony, 2014). The software serves as a multiplatform tool for conducting research in corpus linguistics and data-driven learning. The 'Clusters/N-Grams' function was used to yield a list of lexical bundles out of the two corpora. The criteria for an extraction of lexical bundles in the present study concern: (1) the word length of a lexical bundle and (2) the distributional range of a lexical bundle across a whole corpus. After several pilot experiments, we found that an appropriate length of lexical bundles for the present study is four words for two reasons. First, three-word bundles were usually embedded in four-word ones; for example, the three-word bundles *as the result* and *the result of* were part of the four-word bundle *as the result of*. It seems then of less use to include overlapping bundles of less than four words, which originate from the same longer expression. Second, four-word lexical bundles are more commonly found than five-word bundles. A choice of lexical bundles longer than four words would reduce the number of bundles to be studied. Note that the chosen length of lexical bundles corresponds to that found in several previous studies (e.g. Cortes, 2013, Biber et al., 2004; Cortes, 2002; Hyland, 2008). It must be noted that although a contracted lexical bundle looks like a three-word bundle (e.g. *I'd like to*), it is counted as a four-word one due to AntConc's extraction performance. For example, *I'm going to* is counted as a four-word lexical bundle because the word 'I' and 'm' is counted separately. As for the distribution pattern, no appropriate range of distribution has been suggested. Again, based on our pilot experiments on our relatively small corpora, we decided that the four-word lexical bundles to be analyzed must occur in at least 2% of the total number of texts in each corpus. This is in order to include as many recurrent lexical bundles as possible for our analysis. This results in a focus on the lexical bundles that

occur in at least three texts across TEC and those that were found at least in two texts in ENRON.

With these criteria, a list of four-word lexical bundles in each corpus was generated. Upon an examination of the lists, we manually excluded two types of the bundles: (1) those that contain context-dependent words or proper nouns, e.g. *San Francisco in July* and (2) those that contain lexical items across a clause or sentence boundary, e.g. *me know if you*. In addition, overlapping lexical bundles which originate from the same longer expression were combined into five- or six-word lexical bundles. This is to remove redundant cases where two or more lexical bundles are actually part of the same longer sequence; for example, *look forward to hearing* and to *hearing from you* are four-word lexical bundles that in fact constitute the lexical bundle *look forward to hearing from you*.

Lexical bundles that turn up as a result of the above criteria were analyzed in terms of their functions, based on Biber *et al*.'s (2004) and Biber's (2006) functional categories explained in Section 2. To enhance the validity of functional categories we labeled, two persons were asked to identify functions they considered associated with the lexical bundles on the lists. One of the two raters is an advertising and marketing manager with 10 years of experience and the other an English native speaker lecturer at a university, who has taught business English for four years. The degree of similarity in identification of categories among the two raters and researchers was 79%.

