AN ANALYSIS OF LEXICAL BUNDLES IN BUSINESS ENGLISH
COURSEBOOKS BY DEVELOPING A THAI BELF CORPUS FROM
THE NATION AND *BANGKOK POST* NEWSPAPERS

CHRISTIE MARIE SRIUMPORN

A THESIS SUBMITTED
FOR THE DEGREE OF MASTER OF ARTS
IN ENGLISH LANGUAGE TEACHING
MA-ELT

GRADUATE SCHOOL OF ENGLISH (GSE)
ASSUMPTION UNIVERSITY
BANGKOK, THAILAND.

DECEMBER 2011

AN ANALYSIS OF LEXICAL BUNDLES IN BUSINESS ENGLISH COURSEBOOKS BY DEVELOPING A THAI BELF CORPUS FROM *THE NATION* AND *BANGKOK POST* NEWSPAPERS

CHRISTIE MARIE SRIUMPORN

ID:5319443

ASSUMPTION UNIVERSITY, BANGKOK THAILAND

December 2011

AN ANALYSIS OF LEXICAL BUNDLES IN BUSINESS ENGLISH COURSEBOOKS BY DEVELOPING A THAI BELF CORPUS FROM *THE NATION* AND *BANGKOK POST* NEWSPAPERS

CHRISTIE MARIE SRIUMPORN

ID:5319443

A THESIS SUBMITTED FOR THE DEGREE OF MASTER OF ARTS IN ENGLISH LANGUAGE TEACHING

GRADUATE SCHOOL OF ENGLISH (GSE)

ASSUMPTION UNIVERSITY, BANGKOK THAILAND

December 2011

Thesis Title	AN ANALYSIS OF LEXICAL BUNDLES IN BUSINESS ENGLISH COURSEBOOKS BY DEVELOPING A THAI BELF CORPUS FROM THE NATION AND BANGKOK POST NEWSPAPERS
Name	Ms. Christie Marie Sriumporn
Advisor	Dr. Rajeevnath Ramnath
Academic Year	December, 2011

The Graduate School of Assumption University has approved this final report of the eighteen-credit course, ET 7100 Thesis, submitted in partial fulfillment of the requirements for the degree of Master of Arts ELT
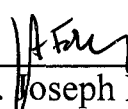
Approval Committee:

_____
(Dr. Rajeevnath Ramnath)
Advisor

_____
(Dr.Passapong Sripicharn)
External Examiner

_____
(Prof. Dr. Joseph Foley)
Chairman

Plagiarism Statement

I certify that this material in this study which is not my own work has been identified and acknowledged, and no material is included for which a degree has been conferred upon me.

Signature: _____

Date: 12|12|11

Acknowledgements

Thank you to all the family, friends, and professors that supported and guided me through one of the most meaningful chapters of my life.

Table of Contents

Chapter 1: Introduction

Chapter 2: Literature Review

Chapter 3: Research Methodology

Chapter 4: Analysis of Data and Discussion

Chapter 5: Conclusion

List of Tables

Abstract

This corpus-based study compares the lexical bundles and vocabulary in Business English coursebooks used in Bangkok, Thailand with a BELF (Business English Lingua Franca) corpus developed from business stories in English newspapers published in Thailand. It also discovers patterns within the Thai BELF corpus that reveal cultural and other attributes of Thai BELF. The repeated use of lexical bundles for various reasons are considered the cultural aspects of Thai BELF and come in many forms such as style, the use of certain words opposed to others, and the high frequency of specific lexical bundles. The use of a newspaper-based BELF corpus has not yet been explored in the teaching of Business English in Thailand.

This study is to better understand the lexical bundles and vocabulary used in Thai business news stories as a means to assess the lexical bundles being taught in Business English courses in Bangkok. A corpus of 800,000 words from March 2011 to July 2011 has been created to explore the bundles and vocabulary used in *The Nation* and the *Bangkok Post*. Two small corpora have been created to determine how the lexical bundles in Business English coursebooks such as, *Intelligent Business* (2005), *Essential Business Vocabulary Builder* (2011) and *New Business Matters* (2004) coursebooks used in Thailand and *BG 2000 English III* and *BG 2001 English IV* coursebooks used in Assumption University reflect the lexical bundles used in the Thai business environment. The top 100 most frequent lexical bundles in the Thai BELF corpus are analyzed to determine common attributes that can be compared to these two corpora.

The results of the study tend to indicate a low correlation between the corpus data and the Business English coursebooks used for this study. Much of this low correlation can be attributed to a lack of authentic texts and regional specific lexicon. Research into the Thai

BELF corpus itself revealed attributes of culture. These attributes include male gender dominance and a collectivist view of society as well as a focus on the Asian region. The use of time indicators and the role of the government in business are also discussed. Further research into the lexical bundles used in Thai BELF can aid teachers in designing syllabus to better meet the needs of Business English students in Thailand.

Chapter 1

Introduction

This chapter will explain the reasons for conducting this study. Major objectives of the study as well as the research questions will be presented. The significance of this thesis will also be explored before explaining key terms used in this study.

1.1 Background

Teaching English in one part of the world is never exactly the same as another. English itself changes and adapts to suite its specific purpose based on the environment of its use. English as a Lingua Franca (ELF) is no new theory. (Jenkins 2008) However, the realities of its existence have not yet been fully applied to English teaching. The English used in business is specialized and courses focused in Business English are in great demand. For this reason it is necessary to teach appropriate vocabulary that is relevant to the actual environment it is being used in. The English in a business context, often called BELF (Business English as a Lingua Franca), used in Thailand often takes place in a company owned and largely managed by a foreign based company. The foreign managers use English, which may not be their first language, to talk with Thai staff. Being that the Thai staff is using English in everyday situations at work but dealing with a majority of their fellow Thai, they may adapt their English to closely match the conversational norms of Thai society. For example, Thai people generally pay far more attention to seniority as a guide for interaction and being careful not to offend anyone than the typical westerner. Use of strong direct language is not usually a common occurrence. More specifically, the addition of Thai vocabulary such as "Khun" as an address term into an otherwise English conversation. These adaptations should be reflected in

the teaching of English. An efficient way to discover relevant vocabulary and other word use is through corpus based analysis.

Corpus as a research tool has been used for the last two decades to analyze spoken and written texts for the benefit of understanding language varieties of language better. The linguistic discoveries made possible by corpus research has aided in a variety of linguistic areas including language teaching. Tan acknowledges this by stating:

> What corpus linguistics offers for education are evidence, from data, of discoursal, socio-cultural and psycholinguistic insights which provide direct applications and even restructuring of syllabuses and materials for teaching and learning.

(Tan 2002: 4)

Corpus linguistics itself is based on real life language use. It is a "methodology rather than an aspect of language requiring explanation and description". (McEnery et al. 2001: 103) Corpus linguistic research is carried out in a scientific manner using empirical data. As explained by McEnery et al.:

> Empirical data enables the linguist it make statements which are objective and based on language as it really is rather than statements which are subjective and based upon the individuals own internalized cognitive perception of language.

(McEnery et al. 2001: 103)

Although most corpora use native speaker data, there are some ELF corpora in use. They have successfully been used to study aspects of language such as lexis. A study in Singapore using a Singapore-English corpus of newspaper articles found lexical variations that were unique to Singapore-English. (Pakir 2009: 86)

These benefits of using a corpus allows the researcher to explore the vocabulary in the *Bangkok Post* and *The Nation*, two major local English newspapers published in Thailand. The articles are written by mostly Thai journalists and represent the BELF environment in Thailand. Then the researcher can use the corpus data to determine how well the chosen

coursebooks used in Bangkok accurately reflect the lexical bundles used in the genuine business environment. From the researcher's understanding, this type of research has never been applied to the teaching of Business English in Thailand.

1.2 Rationale

The driving force of this study is to better understand the linguistic features of Thai business news stories as a means to assess the probability of the same lexical bundles being taught in Business English courses in Bangkok. Corpus study enables the researcher to make discoveries in language phenomena that are a "genuine reflection of the behavior of a language". (McEnery et al. 2001: 76) As stated earlier, the business environment in Thailand often requires its own form of BELF. Cultural norms of Thai society and other non-native or native English speakers as well as the influence of different registers create a unique type of language use that is represented in both written and spoken discourse. Using the corpus to research lexical bundles as opposed to only vocabulary allows the researcher to discover linguistic features that are more context-based and thus more relevant.

Corpus linguistics has been used to improve many areas of language teaching including LSP (Language for a Specific Purpose also known as English for Specific Purposes). (Tan 2002: 2) It is important for vocabulary and other course content to represent the cultural aspects of the societies they are being used in. (Ur 1999: 197) The result of not doing so can be a lack of connection between course content and learners or inadvertently imposing a type of cultural stereotype. (Harmer 1990: 23) (Hedge 2000: 38) Additionally, there is a lack of research in Business English in Thailand and cultural attributes of Thai BELF.

The corpus created from the *Bangkok Post* and *The Nation* newspapers is a relevant source of data for this study. The language used in the articles are written by Thai authors in the business context and follows norms of business news articles and functions in the context of

Thai society. This context has its own rules as to what is appropriate or not. The coursebook corpus is made up of articles chosen based on content in a business context and informative genre (news stories). They can be compared to the larger news stories corpus accurately because the content of the coursebooks is supposed to be a concentrated version of real life vocabulary use. Generally speaking, the authors claim that the content of their coursbooks is what a Business English learner ought to know.

The corpus data can be used to determine the common lexical bundles used in real life and in adherence to the cultural norms of the Thai community. Bundles are focused on over single vocabulary because their structures reveal and language use at the same time instead of just vocabulary. This study can then compare the bundles to texts used in five coursebooks: *Intelligent Business* 2005, *Essential Business Vocabulary Builder* 2011, *New Business Matters* 2004, BG 2000 English III 2010, and BG 2001 English IV 2010 to see unique attributes, commonalities and differences. The corpus itself and the results of this study can be used to evaluate other coursebooks in terms of their appropriateness to the Thai context and to better understand the lexical uniqueness of BELF in Thailand. Once the research is complete, the results of the text analysis can aid teachers to decide what needs to be supplemented or what is inappropriate. The coursebooks have been chosen based on the following criteria: international focus, published in the last decade, appropriate levels, vocabulary focus, actual use in Thailand.

1.3 Objectives

- Create a corpus of authentic business news stories from the *Bangkok Post* and *The Nation* newspapers from March 2011 to July 2011.

- Create a list of lexical bundles from the corpus data.

- Compare the newspapers corpus bundles to the bundles from the *Intelligent Business, Essential Business Vocabulary Builder, New Business Matter, BG 2000 English III* and *BG 2001 English IV* coursebooks.

- Determine the attributes, similarities and differences of lexical bundles in the news stories corpus and coursebook corpora.

- Offer recommendations for Business English syllabus design in the Thai context.

1.4 Research Questions

1. What are the most frequently used BELF lexical bundles and their attributes in the business pages of the *Bangkok Post* and *The Nation* newspapers?

2. To what extent are the most frequently used lexical bundles in the Thai BELF corpus represented within the *Intelligent Business, Essential Business* and *New Business Matters* coursebooks available in Thailand (MAS corpus) and the *BG 2000 English III* and *BG 2001 English IV* coursebooks used at Assumption University (ABAC) in Thailand (LIM corpus)?