4. Findings and discussion

The above threshold resulted in a total of 31 and 36 lexical bundles in ENRON and TEC, respectively. (See Appendix A for a full list of lexical bundles in each corpus). As mentioned above, due to the small size of the corpora, the figures reported here must be interpreted cautiously. They are used mainly to suggest a tendency in which the different types of lexical bundles constitute the discourse of informal business emails in textbooks and authentic use. The quantitative information here can be applied for future large-scale research with appropriate statistical measures, e.g. studies that compare different sub-types of business emails. The distribution patterns of all the four-word lexical bundles in TEC and ENRON are shown in Figure 1 below:
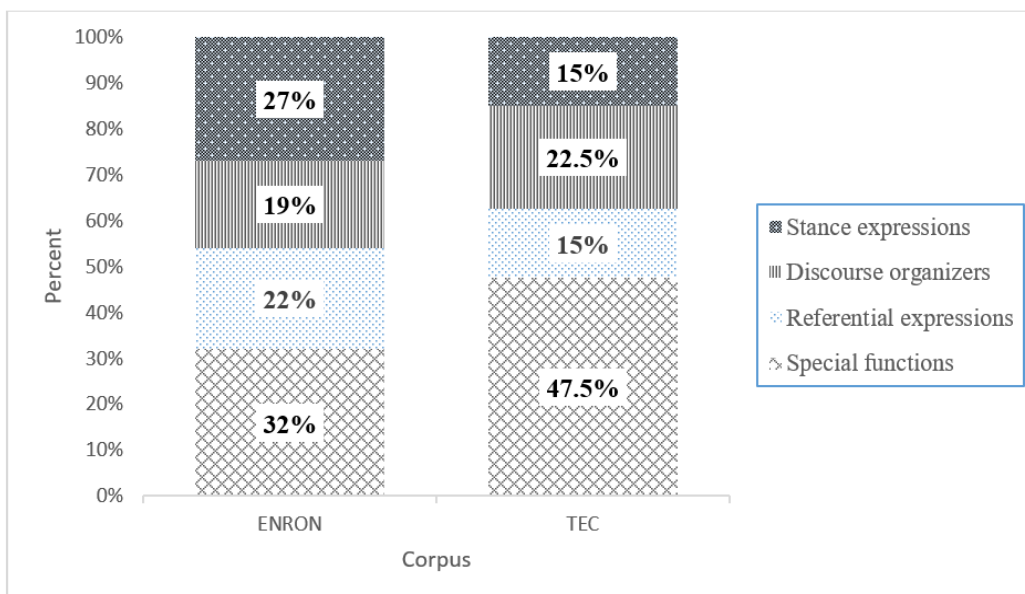
Figure 1. Functional distribution of lexical bundles

As can be seen from Figure 1, the distribution patterns of functional categories on both lists are similar and different in some ways. The only one similarity between the two corpora is that *Special functions* is the most frequent type, with 32% in ENRON and 47.5% in TEC. In contrast, the other three functional types of both corpora go in the opposite directions. While the *Stance* and *Referential expressions* are least frequently found in TEC, they are ENRON's second and third most frequent types, respectively. In addition, the *Discourse organizers* type, which is ranked second in TEC, turns up as the least frequent type in Enron. In this respect, the chi-square test indicates statistical difference between the two sources of business emails at the 0.05 level ($X^2 = 0.391132$, *df* = 3). Details about each functional type of the results are discussed below.

4.1 Stance expressions

Lexical bundles that express stance and the associated sub-categories are shown in Table 1 below. They are the second most frequent in authentic business emails but the least frequent in business English textbooks.

Table 1. Stance expressions in ENRON and TEC

Subcategory	ENRON	TEC
1. Epistemic	I'm not sure	-
2. Desire	I can't wait *I wanted you to be aware would like to keep *I'd like to*	*I'd like to* I just wanted to *it would be good to wanted to give you
3. Obligation/directive	**I don't have** **you should know that**	-
4. Intention/prediction	we're going to	be in touch with
5. Ability/effort	**I don't have** you get a chance	**I haven't got**

Note: (1) The asterisk mark (*) is put in front of a lexical bundle resulting from combining overlapping cases and used throughout the paper.
(2) The **bold** lexical bundles are those that can be found in more than one functional type.
(3) The *italics* lexical bundles mean they are found in both corpora.

With regards to the number of subcategories and lexical bundles, authentic business emails outnumbered textbook ones. In ENRON, ten *Stance expressions* constitute 27% of all the lexical bundles on the list, in which all five subcategories were identified in Table 1 above. On the contrary, only six lexical bundles from three subcategories were found in TEC. It can be inferred that TEC presents learners with *Stance expressions* less in number and types in comparison to ENRON.

Regarding the *Desire stance* bundles, each corpus offers four types of lexical bundles, in which *I'd like to* overlaps in the two corpora. In addition, it is observed that the word *wanted* are repeatedly used to express the wish of the subject as the following concordance lines display:

- Sent this out to my direct reports, but : ***I wanted you to be aware*** as well. Rick Buy (ENRON)
- Hi Hope you had a good weekend. ***I just wanted to*** give you a bit of feedback on the English lessons. (TEC)
- Harrison called this morning to talk about the XTC contract. He also ***wanted to give you*** a name — John Peters 0046 57576984 (TEC)

In addition, there are two *Desire* lexical bundles which share the same second functional category: *I'd like to* and **it would be good to*. Both of them also perform *Topic introduction function*, which will be discussed in the next section. However, in TEC, the bundle *I'd like to* serves only the *Desire* function.

- have an idea on the breakout sessions. It needs some work but **I'd like to** talk it through with you. (ENRON)
- about price, and about customization of our products for the Indian market. **I'd like to** know whether we can offer them a more basic (TEC)
- problems in Milan implementing a new customer service database As a result, **I'd like** to cancel my participation in the European management circle meeting (TEC)

The *Obligation/directive* function is found only in ENRON. They are *I don't have* and *you should know that* as exemplified below. As for the bundle *I don't have*, it must precede *to*, forming a modal verb phrase 'have to', in order to perform this function. The bundle *you should know that* in the example is considered *Obligation/directive* type since the writer wanted the reader to acknowledge the fact of the following proposition of the lexical bundle, making the bundle similar to a directive function. This could be inferred from the use of conjunction 'but' preceding the lexical bundle when the writer wanted to point out something to the reader who presumably lacked such knowledge.

- wanted to let you why and I am doing it via email so **I don't have** to say it many times. (ENRON)
- better because of Wade's involvement. It was a team effort, but **you should know that** Jimmy Mogal also was very involved. (ENRON)

As for *Ability/effort* subcategory, none of lexical bundles in this data contains the modal verb *can* or *could* but have verbs *have, get (got)* instead. This is because possession (*have* and *got*) is considered as subtype of *Ability/effort* for the present study as mentioned in Section 2.

- We discussed the issue. **I don't have** any problems. (ENRON)
- Subject: Video Tape Steve. Did **you get a chance** to look at the video tape I gave you? (ENRON)
- Sorry we didn't get a chance to Talk before you left. But, I'm still here; give me a call when **you get a chance** - 1586. (ENRON)
- The agenda looks fine to me, and **I haven't got** anything to add — except of course to pass on my congratulations to all involved, especially Diego. (TEC)

4.2 Discourse organizers
Discourse organizers are normally used to construct a discourse by introducing a new topic, elaborating the topic, or providing a condition. Eight types out of the total 40 types in TEC, amounting 22.5%, are discourse organizing bundles. On the contrary, this function is the least found in the authentic email corpus, with only five types out of 37 types in total, equaling 19%. The full list of lexical bundles in each corpus is displayed in Table 2 below.

Table 2. Discourse organizers in ENRON and TEC

Subcategory	ENRON	TEC
1. Topic introduction/focus	*to follow up on the email to give you a **you should know that** **I'd like to**	*just a quick note to **I haven't got** I'm sorry to *it would be good to *wanted to let you to tell you that let you know that to let you know
2. Topic elaboration/clarification	so I don't	**-**
3. Conditions	**if you have time** if you have any	if you need anything

It can be seen that the majority of *Discourse organizers* are *Topic introduction/focus*, followed by *Conditions*, while *Topic elaboration/clarifications* is identified only in ENRON. This highlights the fact that authentic business email corpus seems to contain a wider range of functional types despite being smaller in the number of lexical bundles.

Regarding *Topic introduction/focus* function, TEC offers seven lexical bundles while ENRON contains merely four lexical bundles. This might reflect that email writers in this data used limited variations of topic introducing bundles.

Among the *Topic introduction/focus* bundles in TEC, three-word multiword units, *let you know*, are embedded within two four-word lexical bundles – *let you know that* and *to let you know*. It is noteworthy that *let you know that* always follow *to* while *to let you know* can precede either *that* or other linguistic element (i.e. what), which is the reason why it is extracted as a new lexical bundle since this more flexible bundle appears in more texts than *let you know that*.