1.5 Significance of the Study

This study can provide valuable insight into an area of English that is in high demand and ever changing. One of the main purposes of English is to conduct business for financial gain, a top priority in many people's minds. A great number of people in Thailand will benefit from more effective learning of Business English. This study is likely to shed light on the teaching of Business English in Thai context by bringing further attention to the need for teachers to tailor their vocabulary base to suit the needs of their learners. If texts in coursebooks can be chosen to better connect with the actual environment the learners are functioning in then the transition between understanding classroom texts and real life texts will be more seamless.

5

Second, the corpus itself can provide useful insights into the distinctive language features of the Thai business community for other teaching and research purposes. As states earlier, due to the influence of Thai culture, other non-native English users, and native English speakers all working for the same field or company, BELF in Thailand has evolved. Not only do speakers in this environment adopt vocabulary from registers around the world, they do so in a way that fits the nature of Thai communication. The effect of culture on language has already been studied in non-Thai contexts. For example, Tribble's (2002) study investigated aspects of culture in English language education in England. And the use of corpus to has been used for task development in Fuentes and Rokawski's (2002) study, which explored the use of corpus-based task development in an IT genre of Business English in Spain. But, it has not yet been applied to Thai users of BELF.

Another significant contribution of this study is its additional aid to the understanding of ELF and BELF in Asia. The English used in Thailand in general or in business is part of Thailand's people's identity. Jenkins states:

> ...while some speakers may currently wish as individuals to identify themselves linguistically (but rarely in other ways) with NS's [Native Speakers] of English, they may at the same time feel more "at home" in English as part of their own linguacultural community or even an international NNS [Non Native Speaker] community, and wish to signal their affiliation to these groups linguistically.

(Jenkins 2007: 199)

The idea of identity coming forth from one's use of language is important in the understanding of a culture. The many users of this specific International English can benefit from a better understanding of their ELF and BELF characteristics. By continuing to update the corpus, the development of the Thai Lingua Franca identity can be monitored and researched from a cultural awareness perspective.

1.6 Definition of Terms

Attributes (of language): According to the researcher the repeated use of lexical bundles and key words for various reasons including but not limited to the form of stylistic choice, the use of certain words opposed to others or patterns that materialize from these reasons.

BELF: (Business English as a Lingua Franca) A Lingua Franca used in business situations where specific context understanding is needed to succeed in business communication in the areas of national and corporate cultures, knowledge of the other party and his or her organizational role, business competence and business know-how overall. (Kankaanranta et al. 2010)

Business English: Business English is a type of English for Specific Purposes (ESP) but has its own specific terms and "emphasis on particular kinds of communication in a specific context" and is "often a mix of specific content (relating to a particular job area or industry) and general content (relating to general ability to communicate more effectively, albeit in business situations). (Ellis 1994: 3)

Corpus: "Corpora or corpuses (singular: corpus) are simply large collections of databases of language, incorporating stretches of discourse ranging from a few words to entire books."

(Schmitt 2000: 68)

Culture: For the purpose of this research, "culture" is considered to be the reoccurring aspects of society, or "norms", which can be observed through language used in the context of that society. Specifically, the "norms" of language that are found in the Thai BELF corpus, the repeated use of lexical bundles for various reasons, are considered the cultural aspects of Thai BELF. These aspects of culture can come in

many forms such as style, use of certain words opposed to others, and the high frequency of specific words or lexical bundles.

English for Specific Purposes (ESP): This term refers to teaching a specific genre of English for students with specific goals. Examples include English for Academic Purposes (students will enter an English-speaking university), business English (for business people), medical English (for nurses, doctors, other health care professionals). ESP can be (as the name implies) very specific -- a growing market is English language computer support personnel. In sum, ESP is English for vocational purposes, where the word vocation is used loosely to include education and all kinds of employment. (http://www.teaching-english-in-japan.net)

Lemma: Used to mean the base form of a word, disregarding grammatical changes such as tense and plurality. (Biber et al. 1998) For example, the word in bold is the lemma: "**mean**", "**meaning**", "**means**".

Lexical Bundle: Also called word clusters or word chunks, these are recurrent sequences of words commonly found together in texts or spoken language. Some examples are: "ACCORDING TO THE", "YOU MIGHT WANT", and "THINK WE NEED TO". Other characteristics of lexical bundles are that they are "usually not structurally complete and not idiomatic in meaning". (Biber and Barbieri 2007: 263)

Lingua Franca: (LF) The language used between people who don't share a common native language or common nationality. (Seidlhofer 2005) Further explained, Lingua Franca, whether it is English or another language, is the language that two or more people use to communicate with each other in the event that they do not know each other's native language.

Tokens: In the context of a corpus, tokens are the individual words in the corpus.

1.7 Thesis Organization

Chapter two, the literature review, focuses on prior studies that have contributed and influenced the creation of this study. Chapter three is the methodology of this study. The methodology will cover the development of the corpus, data collection, method for data analysis, and a pilot study. Chapter four is analysis and discussion of the research findings. Lastly, chapter five concludes this study with the major findings followed by limitations and some directions for future research.

1.8 Conclusion

Business English is commonly used as a Lingua Franca throughout the world. The coursebooks used in Thailand that aid in teaching this variety of English have yet to be subject to a corpus-based analysis. This study uses corpus-based analysis to objectively determine the common clusters in the *Bangkok Post* and *The Nation* newspapers and compare the *Intelligent Business, Essential Business* and *New Business Matters* coursebooks and the *BG 2000 English III* and *BG 2001 English IV* coursebooks used in Thailand to determine the extent that these coursebooks are corpus-based. The next chapter will discuss research that pertains to this study.

Chapter 2

Literature Review

2.1 Introduction

This chapter will discuss the significance of BELF and its connection to vocabulary and word clusters in business English coursebooks. Some significant studies in regards to Non-native English corpora, corpus-based study and lexical bundles, corpus in English language teaching, and corpus in Business tasks will be explained.

2.2 Business English as a Lingua Franca (BELF)

English as a Lingua Franca (ELF) is a concept that is known to be one of the reasons that English has spread to so many areas of the world including Thailand. According to Seidlhofer:

> ...(ELF) has emerged as a way of referring to communication in English between speakers with different first languages. Since roughly only one out of every four users of English in the world is a native speaker of the language (Crystal 2003), most EFL interactions take place among "non-native" speakers of English. Although this does not preclude the participation of English native speakers in EFL interaction, what is distinctive about ELF is that most cases, it is 'a 'contact language' between persons who share neither a common native tongue nor a common (national) culture, and for whom English is the chosen foreign language of communication'(Firth 1996:240).

(Seidlhofer 2005: 339)

This brings on another aspect that ELF users are subject to, "a user of English as a lingua franca thus has to accommodate to different types of speakers from different linguistic and cultural backgrounds with different levels of competence in each speech situation" (Mollin 2006: 45). The speech situation using ELF most commonly found in the world including Thailand is business.

More specifically, the term BELF has been introduced as a more narrowed view of ELF.

BELF's definition of use in addition to its' LF aspect is:

> As BELF is used in the business domain to get the job done, it automatically implies certain roles for the language users (e.g. buyer, seller, manager), the kind of jobs they do (e.g. negotiate deals, manage projects, lead people), the issues they discuss (e.g. prices, recruiting, finance), and the genres they use (e.g. business email, intranet, meetings).

> (Kankaanranta et al. 2010: 204)

A closer look at Kankaanranta's definition reveals some key points to BELF. First, the motivation for using English is clearly defined as the completion of a task. In contrast to general ELF that's use can be used based on a wide range of motivations, BELF have one specific purpose. The roles implied by the domain are also almost identical to job titles. This takes some of the aspect of personal individuality out of the formula because these roles have certain unavoidable universal responsibilities and expectations placed on them. Job type is also in the same league as roles however, the nature of language use in the job activities can vary depending on the cultural context of use. The cultural context of use can be culture from a national/social background or a "corporate culture" that exists within the company itself. The issues discussed are also bound to a very narrow register as dictated by the role and job restrictions that occur from situation to situation. Finally, the genre's used such as business email, intranet, and meetings all share a purpose that goes back to the main motivation for using BELF, "to get the job done". These modes of communication are meant to function as a tool for completing tasks and business-backed relationship building, not usually as entertainment or purely social reasons.

Just as with the traditional concept of ELF, BELF is needed for successful communication between participants that do not share a native language. However, in BELF, "successful communication" is defined far more detailed. Research suggests that success while using

BELF depends on "business professionals to know the other party's specific context to succeed in communication". Specific context includes national and corporate cultures and, even more importantly, knowledge of the other party and his or her organizational role. Another significant aspect of BELF is the users perception of language "proficiency". In the study, users did not associate proficiency with proper grammar and word use directly, but with "conceptualization of business communication competence, business competence and business know-how overall." Also, in regards to the three main skills of business communication, clarity and directness were rated higher than politeness. (Kankaanranta et al. 2010: 207) This research suggests that more real-use/case study Business English needs to be taught as opposed to purely teaching Business English as a regular language course.

Another aspect of BELF found in a study by Planken, is the emergence of "safe talk". Safe talk occurs during a meeting of people from different cultural backgrounds and native languages. These participants consciously and verbally acknowledge each other's cultural differences and try to create an inclusive new environment made up of all non-native speakers. (Planken 2005, cited in Canagarajah 2006: 207)

2.3 Business English

Business English is a type of English for a specific purposes (ESP). According to Olivera et al., "ESP is an essential component of ELT [English Language Teaching], retaining its emphasis on practical outcomes stemming from needs analysis, genre analysis and preparing learners to communicate effectively." (Olivera et al. 2003: 143) Unlike general English, ESP is tailored to suit the needs of learners interested in using English in a particular field. These fields include English for academic purposes (EAP) and English for occupational purposes (EOP) (i.e. aviation, medical/nursing, science and technology, hospitality, etc). Each of these fields has a specific domain and register.

When determining syllabi, there are three general characteristics of Business English that should be kept in mind. The first is a "sense of purpose". Exchanges in a business context have definite purposes. Ellis et al. explains:

> Language [business English] is used to achieve an end, and its successful use is seen in terms of successful outcome to the business transaction or event. Users of Business English need to speak English primarily so that they can achieve more in their jobs.
>
> (Ellis et al. 1994: 7)

Additionally, the objective of learning Business English for performance sake is more important than for language learning in general.

The second characteristic is a need to be socially acceptable in an international business context. This requires a "way of doing things so that people from different cultures, and with different mother tongues, can quickly feel more comfortable". (Ellis et al. 1994: 8) Being polite and appropriate during business interaction, which requires social knowledge and an understanding of appropriate lexical attributes, can often be a contributing factor to the overall success of the interaction. The third main characteristic is clarity of communication. Due to the nature of business interactions, misunderstanding from either party can have disastrous results. Users of business English need to be concise and choose their words carefully to show logical thought processes.

Besides these general characteristics, some detailed features of Business English have already been found using a one million word corpus of spoken and written Business English to examine semantic word associations. Words in a business environment interact with each other on a semantic level. According to Nelson (2005: 1),

> ...words in the business environment have semantic prosodies, that is they are found to regularly collocate with word groups that share semantic similarity... they [words] have prosodies that are unique to business, separate from prosodies they generate in the 'general English' environment.