- Hi Emma Just a short note to ***let you know that*** the meeting with Mr Gupta and his business associates (TEC)
- Hi Susan You asked me ***to let you know*** what happened in the meeting yesterday. (TEC)

As mentioned in 4.1, two stance bundles from each corpus, totaling four bundles, can serve as *Topic introduction/focus* function. They are *I haven't got, *it would be good to, you should know that,* and *I'd like to.* The following concordance lines illustrate the contexts in which they are found to perform the *Topic introduction* function.

- and the HR management team thought ***you should know that*** participants rated your presentation very high. (ENRON)
- last on which I'll be assessed as a Research guy, so ***I'd like to*** ask if you'd be prepared to act as reviewers for me? (ENRON)
- Williams Subject: Re: meeting with Mr Gupta Dan - thanks for this. Sorry ***I haven't got*** back to you sooner. What's the next step? (TEC)
- Don't forget to check with Marek about transport. And another thing - ***it would be good to*** get some fresh air in the mountains (TEC)

The bundle *I haven't got* is followed by the preposition *back*, constituting a phrasal verb 'got back'. Since the phrasal verb refers to the action of replying, this lexical bundle thus is part of *Topic introduction*, which links to the unfinished business between the two parties. Meanwhile, the lexical bundle *you should know that* in this context serves as a discourse building block by introducing a new topic that the writer wants the email recipient to know. In the case of *I'd like to*, this bundle is used to prolong the question asking the email recipients to do him a favor.

It is interesting to note that a personal pronoun is missing in two longer expressions in *Topic introduction/focus*, *to follow up on the email and *just a quick note to from ENRON and TEC, respectively. Consequently, they appear similar to fixed expressions since they are longer than four words but repeatedly used in the same structures.

Apart from lexical bundles for introducing or elaborating topics, any bundles containing the conjunction *if* will be put into the conditional function. It is found that all three lexical bundles from ENRON and TEC have *you* as their subject as show in the following concordance lines.

- I would recommend reading the whole text ***if you have time***, and if you seek perverse entertainment. (ENRON)
- ***If you have any*** interest, I would like to talk to you sometime. (ENRON)
- They had data problems — IT failure. ***If you need anything*** else, tell me. That's all. BD (TEC)

4.3 Referential expressions

In this functional category, both ENRON and TEC corpora have *Tangible* and *Time reference* but *Identification/focus* and *Intangible referential expressions* are identified only in ENRON. This correlates with the fact that referential expressions are at the bottom of TEC list. Table 3 shows lexical bundles classified as *Referential expressions* in the two corpora.

Table 3. Referential expressions in ENRON and TEC

Subcategory	ENRON	TEC
1. Identification/focus	as one of the	-
2. Tangible	got a call from	the budget of the
3. Intangible	congratulations on your new get a chance to had an opportunity to in light of the	-
4. Time reference	a couple of moths as soon as possible	a couple of weeks end of next week such a long time by the end of in the coming days

In line with the previous two functional categories, ENRON covers a wider range of functional types than TEC. Two missing subcategories in TEC are *Identification/focus* and *Intangible*. As for *Identification/focus*, the bundle *as one of the* identifies the subject of the clause as a member of certain group in its following proposition, as exemplified below. It is noticeable that the lexical bundle in both samples are associated with positive meanings as seen from its co-occurrence with phrases "the skill to assist" and "brave and bold".

- You have been identified *as one of the* few members who has the skill to assist in this area. (ENRON)
- You still go in my book *as one of the* brave and bold for standing up to that sort of intimidation for so long. (ENRON)

The *Time reference* sub-type shares the largest proportion of referential expressions in TEC with five lexical bundles. Although both corpora differ in the number of the lexical bundles in this sub-type, they share the same three-word bundle *a couple of* plus the unit of time as *weeks* (TEC) and *months* (ENRON). This can be taught to learners by suggesting them to use the three-word chunk and selects the time unit suitable their contexts to fill in the gap.