13

There is a real differentiation between Business English and non-Business English, or "general English" (Nelson 2005: 1). This difference has also been recognized by the general public as shown by the numerous courses, coursebooks and training centers focused only on English for business. Most of these courses use situational based approaches as a means of teaching business specific vocabulary. For example, common topics found in these coursebooks are: corporate information, management and leadership, salary and compensation, marketing, finance and business interactions between customers and suppliers. Some courses are also industry specific such as using business English in the hospitality industry, IT, and tourism. These courses can contain many different types of technical words and acronyms not found in general English.

Materials published for Business English generally come in two forms, coursebooks for teaching language and coursebooks for the business skills training market. Coursebooks for teaching Business English share some common traits besides the topic lists mentioned above. Business English coursebooks are based on a syllabus with a situational, functional or structural main focus and aim to provide enough material to not need any supplementation by the instructor. Many coursebooks come with CDs having videos, audio files, and other computer-based programs and/or reference books like dictionaries, workbooks, and answer keys. (Ellis et al. 1994: 117-122)

Business skills training materials are usually for company training programs or business schools. They do not focus on learning the language, but learning the business knowledge in a specific industry. According to Ellis et al., "they are valuable for trainers and learners who have not had business experience because they provide insights into typical business situations, behaviors, and procedures."(Ellis et al. 1994: 122)

There are two main varieties of Business English. This is important when materials are chosen. The first variety is pre-experience. Learners of the pre-experience variety are described by Ellis et al. as:

> Students in colleges or universities will have gained their knowledge of business largely from books and, as a result, such knowledge will be incomplete and theoretical rather than practical. They will be less aware of their language needs in terms of communicating in real-life business situations, and their expectations of language learning will be molded by their experiences from school, and thus by the educational policies of the country in which they grew up.

(Ellis et al. 1994: 5)

The second variety is job-experienced learners that:

> ...will also be influenced by their educational backgrounds, but they will, in most cases, have gained some practical experience of having to communicate o the job. The experience has the effect of focusing their attention on what they perceive as their own short comings in terms of fluency, getting the message across, and being able to understand the people from the other countries that they have to deal with.

(Ellis et al. 1994: 5)

The learners of these two varieties have different needs that greatly affect syllabus planning and material selection. Pre-experienced learners may need to use coursebooks and lecture to develop all of their language skills and need to be prepared for working in the future. They may also be preparing for exams. Job experienced learners may have specific needs in order to improve their capability that can be met through learning the use of the language and the theoretical knowledge about the language. These learners will also be subject to the needs of their organization that has most likely set guidelines to content being taught to their employees. Additionally, if the intended use of Business English is for academic purposes either type of learner will need to be familiar with yet another special sub-characteristic of Business English. Because of the differences between these two varieties, the Business English course will need to have different objectives, course content, and teaching methodology. (Ellis et al. 1994: 5-7) Business English carries its own set of components in

15

order for it to function as a "lingua franca of international business and communications" "(Chujo et al. 2004: 1) were in learners sources of input and requirements for output are determined by the business environment. Some significant features of three business oriented vocabulary sets: "monthly textbook, used in a radio Business English program-"NHK Business Eigo", TOEIC practice tests, and the British National Corpus (BNC) spoken data from business meetings, consultations and interviews" (Chujo et al 2004:4) was used to show that, in regards to business ELF in Japan use and the input being given to those learning English, English textbooks did not cover enough vocabulary. In regards to further research, the authors stated:

> It is hoped that the observation of both spoken and written business-oriented vocabularies using the same approach suggested in this paper will further contribute to preparing a steady and effective method for meeting the increased need for pedagogical materials, as well as address the demands of a rapidly growing global business community.
>
> (Chujo et al.2004: 12)

As stated earlier, BELF has vocabulary that is unique to the business environment. This vocabulary can be studied and taught in the same manner with the same corpus-based techniques as vocabulary in general English.

2.4 Domain and Register

Business English or any other ESP require attention to be paid to the specific language and situations for that language to be used. Genre, text type and register are some ways of classifying texts used in corpora as well as in language in general. Whereas domain is the specific subject that the texts are applied in. Baker explains the differences in genre, text type and register as stated by Biber:

Biber (1998: 208) uses the term 'genre' for classes of texts that are 'determined on the basis of external criteria relating to author's or speaker's purpose'. On the other hand, he uses 'text type' to refer to classes of text that are grouped on the basis of 'similarities in intrinsic linguistic form, irrespective of their genre classifications'. For example, particular texts from press reportage, biographies and academic prose might be very similar in having a narrative linguistic form, and thus would be grouped together as a single text type, even though they represent different genres. Register (Biber 1998: 135) refers to variation in language arising from the situations it is used in, and depends on such things as its purpose, topic, setting, interactiveness and the intended addressee. Registers may be highly specific, such as novels by Jane Austen, or the 'Methods' sections of academic papers describing biological research.

(Biber cited in Baker 2009: 173)

Cultural-level factors have influence on the communication process just as individual and universal factors do. (Ammon et al. 2006: 807) However, domain and register should not be confused with dialect which refers to language varieties grouped together by geographic region, education level, social class, gender, etc.

Different domains used by an individual person are not necessarily equal in frequency of use (Boxer 2002: 211). In ESP, learners must adjust to a new balance between their domains as their use of the new domain becomes more dominant. This adjustment can be compared to the accomplished feeling a learner gets when he or she achieves a higher understanding or level of use in their non-native language. In the teaching of new domains in text form, it is important that the texts carry some sort of cohesion through vocabulary, genre, etc. This is known as "a set of domain-specific disciplinary genres" (Bhatia 2004, cited in Bremner 2008: 307). In order to "teach students the rhetorical and cultural significance of one genre will require teaching the significance of its genre set and the place of that genre within that set" (Devitt 2004: 188, cited in Bremner 2008: 307) so that a clear view of the domain can be understood. According to Campoy et al. (2010: 22), "The compilation, description and analysis of domain-specific corpora is one of the widest areas of research in corpus linguistics, especially as regards academic and professional settings."

Business English itself is a complex domain requiring interaction between many different registers and sub-domains especially in an international setting (Camiciottoli 2010).

Business communication coursebooks specifically have been looked at in terms of their texts and intertexuality functioning in the appropriate register (Bremner 2008). The importance of this type of connection between texts helps to create a more true to life domain in the coursebook.

2.5 Vocabulary

There are many aspects to vocabulary and vocabulary learning/teaching in general. When thinking about vocabulary in a general or ESP context, it is important to understand that "Vocabulary knowledge enables language use, language use enables the increase of vocabulary knowledge, knowledge of the world enables the increase of vocabulary knowledge and language use…". (Nation et al.1997: 6)

Vocabulary is complex in the sense that one must understand a variety of characteristics about the word in order to use it and understand it correctly. A person understands a word in two forms. The first, receptive knowledge, is used when someone simply understand the word when listened to or read. The second is productive knowledge which is when a person speaks the words themselves. Generally, people learn a word receptively first and later productively. There are also various "word knowledge" that needs to be learned about each word. These are: the meaning(s) of the word, the written form of the word, the spoken form of the word, the grammatical behavior of the word, the collocations of the word, the register of the word, the associations of the word, and the frequency of the word. (Schmitt 2000)

The nature of vocabulary acquisition is incremental through repeated exposure to the word. This idea can be applied to both spoken and written vocabulary (spelling). Some already acquired vocabulary can be forgotten from lack of use or exposure. This is called lexical

knowledge attrition. (Schmitt 2000: 4-6) Also, "the findings of a casual relationship between reading and indirect vocabulary learning in EFL context is consistent with research into vocabulary learning by children in their first languages." (Day et al. 1991: 545) which suggests that the vocabulary content in texts being read has impact on a learners ability to acquire vocabulary.

Teaching vocabulary using corpus techniques is another method that helps learners learn and teachers teach. One way is to use corpus information to teach better is by studying corpus generated language elements as a way to better inform teachers of the concepts they are teaching. As Schmitt (2002) explains, a teacher can take a form of one vocabulary word or words and see how its use changes depending on what form of the word is being used. Another way to use the corpus information is to provide students with concordance lines, fragments of sentences around a common vocabulary word or phrase, as a means for the students to come to their own conclusions about how the word of phrase is used. Corpus techniques can also be used for vocabulary assessment. By inputting student work into a corpus program, the teacher can see the range of vocabulary the student uses as well on the sophistication of the vocabulary used. (Schmitt 2002: 31-41)

Vocabulary itself can carry significance by means of word choice. Gendered words such as "her" and "him" can show how gender is constructed in society. According to Butler, the use of "gender-acts", which are similar to sociolinguistic "speech-acts", give meaning through shared "collective action" in society. (Butler 1998: 525) Therefore, the use of "she" and "him" define and confirm gender-acts that are considered normal in a particular society. Word choice can also show collectivism or individualism in society. Collectivism is a type of social outlook that values society's members as a group more than an individual and vice versa for individualism. Talja et al (2005) discuss how domain is affected by collective nature of certain professional fields such as business management and educators. This collective

19

nature is established through the expectations of the individual as imposed by the other members of the society. It is demonstrated externally through actions and speech as well as internal thought processes. (Talja et al. 2005: 86)

2.6 Corpus and Linguistics

Corpus study can be described as follows:

> Corpora or corpuses (singular: corpus) are simply large collections of databases of language, incorporating stretches of discourse ranging from a few words to entire books. The exciting thing about corpora is that they allow researchers, teachers, and learners to use great amounts of real data in their study of language instead of having to rely on intuitions and made-up examples. Insights from corpus research have revolutionized the way we view language, particularly words and their relationships with each other in context.

(Schmitt 2000: 68)

Corpora can be made from any texts or spoken transcriptions. Depending on the aims of the researcher, genre, typicality, period of time samples are taken from and the overall size of the corpus can vary. (Sinclair 1991) There are four essential characteristics of corpus based analysis. First, it can analyze the actual patterns of use in natural text, second, it uses the corpus as a basis for analysis , third, it is conducted on computer programs, and last, it can require both quantitative and qualitative analytical techniques. There are two ways to look at linguistic patterns in a corpus. On one hand, lexical associations can be investigated to see how it is systematically associated with other words. On the other hand, grammatical associations investigate the grammatical features in a specific context. (Biber et al. 1998) The three main research functions in a corpus computer program are wordlists, collocation and concordance. Wordlists are a list of words that show the frequency that the word or only its lemma occurs in the corpus. Frequency lists can be ranked in terms of number of occurrences. In corpus, word occurrences are called "tokens". Collocation is when a word is researched to show the other words it is commonly positioned by. In other words, the group

of words that appear together frequently. There are two main types of collocations, grammatical/syntactic and semantic/lexical. Lexicographical use of corpora is used for input in dictionaries such as the Longman Dictionary of Contemporary English and the Oxford Advanced learners Dictionary. (Schmitt 2000: 68-82) Concordance is a set or example of a given word or phrase showing the context. (Scott 2010) Corpus research can show you empirical data, however the inference of meaning, etc. it entirely up to the researcher.

Discourse analysis can also be carried out using corpora. Four types of studies using corpora have been identified to examine discourse attributes: 1. examine a feature of language in use, 2. realizations of a function of language, 3. variety of language and 4. tracking an occurrence of a linguistic feature throughout a text. (Conrad 2002, cited in Charles 2009: 1) Using corpus based discourse analysis has the benefit of allowing for generalizations across a large volume of texts to be made systematically where as in typical discourse analysis, large scale generalizations across registers is not typically possible.

Corpus can also be a tool for individual learning. According to Tan (2002: 3), "Learners can use the evidence in corpora for introspection. Introspection is seen to a behavior desired from learners where learners are viewed as active participant of language from a textual and discoursal perspective." Learners can also become aware of unfamiliar usages of language and investigate and experiment with vocabulary on their own. Learners can also "induce language rules and regularities for themselves". (Schmitt 2002: 33)

2.6.1 Corpus Based Studies

The number of BELF corpus studies is still relatively low. One corpus of note was used to compare the letters of request in the Thai business context. In this study, Chakorn (2006) looked at 80 authentic letters of request written in English by Thai speakers and native English speakers. By using both quantitative and qualitative investigation it was determined that persuasive strategies and politeness strategies were different from the native English speakers letters. Some of the reasons for this were found to be culture bound. Thai letters particularly showed a "delayed introduction of purpose" and "use more negative politeness in that they include more indirect, deferential and self-effacing strategies". (Chakorn 2006: 103)

Currently there are several ELF type corpora. The International Corpus of English (ICE) and the Vienna Oxford International Corpus of English (VOICE) are two LF type corpora. According to the ICE website:

> ICE began in 1990 with the primary aim of collecting material for comparative studies of English worldwide. Twenty research teams around the world are preparing electronic corpora of their own national or regional variety of English. Each ICE corpus consists of one million words of spoken and written English
>
> (http://ice-corpora.net)

Additionally, ICE's creators believe in the theory of "World Englishes" and have created their corpus to have the ability to separate different World Englishes from each other. Similarly, the VOICE corpus follows the same LF beliefs stating:

> In the early 21st century, English in the world finds itself in an "unstable equilibrium": On the one hand, the majority of the world's English users are not native speakers of the language, but use it as an additional language, as a convenient means for communicative interactions that cannot be conducted in their mother tongues. On the other hand, linguistic descriptions have as yet predominantly been focusing on English as it is spoken and written by its native speakers. **VOICE** seeks to redress the balance by providing a sizeable, computer-readable corpus of English as it is spoken by this non-native speaking majority of users in different contexts.
>
> (http://www.univie.ac.at/voice/)

An additional LF corpus that is based in Asia is the Asian Corpus of English (ACE). The ACE page in the University of Hong Kong website states:

> The major role of English in Asia today is as a lingua franca. English is the de facto lingua franca of the grouping of Association of Southeast Asian Nations (ASEAN) and with the signing of the ASEAN Charter in December 2008 will assume official status as the working language of ASEAN. English is also the working language of the extended grouping known as ASEAN + 3, which includes the ten states of ASEAN plus China, Japan and Korea… Our aim is to collect a corpus of one million words of naturally occurring, spoken, interactive data of English being used as a lingua franca in Asia.

<div align="right">(http://www.ied.edu.hk/rcleams/view.php?secid=227)</div>

Another corpus, the Hong Kong Corpus of Spoken English as an addition to the HKCSE, purpose is to enrich the intercultural resources available for learners and teachers. Some studies based on this corpus are in the areas of discourse analysis, intercultural pragmatics, pragmatics and intercultural communication with a focus in discourse intonation. (Cheng 2005: 47) This research is significant because it gives a data backed view of the authentic use of English as a lingua franca for application in English language teaching materials and other types of research. They have used actual "lingua franca English" to create this corpus and thus made a data source that can accurately represent the specific uses of English in different areas including Business.

It is important to note some specific implications of this study. First, this corpus in made of intercultural speech. This is because researchers understand that Hong Kong, like many other areas in Asia that use English when conducting various types of business, realizes that Hong Kong itself is composed of people from many different backgrounds. All of these different backgrounds shape the use of English within the context of the specific area.

Using the BLC (1 million words Business Letter Corpus) and comparison to the 4000 Basic Words List as defined by JACET (Japan Association of College Teachers researchers found

that English for Business Purposes (EBP) is characterized by a high degree of lexical closure (15,000 to 20,000 words range). This means that a large percentage of word types are represented by a relatively low amount of all occurrences.

Also, EBP is characterized by a low level of lexical difficulty. This is shown by "…77.5% of the first 3,000 word types in the BLC, which cover over 95% of all BLC word tokens, are within the 4000 Basic Words as defined by JACET." (Someya 1999: 2) Many of the words in written business English seem to incorporate spoken English features because people seem to "write as they talk". There is also a connection between common errors in Japanese users of EBP with high frequency lexical items which suggests that something is "going very wrong with the way these lexical items are taught in the classroom." (Someya 1999: 3) Further research is encouraged by this study:

> The discussions are not exhaustive by any means, but enough to substantiate the claim that Business English is a "sublanguage" with its unique lexico-grammatical patterns and that the identification and description of which in a systematic way will help the learners of EBP to learn what need to be learned more effectively than it used to be.
>
> (Someya 1999: 3)

A corpus created from e-mail distributed news articles from *The Guardian International* starting in October 2006 and ending in October 2001 was developed to analyze the text from a historic and cultural standpoint. Tribble (2002) approached texts by means of using wordlists and keyword lists, identify words of phrases that have prominence in the texts and assess their salience, use collocation and clustering to elaborate on significant features, and re-connect the analysis with the texts.

Using keyword lists can give instant access to changes in the national agenda and a list of words and people that "mattered" during a specific time period and a specific context of culture (Tribble 2002: 153-154). Corpus investigation in newspapers can give learners higher

understanding of concerns of a news organization or when compared across international newspapers, corpora can enrich areas of cross-cultural study.

Evaluating the validity of corpus-driven information for ESP courses has already been shown to be a plausible application on corpus data. "Corpus data enables the realization of important language functions, which 'are widespread and consistent in usage, and most importantly, meaningful.'(Guest 1998: 31, cited in Fuentes 2002: 6) and that contents of a corpus should be consistently revised for effective integration in the ESP setting.

The lexical analysis of written business science and technology texts directed towards word behavior acquisition and their implication in regards to use of lexical chunks during tasks showed that "lexical command is highly significant...as learners realize the need of specific words to suit specific contexts." and "Good linguistic skills demonstrate lexical knowledge in the tasks, while subject competence contributes to acceptable performance.". (Fuentes 2002: 28)

2.7 Corpus Based Study of Lexical Bundles

The study of lexical bundles using corpora is useful in understanding the writing or speaking styles of different registers. The following studies have shown that the use of bundles and the composition of the bundles themselves give insight into the attributes of learners of English as a second language writing and speaking styles. Research of lexical bundles also serves as a tool to differentiate registers. When specifics about a particular register and its lexical bundles are determined that information can be used to enhance teaching methods.

Biber et al (2004) researched lexical bundles in university teaching and textbooks in order to find the theoretical status of lexical bundles using corpus data. They decided to use a frequency-driven approach to multi-word units, based on analysis of the most frequently

occurring lexical bundles. They then compared the patterns of use in classroom teaching and textbooks to those found in conversation and academic prose in order to find unique distinctions and similarities. The corpus was made from texts from classroom teaching and textbooks in six major academic disciplines: business, education, engineering, humanities, natural science and social science over three levels of education: lower division undergraduate, upper division undergraduate and graduate. The results of their analysis suggested that lexical bundles "should be considered basic linguistic constructs with important functions for the construction of discourse" and "lexical bundles differ dramatically from other linguistic features". (Biber et al 2004: 398 – 399) The researchers also found that lexical bundle patterns of use show that they are very different from other kinds of linguistic constructs and productive grammatical constructions.

The researchers also found a significant lack of lexical bundles in the textbooks. According to Biber et al,

> Its surprising that textbook authors do not incorporate more lexical bundles in their writing, given the heavy reliance on bundles in classroom teaching. Reasons for this absence might be that textbook authors tend to use fuller expressions, preferring full clauses rather than phrasal lexical bundles, as well as the fact that textbook authors are free from real-time production constraints of face-to-face teaching and therefore make more diverse language choices.

(Biber et al 2004: 383)

In conclusion, this analysis of lexical bundles showed that there is significant potential to gain new linguistic knowledge about language features by studying lexical bundles using corpus methods. It also showed that lexical bundles are unique depending on the register they exist in. The final statement from Biber et al explains that "this study illustrates how an exploratory corpus approach facilitates the identification of language features that would go unrecognized otherwise, but that turn out to be a fundamentally important part of writers' and speakers; communicative repertoire" (Biber et al 2004: 400)

In a study of advanced academic writing of Chinese EFL learners comparing a corpus of doctoral dissertations to published articles by professional writers found that significant differences of lexical bundle use occurred between the two corpora. The advanced learners of English used more lexical bundles and more types of bundles in their academic writing than the professional writers. Also, after the researcher reviewed the structure of the bundles that were in use, it was determined that both the learners and the professional writers used the same amount of prepositional phrases, noun phrases, be + noun/adjective phrases and more. But, the learners relied much more on passive structures where the professionals used more anticipatory bundles. The learners and professionals used the same amount of research oriented and text oriented bundles but the learners used less participant oriented bundles. Research oriented bundles "help writers structure their activities and experiences which include bundles referring to location, procedure, quantification, and topic of research." (Wei and Lei 2011: 161)Text oriented bundles use bundles for transition, results and framing. The researchers came to the conclusion that learners preferred to keep their academic writing more impersonal. The researchers also noted that, contrary to other studies, lexical bundles were not easily acquired in the natural course of learning a language. This means that specific teaching of lexical bundles is required to some extent. (Wei and Lei 2011)

Another study that uses both spoken and written data for corpus study from university registers such as teacher instruction and student advising/management (such as written syllabi) found that lexical bundles are more prevalent in non-academic university registers than they are in core instructional registers. This research also found that contradictory to other research in lexical bundles; lexical bundles themselves were more common in written registers than spoken ones. The researcher was also able to generate a functional framework for distinguishing three major types of bundles: "stance bundles" that show attitude or an assessment of certainty, "discourse organizers" that signal relationship between prior and

coming discourse, and "referential expressions" that identify entities or specific part of entities. By using this framework, the researchers were able to see that each register actually used distinct types of lexical bundles. (Biber and Barbieri 2007)

Research done by Cortes (2002) also showed the potential for register-distinct bundles. This study focused on examining a corpus comprised of University Freshmen compositions. Contrary to other studies that assumed students would write closer to conversational prose, the students used lexical bundles more closely related to academic prose. That is to say the bundles were more nominal or phrasal than clausal showing a closer match to academic registers. Further research into the context of these bundles showed that although the students' bundles shared more commonality with academic prose, the bundles themselves were not as varied as true academic lexical bundles. The bundles were mostly just for location and temporal markers that can be attributed to the instructional tasks designed for the particular composition the students were to complete. So, the task itself had great influence on the lexical bundles produced by students.