4.4 Special functions

Lexical bundles in the *Special functions* category constitute the largest proportion of functional types in both ENRON and TEC, with 32% and 47.5% of total lexical bundles, respectively. A full list of this type of lexical bundles is given in Table 4 below.

Table 4. Special functions in ENRON and TEC

Subcategory	ENRON	TEC
1. Politeness	doing a good job enjoyed working with you thanks for your help	working with you and you've done a thank you for your thanks for your email it was good to
2. Request	**if you have time** *please let me know*	*could you let me know get back to me please let me know you let me know let me know what
3. Opening up for further communication	*let me know if*	*let me know if*
4. Expectation	*look forward to hearing from you *look forward to seeing you I look forward to*	*hope to hear from you soon to seeing you again *look forward to hearing from you to hear from you *look forward to seeing you I look forward to*
5. Offer	*let me know if*	*let me know if*
6. Hybrid function	*let me know if*	*let me know if*

As shown above, five out of six overlapping bundles in the two corpora are found in this category (see those in italics). It thus significantly shows how essential special functional lexical bundles are in the business email discourse since the majority of overlapping lexical bundles are categorized into this function.

With respect to the *Request* function, TEC offers a wider range of lexical bundles than ENRON. The present study classified a lexical bundle into this subcategory on the primary condition that the word *please* or other linguistic element showing the similar connotation is apparent. One lexical bundle overlapping in both corpora is *please let me know*. Similar to *Topic introduction/focus* and *Time reference* in 4.2-4.3, the fact that a larger number of lexical

bundles under the same subcategory is found in TEC supports findings from previous studies (e.g. Sánchez, 2014) that textbooks contain a larger proportion of lexical bundles.

The *Expectation* function concerns the writer's expectation to receive a certain action from the reader. All three identified lexical bundles in ENRON overlap with those in TEC, including **looking forward to seeing you*, **look forward to hearing from you*, and *I look forward to*. This again supports the argument that textbooks tend to teach more types of forms rather than functions compared with the authentic email corpus in the present study.

The last three subcategories found in a single lexical bundle are *Offer, Opening up for further communication* and *Hybrid function*. In both corpora, a single lexical bundle, *let me know if*, shows three different functions as in:

- ***Let me know if*** there is anything I can do to help. (ENRON)
- Hope this isn't going to cause you too much trouble. ***Let me know if*** I can be of assistance. (TEC)
- More information can be found in the document below. Please ***let me know if*** you have questions or comments. (ENRON)
- The total cap is definitely the better way to go. ***Let me know if*** you need more. (TEC)
- ***Let me know if*** it begins to interfere unduly with your new responsibilities. (ENRON)
- ***Let me know if*** there are any problems. Thanks (TEC)

The lexical bundles in the first two concordance lines serve to offer in the form of conditional sentence. This is because the two bundles are followed by a phrase where the writers offered to help the recipients on a condition as in "there is anything I can do to help" and "I can be of assistance". The next two instances are categorized into the *Opening up for further communication* function since the writers encouraged the email recipients to get in touch if they "have questions or comments" in ENRON or "need more" in TEC. The final two concordance lines shift to a directive tone and are classified as belonging to the *Hybrid function* since they directed the email recipients to perform the action of notifying the email writer if something happens as in "it begins to interfere unduly with your new responsibilities" and "there are any problems".

5. Pedagogical implications

Based on the findings, the similarity between textbook emails and those in real business contexts is that both corpora mostly contain lexical bundles in the *Special functions* category, i.e. those that deal with pragmatic aspects of the email, such as *Politeness* and *Expectation*. This functional group of lexical bundles have been found particularly frequently in Conrad and Biber's (2005) study of conversational discourse, when compared with academic discourse. Based on this correspondence, it can be argued that the predominance of *Special function* bundles in both ENRON and TEC reflects an informal writing style of this email type. However, there are still several differences between sample emails in business English textbooks and authentic emails. Nevertheless, some types of lexical bundle such as *Obligation/directive*, *Topic elaboration*, *Intangible*, etc., can be found only in ENRON, not in TEC. Although textbooks attempted to include a variety of forms, they cannot cover all functions found in the real business contexts. However, it should be reminded that this research study examines only informal business emails in two small corpora. The figures revealed above show possible tendencies of lexical bundles that tend to be found in this type of email. As a result, some missing functional types in the textbooks might be found in other genres of business emails from teaching materials as well.