A study by Cortes (2006) aimed to take a direct approach at teaching common register specific lexical bundles to university students and evaluating that teaching using a corpus of the students' writing. Cortes designed a mini-training course as part of a writing course taken by native speakers of English. The mini-course taught a group of functionally related lexical bundles similar to the way that collocations are commonly taught. The results showed no major improvement in the frequency or variety of use of lexical bundles. Instead of using the bundles, students still relied on simple conjunctions, conjuncts and adverbs to express things that professional authors would use lexical bundles for. The training courses were only successful of raising the students' awareness of lexical bundles and their interest in learning them. Cortes (2006) does suggest that the fault may be in the mini-training course itself. The students may not have had enough time to acquire the lexical bundles into their actively used

knowledge. They may also need more exposure to lexical bundles in context, perhaps through the use of corpus based instruction, in order to use them frequently and correctly in their own writing. This shows that there is still a possibility of actively teaching lexical bundles to students given more frequent exposure and better conducted instruction.

Hyland (2008a) has also come to the conclusion that lexical bundles are register specific and should be taught according to the particular field that students are most concerned with. Hyland used three corpora of data that could be separated into four distinct disciplines: Electrical Engineering, Biology, Business Studies, and Applied Linguistics. When comparing the most frequent four-word lexical bundles from each discipline, Hyland found that "over half of the items in each list do not occur at all in any other discipline and only 30 per cent of the strings in each discipline are found in two other fields." (Hyland 2008a: 12) Based on these findings, Hyland expresses that increased pedagogical focus on bundles is needed and that there is no single core vocabulary list that can be applied across all academic study. Furthermore, "the best way to prepare students for their studies is not to search for universally appropriate teaching items, but to provide them with an understanding of the features of the discourse they will encounter in their particular courses." (Hyland 2008a: 20)

Hyland (2008b) continued his research into lexical bundles by focusing research in the areas of academic writing across different disciplines as well as from different writing genres in order to further show the importance of lexical bundles in student's writings. His motive for this research is based on the belief that,

> ...there may be advantages to identify these clusters so as to help learners acquire the specific rhetorical practices of the texts they are asked to write. In order to accomplish this, writers need a familiarity with both clusters which characterize their disciplines and those which are valued in the particular genres of those disciplines.

> (Hyland 2008b: 42)

Hyland used research articles, PhD dissertations and MA/MSc theses from Electrical engineering, Business studies, Applied Linguistics and Microbiology disciplines as data for his three corpora. He found that the lexical clusters for each genre (research, PhD, and MA/Msc) were different as well as across disciplines. He also found the audience that the written document is intended for greatly influence the use of lexical bundles. In regards to genre, Hyland found that "Clusters, in other words, should not be regarded as a basic linguistic construct; their distribution can also be an effective way of characterizing genres with in a single register." (Hyland 2008b: 59)

He has also made some strong implications for pedagogic practice. First,

> ...evidence from the learner corpora help improve descriptions of the target language and provide more realistic models for students. It alerts us to the need to understand the kinds of text our students need to write rather than rely on the massive literature which describes the research article.
>
> (Hyland 2008b: 60)

Second, "learner corpus data can play an important role in the selection, sequencing and structuring of teaching content." Third, "teaching materials can benefit from the findings of cluster research in different genres, allowing teachers to focus on the specific ways of creating meanings appropriate to particular kinds of writing." And finally, "relevant genre data refocuses instruction on form, and can provide a basis for methodological practices which involve data-driven learning." (Hyland 2008b: 60)

2.8 Development of Corpus Based General English Materials

There are many coursebooks and other teaching material available that are influenced by corpus data. These materials are designed using one of two approaches, "modeling" and "corpus-driven". (Ranalli 2003: 5) Modeling is a less corpus dominated approach. It can include using corpus data to evaluate existing material such as the degree that vocabulary

content in coursebooks correlates to vocabulary used in "real-life" by using word lists or collocations. The findings can then be used to create better syllabus designs. An example of this is a corpus-based lexical coverage analysis of college level general English coursebooks in Taiwan. Hsu was able to determine, based on the corpus, the "importance of considering the appropriateness of book levels when choosing college GE [General English] textbooks, especially when a student's vocabulary size has reached a certain level." (Hsu 2009: 59)

The "corpus-driven" approach uses corpus data to strongly change syllabus design. Ranalli states, "in this view, CL [Corpus Linguistics] are not simply a contributor or arbiter of syllabus content but the primary source." (Ranalli 2003: 7) In other words, the syllabus is not based on situational context or academic requirement but the relationships between words.

In addition, materials corpus can be used to determine if the contents (text, language, etc.) used in a coursebook model that of a native speaker or, in the interest of the researcher, a speaker of BELF. Text in coursebooks can be questioned in terms of language used, text types, and even authenticity of tasks that can be used to improve levels of authenticity. (Campoy et al. 2010: 27) Research from Partington (1998, cited in Camiciottoli 2010: 115) using a corpus made from newspaper texts "showed that if-constraints were much more varied and complex in this authentic usage with respect to traditional forms presented in most ELT materials." Research such as this suggests that the content materials developers and teachers deem "native speaker" or "authentic" may not actually be completely true.

In order to create corpus-modeled or corpus-based materials, material developers should know what corpus data is appropriate and how to incorporate it into the materials to create the most benefit for learners. Generally, what to teach can be determined by three categories, the "situation" the language will be used, the "functional or pragmatic use of the language" and the specific" lexico-grammatical features of the discourse." (Simpson-Vlach et al.2006: 268)

Once these items are determined, the materials developer can use the corpus accordingly. Another aspect to corpus-based materials is "consciousness-raising (C-R)" (Ranalli 2003: 10) This is another form of inference induced learning but not from a set of texts. C-R takes place when learners are given a corpus created list or collocation and left to discover the meanings and grammar rules themselves.

2.9 Corpus and Business English Materials

Corpus has been used as a base for General English coursebooks. The reason to use the corpus is to have factual backing for the content in the cousebook. Due to the differences between general corpora and English for specific purposes (ESP) corpora, ESP coursebooks may not accurately represent the language that people are actually using. This is especially so if one is creating Business English materials with a focus on the local BELF domain. Some of the differences can include glossing over of important aspects of usage and variation or leaving out less frequent stylistic choices for more common ones. This makes the materials they are presenting very misleading. Domain specific corpora can make materials that fit the area of application they will be used in which addresses particular student needs better. (McEnry et al. 1996: 120-121)

Corpus has successfully been used to create Business English materials such as worksheets, syllabi, and vocabulary lists by incorporating corpus driven knowledge to understand the relationship between academic corpus-based data evaluation and corpus material design. Corpus created material reflects "…1.The need to make EAP teaching material reliable in terms of novelty and content reliance, and 2.The need to provide compatible exploitation of such content according to corpus-driven itemization. (Fuentes et al. 2002: 3)

Corpus based materials can also account for changes in language, language change occurs in accordance with the evolution of the business industry and cause linguistic problems for

learners. (Fuentes et al. 2002: 4) In addition, corpus can act as a content reference for Business English schematic knowledge growth and to raise student motivation giving students the opportunity to notice content relationship in language use.

Communicative tasks based on the corpus data can make the tasks themselves more effective because of the higher exposure to relevant vocabulary. Fuentes et al. notes that "the aim lies in exposure to the key language during both the project work and corpus-driven classroom activities, in order for learners to build up suitable lexical competence for BIT [Business Information Technology English]." (Fuentes et al. 2002: 9)

Business English coursebooks are similar to general English coursebooks with a heavy focus on situation-based structures. As with any coursebook, "a coursebook must perform in the context which it will operate" and should be looked at "in what way and to what extent it addresses: teachers' needs, learners' needs, syllabus outcomes/guidelines, publishers' needs, and writers' needs."(Wala 2003: 58)

2.10 Conclusion

The realities of BELF is deserving of being researched and analyzed in the same way that general English has for the past few decades. The nature of vocabulary and acquisition shows a strong need for coursebooks to use the most useful vocabulary in the correct context. Sometimes in general English, this vocabulary has been taken from a corpus, but Business English coursebooks seem to not have the same opportunity. Research into lexical bundles using corpora has proven that lexical bundles are register distinct and that the teaching of specific bundles can aid greatly in the teaching of English. Business English materials, especially ones that could be developed specifically for use in Asia international hubs like Hong Kong, India, or Singapore are teeming with the use of English by a high proportion of

non-native speakers, or, described another way, speakers from the outer and expanding circle (Kachru 1985) would benefit greatly from corpus-based development approaches.

The characteristics in Business English vocabulary that have already been found in western studies should be apparent in BELF used in Thailand as one study uncovered. The Hong Kong Corpus of spoken English has touched on the uniqueness of ELF in Asia, but not in a strictly business domain. A corpus of Business English used in Thailand could enrich the intercultural resources for teachers and learners indefinitely. Similarly, a gap has been discovered between Business English lexical bundle use and teaching methods. What this gap is exactly requires a comparison of materials and to the corpus information in order to find the extent of the issue.

Tribble's (2002) research from corpora using newspaper articles input shows the words that matter most in a specific context of culture and thus can enrich cross-cultural studies could reflect the vocabulary needed in Bangkok's Business English environment and in turn what should be taught in the classroom. Using corpus-based data in a business domain to investigate language and content needs of the learners has already been shown as an effective method to assess communicative tasks and can be expanded to assess vocabulary.

Corpus based research into lexical bundles found unique distinctions in bundle use depending on the register or genre they are being used in. There is also research suggesting that coursebooks show a noticeable lack in lexical bundles. Those who are either learning to write in English or simply learning to write according to their genre tend to you less and different bundles than professional writers. Also, lexical bundles are found more frequently in written English than spoken which contradicts popular belief. The importance of lexical bundles in written texts also suggests that further investigation into teaching lexical bundles explicitly should be considered.

The influence of these studies on this research comes in many forms. The most significant ones support the base concepts of BELF, cultural influence on registers, and the existence of register specific vocabulary and lexical bundles as well as the importance of exposure to them. Mollin (2006) recognizes the significance of BELF as being a language of different types of speakers from different linguistic cultural backgrounds. Ammon (2006) also came to the conclusion that culture level factors have great influence on communication. A study of Thai business letters written in English by Chakorn (2006) found strong evidence that major differences exist in Thai BELF English and western English. Specifically, delayed introduction of purpose and indirectness showed clear cultural based motivation on language use.

The Hong Kong Corpus of Spoken English made from a collection of ELF data and has already showed great potential in the study of ELF. Olivera et al (2003) found that Business English has its own specific register. Chujo et al (2004) came to the same conclusion and found evidence that Business English textbooks don't cover enough specialized vocabulary. The specialized vocabulary, she believes, is important input that learners need in order to produce appropriate output. Hyland (2008a) and Cortes (2002) found that registers have their own distinct lexical bundles. Biber et al (2004) went further to investigate lexical bundles in corresponding coursebooks and found that there was a lack of lexical bundles. In carrying out the data analysis for this study, the researcher followed Tribble's (2002) method. First, the researcher identified words and lexical bundles of prominence and then elaborated on them in the text themselves. The researcher will also explore the Thai BELF corpus as a means to find linguistic features that would otherwise go unnoticed using conventional sociolinguistic methods (generally without corpus use) and typical non-culture based corpus linguistics.

These studies have shown that corpus-based research can be an effective means in understanding language and can be used to create and evaluate course materials. Currently, there has been no attempt to formally, corpus-based or not, assess the Business English coursebooks used in Bangkok in terms of vocabulary or lexical bundles. Because the English used in business communication is always changing, materials need constant updating to be a reliable source of input to learners. The next chapter will go into detail regarding the methodology of this study.

Chapter 3

Research Methodology

3.1 Introduction

This Chapter will discuss the research methodology of this study. First a restatement of the research questions is presented followed by details about method of data collection, data analysis including the coursebooks that are evaluated. A Pilot study is also included in this Chapter.

3.2 Restatement of Research Questions

1. What are the most frequently used BELF lexical bundles and their attributes in the business pages of the *Bangkok Post* and *The Nation* newspapers?

2. To what extent are the most frequently used lexical bundles in the Thai BELF corpus represented within the *Intelligent Business*, *Essential Business* and *New Business Matters* coursebooks available in Thailand (MAS corpus) and the *BG 2000 English III* and *BG 2001 English IV* coursebooks used at Assumption University (ABAC) in Thailand (LIM corpus)?

3.3 Development of the Corpus

The corpus is created in the program WordSmith published by Lexical Analysis Software Ltd. and Oxford University Press. This program uses text data to perform research functions such as generating word lists, collocations and key words. Business texts in the form of news articles are taken from the *Bangkok Post* and *The Nation* online newspapers. The researcher selected articles from March 2011 to July 2011. The corpus is made of approximately 800,000 words from almost 2,000 newspaper texts.

Texts are taken from *Bangkok Post* and *The Nation*'s business section published online Monday through Friday. *The Nation*'s business section has six sub-categories: Macroeconomics, Investment and Finance, Property, ETC, Corporate, and Management and Innovation. The *Bangkok Post's* sub-categories are: News, Economics, Telecommunication, Tourism, Aviation, Retail, Marketing, and Financial Advice. Approximately ten texts of around 400 to 600 words each has been selected from each newspaper daily and saved in text file format for compilation in the WordSmith program. The content of the texts chosen include the article content and title but exclude and advertisements, etc. that is also on the webpage.

3.4 Method of Data Collection: Coursebooks

The coursebooks, which are available at Chulalongkorn University in Bangkok and used at Assumption University, are chosen using a number of criteria. First, in regards to the coursebooks from Chulalongkorn University, an international focus is important because these books are being used in non-native English speaking environment that is multinational. Comparing news stories using BELF to international coursebook vocabulary is likely to even the ground when comparing the corpus data to the coursebooks. The mini corpus made from these books is called MAS (mass produced). The books used at ABAC are the highest level of English before being subdivided into different business focuses such as Hospitality and Music Industry. These two coursebooks include articles written in a business context for informational purposes. The mini corpus for these books is called LIM (limited production).

Books published in the last five to six years, 2005 (*Intelligent Business*), and 2004 (*New Business Matters*), 2011 (*Essential Business Vocabulary Builder*), are more relevant than books published earlier because they represent a range of books used today. Since some schools still use older books, books from the undergraduate Business English required

courses at Assumption University, *BG 2000 English III* and *BG 2001 English IV*, are also included. These books are currently used in classes and represent the actual learning situation in a major University in Bangkok.

This research uses books from the two levels, Intermediate (*Intelligent Business* and *New Business Matters*) and Pre-intermediate to Intermediate (*Essential Vocabulary Builder*, *BG 2000 English III* and *BG 2001 English IV*), in order to get a more complete view of vocabulary use. The levels of English books in general are determined in a more subjective way rather than by using concrete criteria. For this reason, some pre-intermediate books from one source and intermediate books from another seem very similar. Using this level pairing should account for level variation within the individual books.

The five books have a focus in written text vocabulary. It is important to have books that focus in vocabulary since this research uses vocabulary-based analysis. If the books didn't have a specific vocabulary focus, the research would be, in a way, comparing apples to oranges. Specifically, if the authors of the coursbook intended the book to practice speaking, then their focus when designing the book would not be written text vocabulary, but vocabulary used in utterance.

Each coursebook has its own description on the back side of the cover. *New Business Matters* is described as:

> …a stimulating language course for students of Business English. Its unique lexical syllabus precisely identifies what business students need to learn in order to increase their understanding, fluency, and communicative power in English. Each unit is based on a specially written article rich in the language of company life, followed by language activities which ensure that learners engage with the language and interact with each other.

> (Powell et al. 2004)

Intelligent Business is described as a topic-based coursebook with "key concepts of today's business world". It "uses up-to-date authentic material from the Economist". The authors also

39

claim it is "fully benchmarked alongside the Cambridge BEC exam suite". (Trappe et al. 2005) *Essential Business Vocabulary Builder* states:

> ...is a business vocabulary and practice course for students with a pre-intermediate level of English or above....the vocabulary is organized by business topic (first part), business skill (second part) and word family (third part), making it simple for you to target the language and skills you need....the topics and skills are covered from a 21st-century perspective, teaching up to date language in modern business contexts.
>
> (Emmerson 2011)

BG 2000 English III states in the introduction:

This latest edition of the English III textbook has been revised to give strong emphasis to reading comprehension and understanding of written texts. The purpose is for students to be able to discover and explain the meaning of news and technical reports in their own words, without blindly repeating poorly understood details.

(Kumar et al. 2010b)

BG 2001 English IV states in the introduction:

It is projected then that the reading passages used in the textbook will function as models on which writing can be based as appropriate. The passages are to aid reading inside and outside the classroom and give input for the development of writing skills.

(Kumar et al. 2010a)

This study uses an article, written for informational purpose, from each unit of *Intelligent Business*, *New Business Matters* and 15 relevant units from *Essential Business Vocabulary Builder*. *Essential Business Vocabulary Builder* has many more units however they consist of non-article topics such as how to write formal letters and speaking. All articles in a business context from *BG 2000 English III* and *BG 2001 English IV* arealso included. A list of unit titles and articles is below:

Table 3.1: Units/Articles for Analyses

Intelligent Business	New Business Matters	Essential Business Vocabulary Builder	BG2000 English III	BG2001 English IV
1. Companies	1. Career Management	1. Company types	New Threats	Avatar is the Highest Earning Movie
2. Leadership	2. Enterprise	2. Start-up and Growth	Change in Thai Shopping Habits	The Global Business Leader
3. Strategy	3. E-business	4. Individual Company Profile	Protests over the Building of Superstores	China-made Toy Recall
				Food Demand
4. Pay	4. Brand Management	6. Markets and Competitors	Telework	Subway Sandwich Business
5. Development	5. Prices and Commodities	7. Marketing-the four P's	The End of Light Bulbs	The Galactic Suite Space Resort
6. Marketing	6. Corporate Entertaining	8. Marketing Strategy and Brands	Put Your CV to Work for You	"Le Whif" Business
7. Outsourcing	7. Innovation	10. Customers		Legal Career
8. Finance	8. Public Relations	11. Orders, Invoice and Payment		NTUC Fairprice
9. Recruitment	9. Cultural Awareness	12. Money		Public Companies vs. Private
10. Counterfeiting	10. Global Advertising	13. Company Finance		Flooring Market
11. Markets	11. Management Styles	24. Strategy and Leadership		Retail Hypermarkets
12. Lobbies	12. Mergers and Acquisitions	25. Hiring, Firing, Probation		KFC Sales Growth
13. Communication	13. Business and the Environment	26. Pay, Benefits and motivation		Unilever Thailand
				Toy Test
14. Logistics	14. Finance and Credit	27. Cross-cultural Communication		Electricity Demand
15. Innovation	15. Economic Issues	28. Business and Environment		Sony Thailand
*Units 15/15	*Units 15/15	*Units 15/70	*All relevant	*All relevant

*Units used out of total units unless units are not applicable.

The material in the three globally mass produced coursebooks, *Intelligent Business*, *New Business Matters*, and *Essential Business Vocabulary Builder*, have articles that the authors claim have very appropriate content. For instance, *Intelligent Business* takes article from *The Economist*, a famous and respected business magazine. *New Business Matters* claims each of their articles is "specially written" and "rich in the language of company life". *Essential Vocabulary Builder* claims that its articles teach "up-to-date language in modern business

contexts." The table below gives an example of the articles that can be found in each mass produced coursebook.

Table 3.2: Sample Mass Produced Coursebook Texts Excerpts

Essential Vocabulary Builder	"Senior managers, with a strong involvement from the Board of Directors, have to develop a long-term strategy for the company. The first step is to analyze the current situation. There are a variety of tools and techniques available to do this."
New Business Matters	"But an entrepreneur is not what you are, it's what you become, and real entrepreneurs exist only in retrospect. At first, nobody takes them seriously. They're crackpots, dreamers, unemployables. And by the time they've finally earned the respect of the business community, they've already made it."
Intelligent Business	"When the public mood changes, the realisation can take time to sink in. Behaviour that was once acceptable can overnight come to be seen as outrageous. The board of GlaxoSmithKline, a big pharmaceutical company, has found itself at the sharp end of such a mood change."

The materials from the limited production coursebooks, *BG 2000 English III* and *BG 2001 English IV*, have articles that have been taken from various sources such as the *Bangkok Post* newspaper, *Times* magazine, and *The Nation* newspaper. However, the articles have been adapted to suite the authors' purpose to some extent. The authors note that the articles have been "extensively adapted, often with several articles from different sources being combined and re-organized to compile articles of appropriate length and simplicity for our use." (Kumar et al. 2010b: 35) Typically the articles are used for teaching reading comprehension skills and vocabulary. The table below gives an example of the articles that can be found in the limited production coursebooks.

Table 3.3: Sample Limited Production Coursebook Texts Excerpts

BG 2000 English III	If you have been dreaming about working in your own home, now might be the time for it to happen. With rising oil prices and global warming problems, many companies are seriously thinking about implementing telework program. Telework or telecommuting is a type of arrangement where employees can work flexibly outside their central offices.
BG 2001 English IV	KFC has enjoyed 16-per-cent sales growth due to its continuing popularity, the launch of new menu items and marketing campaigns, its policy of maintaining current prices for food and its delivery service. The company has opened 19 new KFC branches nationwide, which accounted for 9 per cent of the 16-per-cent sales growth, Sran said.

Data is taken directly from the coursebooks by first electronically scanning the article, converting it into a word document and then, after being checked for scan errors, saved as a text file for WordSmith. This data is turned into an additional corpus for comparison. The corpus made from the mass produced coursebooks found at Chulalongkorn University is called "MAS". Similarly, the corpus made from the limited production coursebooks used in ABAC University is called "LIM".

3.5 Method of Data Analysis

The mini corpora from the coursebooks and the large corpus from the *Bangkok Post* and *The Nation* is used in the following ways:

Table 3.4: Corpus Analysis

Function	Description	Details	Special Note
Wordlist	Creates a list of the most frequently used words in the news stories corpus and the coursebook corpus.	The list shows words in order from most frequent to least. It is used to determine word clusters.	Three lists will be made. The first represents the entire newspaper corpus, the second the Chulalongkorn coursebooks, the third the ABAC coursebooks.
Word Cluster	Creates a list of word clusters in each corpus. This study uses 3-word clusters.	The program is set to search for all four-word clusters, excluding ones with numbers because the focus is on vocabulary use.	Frequency in the coursebook corpus is at 3+ because the smaller corpus doesn't not reflect high frequency but is expected to show a concentrated amount of data due to the concentrated content of coursebooks.