Therefore, the findings of the present study have pedagogical implications for teaching informal business emails in two aspects. First of all, the similarity between the two corpora

shows that this particular type of emails is characterized by a certain group of lexical bundles and their associated functions. Accordingly, these lexical bundles can be considered essential input that should be emphasized in business English writing instruction. Furthermore, although business English textbooks offer a wide range of lexical bundles, some functions found in the authentic email corpus are missing in textbooks. It is then necessary for classroom teaching to expose students to expressions that occur in real use. Consequently, business English teachers can consult with a corpus of authentic emails in order to supplement those missing functional categories in business English textbooks. This is to ascertain that learners of business English know essential phrases that are truly and frequently used in the real business contexts.

6. Conclusion

This study compares functional types of lexical bundles found in business emails used in a corporation and email samples in business English textbooks. It contributes to business English research in applied linguistics in shedding light on linguistic patterns and functions in informal business emails, as well as has pedagogical implications as mentioned above. The study also has a methodological contribution in that it shows that lexical bundles can be used as a descriptive tool to examine linguistic features in business English in both real and pedagogical contexts. Further research can look at lexical bundles in different types of email samples, e.g. complaint, request and inquiry ones, both in textbook and authentic emails.

References

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. (2014). AntConc (Version 3.4.3) [Computer Software]. Tokyo, Japan: Waseda University. Available from http://www.laurenceanthony.net/

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

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., & 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.

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. (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

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.

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

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

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

Koester, A. (2006). *The language of work*: Routledge.

McEnery, A. M., & Wilson, A. (2001). *Corpus linguistics: an introduction*: Edinburgh University Press.

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

Sabater, C. P., Turney, E., & Fleta, M. B. M. (2008). Orality and literacy, formality and informality in email communication. *Ibérica: Revista de la Asociación Europea de Lenguas para Fines Específicos (AELFE)* (15), 71-88.

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

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*.

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.

Waldvogel, J. (2007). Greetings and closings in workplace email. *Journal of Computer-Mediated Communication*, 12(2), 456-477.

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

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.

Zhang, H.-q. (2007). The pragmatic function of hedges in business English. *Journal of Hunan Business College*, 3, 117-120.

Appendix A

A list of lexical bundles found in informal business emails

ENRON	TEC
a couple of months	a couple of weeks
as one of the	*just a quick note to
congratulations on your new	be in touch with
doing a good job	*could you let me know
enjoyed working with you	end of next week
*look forward to seeing you	* hope to hear from you soon
get a chance to	**I'd like to**
got a call from	I haven't got
had an opportunity to	I just wanted to
I can't wait	I'm sorry to
I don't have	if you need anything
I'm not sure	*it would be good to
*I wanted you to be aware	such a long time
if you have time	the budget for the
in light of the	to seeing you again
so I don't	wanted to give you
thanks for your help	wanted to let you
*to follow up on the email	working with you and
to give you a	you've done a
would like to keep	by the end of
you get a chance	get back to me
you have any questions	in the coming days
you should know that	***look forward to hearing from you**
as soon as possible	**please let me know**
*look forward to hearing from you	to tell you that
let me know if	you let me know
please let me know	let me know what
we're going to	thank you for your
I'd like to	thanks for your email
I look forward to	to hear from you
if you have any	let you know that
	it was good to
	*look forward to seeing you
	I look forward to
	let me know if
	to let you know

* the bundles which are longer than four-word are indicated.
The **bold** bundles mean they appear in both corpora.