Only the coursebook 3-word bundles with a frequency of three or more are researched. A 3-word bundle has been chosen because it offers the right balance between data quality and amount of data available for research. For example, 2-word bundles don't give enough insight into language attributes and 4-word bundles do not produce enough actual bundles for study. The entire newspaper coursebook is used to compare to the coursebooks so that there is a higher chance that the 3-word bundles in the much smaller coursebook corpora could be found. The coursebook bundles that do appear in the newspaper corpus are also looked at for what frequency the bundle has in the newspaper corpus. This helps to determine if the most frequent and thus most important bundles are represented. To research the unique attributes of the newspaper corpus the top 100 3-word bundle are looked at.

3.6 Pilot Study

A pilot study has been done comparing lexical bundles from six chapters from each of the coursebooks in the MAS corpus and the full newspaper corpus. The reason for this is to

determine whether research results and data are sufficient to proceed to a full scale main study. The newspapers corpus produced a list of 65,500 bundles from 851,207 tokens. The coursebooks produced 28 bundles from 9,111 tokens. Below are the twenty common bundles:

Table 3.5: Common Bundles in the Incomplete MAS and Thai BELF Corpora

Bundle	Frequency in Newspaper Corpus	Bundle	Frequency in Newspaper Corpus
PER CENT OF	558	IF THEY ARE	9
THERE IS A	59	IN THE BUSINESS	26
OF THE COMPANY	40	NEED TO TAKE	3
THE COMPANY CAN	8	OF THE WORLD	27
CENT OF THE	168	OF THE WORLD'S	34
IN THE PAST	33	POWER OF THE	5
THE NUMBER OF	228	THAT THE COMPANY	68
A HANDFUL OF	6	THE COST OF	87
AS WELL AS	436	THE FORM OF	21
FIVE PER CENT OF	5	TO MAKE MONEY	4

The pilot study results show that a high percentage of the coursebook bundles are represented in the newspaper corpus. It is important to note that although some frequencies are as low at three, they represent a proportionate value for the scope of this particular study. A high percentage is anything with a frequency of 100 and above. However, when looking at the frequency of the bundles in the newspaper corpus, the bundles are not from the top 100 most frequent bundles. This means the bundles in the coursebooks tend do not represent the more frequently used bundles and are generally slightly less important in terms of language teaching.

The bundle lists also showed some interesting attributes in at first glance. First, there is a large number of bundles with "PER CENT". This shows a focus on percentages instead of absolute numbers which can hint that exact numerical details may not be very important. Second, there is a high number of time indicator bundles such as "IN THE FIRST" and "IN

THE NEXT" and "OF THE YEAR". Time indicator bundles do not show up at all in the pilot MAS corpus. This could reflect the focus of the Thai BELF corpus existing in a continuous state because business is most reflected in the present situation or speculation to the future. These and other attributes can be looked at further in the main study to enhance the understanding of Thai BELF. Further look into the lexical bundles as they appear in the text are also used to give a better understanding of the bundle use in context. **3.7 Conclusion**

In conclusion, this research has identified common lexical bundles in *The Nation* and *Bangkok Post* by using a corpus based on the business news stories from the two online versions of the newspapers from March 2011 to July 2011. Next it can determine if the bundles in the newspapers are represented in the *Intelligent Business, New Business Matters, Essential Vocabulary Builder, BG 2000 English III* and *BG 2001 English IV* coursebooks by comparing the cluster lists. The lexical bundles used in each corpus is analyzed for differences and similarities of note as well as unique attributes. The results of these comparisons give insight into the vocabulary and common bundles used in coursebooks and the real BELF being used in Thailand today. The next Chapter will discuss the results of the study.

Chapter 4

Analysis of Data and Discussion

4.1 Introduction

This Chapter attempts to determine to what extent Thai BELF 3-word lexical bundles from a corpus of business news stories from the *Bangkok Post* and *The Nation* newspapers appear in the texts of business English coursebooks used in Bangkok. A closer look at the corpus created bundles reveal features of Thai BELF as discussed in Chapter three. The results of this analysis can be used to not only check how well the coursebooks used in Bangkok match the content in the real Thai BELF environment, but also the unique attributes of Thai BELF. First, a discussion of the pilot study findings is given. Second, a culture based attribute analysis for the entire BELF corpus lexical bundle list is given. Culture in this sense is defined as in Chapter one as the reoccurring aspects of society, or "norms", which can be observed through language used in the context of that society. Specifically, the "norms" of language that are found in the Thai BELF corpus, the repeated use of lexical bundles for various reasons, are considered the cultural aspects of Thai BELF. These aspects of culture can come in many forms such as style, use of certain words opposed to others, and the high frequency of specific words or lexical bundles. These forms reveal recognizable patterns that are looked at in depth to understand the Thai BELF corpus thoroughly.

The third analysis is on the general attributes of the Thai BELF top 100 bundles that are used to compare the mini coursebook corpora from *Intelligent Business, Essential Business and New Business Matters* coursebooks available in Thailand and the *BG 2000 English III and BG 2001 English IV* coursebooks used at Assumption University. These general attributes are determined by observing which bundles occur the most and the least and the context that these bundles appear in the texts. General attributes include not only frequency, but also salience, patterns of lexical use, culturally derived lexical use, and other standout

occurrences. Next, is an analysis of the bundles from the *Intelligent Business*, *Essential Business* and *New Business Matters* coursebooks corpus (MAS) found at Chulalongkorn University followed by the *BG 2000 English III* and *BG 2001 English IV* coursebooks corpus (LIM) used at Assumption University.

As stated in Chapters one and three, the research questions are as follows:

1. What are the most frequently used BELF lexical bundles and their attributes in the business pages of the *Bangkok Post* and *The Nation* newspapers?

2. To what extent are the most frequently used lexical bundles in the Thai BELF corpus represented within the *Intelligent Business, Essential Business* and *New Business Matters* coursebooks available in Thailand (MAS corpus) and the *BG 2000 English III* and *BG 2001 English IV* coursebooks used at Assumption University (ABAC) in Thailand (LIM corpus)?

The analysis is organized into eight main sections. Section 4.2 Findings of the Pilot Study in the Main Study, discusses the correlation between the two studies, 4.3 Analysis of Culturally Derived Attributes of the Thai BELF Corpus, discusses attributes of gender, social outlook, global business focus, the use of negative vocabulary and governmental influence, time and business specific bundles. Section 4.4 Analysis of Thai BELF Top 100 Bundles, discusses what the bundles and their context are as well as patterns of use and non-use. Section 4.5 Analysis of MAS and LIM corpora discusses the attributes of note in the two corpora as well as example bundles in context. A comparison of the two corpora is also given. Section 4.6 answers research question one and section 4.7 answers research question two. Section 4.8 gives concluding remarks on the previous sections.

The data for the lexical bundles can be found in Appendix I: Top 100, Appendix II: MAS Bundles, and Appendix III LIM Bundles. The texts used for the main BELF corpus and the

two mini coursebook corpora as well as the BELF corpus full bundle list can be found in Appendix IV in the CDROM.

4.2 Findings of the Pilot Study in the Main Study

The first finding of the pilot study showed that the incomplete MAS corpus at only 9,111 tokens, shared a high percentage of the same bundles with the Thai BELF corpus. Specifically, 20 bundles out of the 28 bundles in the incomplete MAS corpus produced are found in the Thai BELF corpus. It also showed that of those 20 common bundles only four bundles were in the top 100 most frequent bundles from the Thai BELF corpus. The main study produced similar results. Although the MAS corpus had more bundles in common with the Thai BELF corpus at 80 per cent versus 67 per cent in the LIM corpus, the corpus with the most bundles in the Thai BELF top 100 is the LIM corpus. This observation is discussed further throughout this chapter.

The second finding of the pilot study, the high use of "PER CENT OF" in the Thai BELF corpus which occurred 558 times, indicated that using percentages instead of absolute numbers was more favored. In the main study, this bundle made up ten per cent of the bundles in the LIM corpus and seven per cent of the bundles in the MAS corpus. In the MAS corpus, "PER CENT OF" appeared as the most frequent bundle at 23 occurrences. In the LIM corpus, it appeared as the second most frequent bundle at eight occurrences. These results show that using percentages is favored throughout the business English coursebooks and Thai BELF and additionally reflect the language of business.

The third finding from the pilot study, the use of time indicator bundles such as "IN THE FIRST" and "OF THE YEAR", did not appear in the incomplete MAS corpus. Based on this observation, it was speculated that Thai BELF exists in a continuous state, focused more on the present situation and the future rather than a specific day or month. After a more detailed

study of the Thai BELF corpus, bundles involving days, months, years, and quarters were more frequent than time indicator bundles without time words such as "IN THE NEXT". The use of time indicator bundles and key words using "YEAR" is most frequently used, however, "MONTH", "DAY" and "QUARTER" were used at relatively the same frequency (78, 76, and 70 respectively). The MAS corpus and LIM corpus showed results similar to the pilot study. The MAS corpus lacked time indicator bundles using days, months, years or quarters. The LIM corpus had only one bundle, "OF THE YEAR". A reason for this non-use of time indicator bundles in the coursebooks could be attributed to the use of only static information. In other words, the texts in the coursebooks generally followed a time line within the scope of the text, but in the newspaper articles, the time scope of the text is often extended to report on present conditions that could be changing and forecasting which also exist in a state of flux. This is further supported by the existence of bundles containing "WILL BE" occurring 86 times. Time indicators in the Thai BELF corpus are discussed further in this chapter.

4.3 Analysis of Culturally Derived Attributes of the Thai BELF Corpus

The news stories corpus is made of 810,662 tokens from 1,768 texts. It produced 65,500 3-word lexical bundles. From these bundles five culture-based observations can be made. First, gender revealing word bundles, although used less frequently than entities used for the same purpose such as "THE COMPANY SAID", revealed implications of gender and collectivism superiority. As discussed in Chapter two, gendered words are those that show if the speaker or subject is male or female. These include "he", "she, "his" and "hers". Collectivism can be considered a type of social outlook where members of a society value the society's members more as a group instead of as individuals. The gendered words "HE" and "HIS" occurred in a total of 131 bundles. The gendered words "SHE" and "HER" occurred in only 14 bundles. The ratio of male to female words use is a little over nine to one. This high ratio to males

over females suggests that the Thai BELF environment is male dominant. However, two of the gender type bundles included both genders, "HIM OR HER" and "HE OR SHE" which can suggest an attempt at creating a more gender neutral environment.

Following are some in-text examples of the use of male gendered words:

> The Commerce Ministry was confident the country's inflation rate for this year would be in the range of 3.2 to 3.7 per cent, **he** said.
>
> A total of 36.3% of these agreed the high cost of living was a difficult problem and not easily solved, as past governments had failed to settle it, while 19.0% believed the prices of consumer goods would continue to rise, **he** added.
>
> Energy Minister Wannarat Channukul yesterday broke **his** silence on plans for Thailand's first nuclear power plant scheduled to operate from 2020, admitting the project should be reviewed by energy policy planners.
>
> Prasert Bunsumpun has only six months left in **his** term as president and chief executive of the PTT Group, Thailand's largest income generator and biggest company by market capitalisation.
>
> Jon Eddy Abdullah has become an overnight sensation in the local telecom industry after less than two months as chief executive of Total Access Communication (DTAC), with **his** surprise court appearances aimed at fighting for fairness and transparency.

Following are some in-text examples of the use of female gendered words:

> **She** said daily deals online had become an especially attractive tool to help companies increase sales volumes.
>
> Brian separately asked three directors who also report directly to Khun Vanida about specific behaviours they thought **she** should change.
>
> The two new hypermarkets will range in size from 4,500 to 12,000 square metres each, **she** told a shareholders' meeting yesterday.
>
> A former advertising film producer who quit **her** job to pursue a green business venture, Ms Orapin founded Kokoboard, which uses rice straw, coconut waste and vetiver grass to make bio-composite wood panels.
>
> Pimjai Matsumoto, the managing director of P&E Techno Co, a leading mango exporter to Japan, said **her** company had spent 60 million baht to install a vapour heat-treatment facility at Kasetsart University to control mango quality prior to export.
>
> Prapa Puranachote, the new president, told **her** first press conference since taking the helm that to achieve its goal, MFC would expand its number of fund sales agents.

The use of gendered words appears to be non-topic specific meaning that regardless of the content of the article, company profile, finance, product launches, etc, no gender occurs more often than another. The female and male speakers are all people in management or other decision making and company representation roles. From a linguistic perspective, both gender words occur in reported speech such as "she told" and "he said". Gendered words also occur to show possession of a company or position such as "her job" and "his term". Similarly, no patterns related to gender can be seen between international versus domestic companies and articles.

Next there are a total of 95 bundles using the word "WE" but only 26 bundles using the word "I". The ratio here is a little over three to one. This ratio is not as high as in the gender aspect, but it is still a noticeable imbalance. This imbalance suggests that a collective "WE" is more accepted in the Thai BELF environment than a singular "I". The root cause of this could be attributed to the need for the company spokespersons to represent the company's interests instead of their own. Additionally, one bundle, "I THINK WE" uses both "I" and "WE" but the focus can still be put on the "WE" aspect of the bundle since the writer chose to include it. Some of the "I" and "WE" bundles are similar to each other, but the "WE" bundles at over 500 bundles still outnumber the "I" bundles" at 217 bundles. Some examples are in Table 4.1. The dominant structure of "I" and "We" bundles is pronoun + infinitive + auxiliary verb as opposed to nouns.

Table 4.1: "I" and "We" Lexical Bundles

WE HAVE TO	I HAVE TO
WE NEED TO	I NEED TO
WE HAVE BEEN	I HAVE BEEN
WE ARE CONFIDENT	I AM CONFIDENT
WE WANT TO	I WANT TO

Following are in-text examples of "WE" and "I" bundles and key words:

'In order to manage properly, **we need to** focus on four areas," says Vanus Taepaisitphongse, the chief executive of Betagro Group since 1988.

"If **we have to** sell at prices set by the ministry, the loss will be more than 2,000 baht a tonne.

"**We have been** promoting the Le'Sasha brand in Thailand's hair-equipment sector for almost 10 years.

"**We are confident** that launching marketing activities via online media will allow Philips to communicate with target consumers directly and immediately," said Ms Siriwan.

"**We** grew 18% year-on-year in the first quarter based on total premiums and expect to main continued strong growth until year-end," said chief executive Binayak Dutta.

"Even in Thailand **we** have a risk-prevention system, where in a worst-case scenario we might change our production plan to seek products from other manufacturers that produce the same model," said Suparat Sirisuwanangkura, president of the Thai Automotive Industry Association.

The first four examples of full "WE" bundles show that company representation is the main reason for the use of "WE". In the first two "WE" examples, "WE" represents a company. In the third example, "WE" represents a national identity. These two uses occurred the most often with only three examples of "WE" representing other institution such as family such as "WE HAVE A lot more family members".

"When I wake up first thing in the morning, **I have to** be the first one to know price changes in the market."

What is her preferred listening style: detail, big-picture, story, or do **I need to** build a lot of rapport first?

"What **I want to** see is the new government throw its support behind franchises, as there are new entrepreneurs wishing to set up their businesses," said Mr Suwit.

"**I am confident** that my previous experience in various industries of multinational companies will contribute business and will continue to bring strong advantages to our partners and the country as a whole."

'I normally employ active listening in my management style regardless of whether **I** agree with an issue.

"**I** plan to change the structure, system, process and people all at once.

"**I** think there are two critical qualifications missing: sound judgment and maturity.

The first three examples of "WE" bundles are focused on individual opinion or self evaluation. The fourth example combines individual opinion with the business unit of "our partners" and the country. This is the only example of the use of "I" where the speaker directly shows his connection to other entities. All three examples of the key word "I" are used to show direct connection to the individual opinion of the speaker. The speaker's opinion does not apply to that of any company or institution.

Bundles containing "YOU" and "THEY" were not as frequent as the prior discussed bundles. "YOU" occurred in 55 bundles and "THEY" occurred in 67 bundles. These two types of bundles were generally used in different ways all together. "YOU" bundles are mainly question type clusters such as "WHAT DO YOU", "DO YOU THINK", HOW DO YOU", and "DO YOU HAVE". The "THEY" bundles followed along more similar in use to "I " and "WE" as being statement type bundles such as "THEY HAVE TO", "THEY NEED TO", and "THEY WANT TO". In this respect, the focus on the individual "YOU" is more towards introspective passiveness rather than collective assertiveness.

Following are in-text examples of the key words "YOU" and "THEY":

> Have **you** ever received an intriguing message posted on your Facebook wall - presumably from one of your friends - asking you to click the link provided?

> An impressive figure, especially when **you** consider that there are 6.8 billion people on Earth.

> Have **you** ever circled a problem, only to find the same flawed solutions over and over again?

> Japanese restaurants and hotels in Thailand say **they** are taking pains to ensure their seafood is free from radioactive contamination.

> Most parties to a contract may not appreciate the importance of a properly drafted arbitration clause until **they** find themselves in a dispute.

> Businesses claim **they** have now waited seven to eight months for compensation payments and still have no idea whether or not **they** will receive compensation for the floods that hit Thailand last year.

The three examples of "YOU" give a suggestion to think internally about whatever the statement or question is in regards to as opposed to actively giving an answer or opinion. In other words, the use of "YOU" is meant for reflection and not action. The "THEY" examples show that "THEY" represents a group in the form of a company or an organized group of individuals. The context of "THEY" is also one of immediate state of being in various situations as opposed to the context of "YOU" which is not directly connected to an immediate state.

The second attribute of note observed in the Thai BELF corpus is an Asia focus in business as opposed to a global or western focus in business. Bundles including Asian countries, association to the so called ASEAN region and Asia in general totaled 242 bundles. Table 4.2 shows examples of these bundles. Europe accounted for only 10 bundles and the United States totaled 43. The Thai BELF corpus shows that inter-Asian regions have more business focus than full global or western regions. Japan has the highest number of clusters at 70 which coincide with their status as the number one investor in Thailand. Interestingly, the use

of Europe and US (United States) occurred in many bundles that showed distinct separation such as "ASIA AND EUROPE" and "US AND EUROPE". The use of "GLOBAL" only occurred in 40 bundles, slightly less than that for the United States. Again, the use of Asian type bundles is far greater than bundles containing reference to "GLOBAL". Cheng et al. (2005) used the Hong Kong Corpus of Spoken English to also come to a similar finding that the influence of the specific area the language is used in influences the language itself.

Table 4.2: Asian Bundles

ASEAN ECONOMIC COMMUNITY	MALAYSIA AND SINGAPORE
THE ASEAN MARKET	IN THAILAND MALAYSIA
COUNTRIES IN ASIA	OF THAILAND BY
OF THE ASIA	THE JAPAN THAILAND

The third attribute of Thai BELF is a lack of negative key words. A search through the bundles revealed the following information shown in Table 4.3.

Table 4.3: Negative vs. Positive Key Words

Key Word	Frequency	Key Word	Frequency
SUCCESS	9	WILL BE	86
FAIL	7	WILL NOT	8
GOOD	13	CAN	30
BAD	0	CAN NOT (CAN'T, CANNOT)	0

As seen in Table 4.3, positive vocabulary is generally used more frequently than negative vocabulary. The frequency of these key words ranges from zero to 86. The frequency range

for the Thai BELF corpus is two to 558. Keeping this in mind, low frequencies such as nine and seven are significantly low comparatively. "SUCCESS" and "FAIL" are almost the same, but "GOOD", "WILL BE" and "CAN" are used far more often than their negative counterparts or, in the case of "BAD" and "CAN NOT" do not show up at all. This suggests that strong negative comments may be unacceptable in the Thai BELF. This in-turn shows that effort is taken to spare people or organizations from losing face on account of poor performance. Supporting this is the use of "IMPROVE" which occurred seven times in clusters intended to acknowledge a negative event by focusing on recovery such as "EXPECTED TO IMPROVE" and "IMPROVE THE EFFICIENCY". This attribute coincides with Ellis' findings that in order for BELF to be used appropriately, one must be aware of the social appropriateness of using specific language and corresponding lexical knowledge in order to communicate properly. (Ellis et al., 1994)

Following are in-text examples of negative and positive key words and bundle use:

> "The risks of doing business are escalating, and we must review our second-half performance outlook. As yet, we cannot say with certainty that our full-year target growth of 12-14% **will be** achieved."

> It's pretty simple but so easy to forget that without a buy-in from the people involved in the change programme, it **will not** be successful," said Mr Dean.

> The third generation has a mere 5% chance of **success**.

> For example, if the authorities direct a service provider or a system administrator to remove pornographic or inappropriate content within a certain period, and they **fail** to do so, then they should be held culpable and subject to penalties.

> "It is clear that political parties are focusing more on people's welfare, so we have to **improve the efficiency** of our tax collection, such as upgrading our IT system," said Satit.

> The company's debt to capitalisation ratio will peak by 2011, after paying for the investment in the KKD project and making the heavy capital expenditures of Bt94 billion planned for production and development in 2011. However, the level of leverage is **expected to improve** in 2012 and 2013, as PTTEP plans to start production in many new projects during 2011-2013.

The examples of "WILL BE" and "WILL NOT" are fairly straightforward uses of the key word. However, the example using "SUCCESS" shows a very particular use of the word. The chance of success is a low five per cent which means that actually there is a greater, 95 per cent, chance of failure. Yet, the writer has chosen to use the positive word success. The example for "FAIL" uses "FAIL" in an "if" circumstance as opposed to a more definite form such as "Failing to do so will hold them culpable and subject to penalties." In the first example with a bundle, the use of "IMPROVE THE EFFICIENCY" is relatively neutral in terms of being positive. The second example however, shows a clear attempt to recover the negative situation stated in the sentence prior to the sentence containing the bundle "EXPECTED TO IMPROVE". In the prior sentence a problem is presented. Following that is a "however" statement that draws attention to positive future endeavors as an possible attempt to distract readers from the negativity of the prior statement.

Just as in Chakorn's (2006) research on Thai Business English e-mails, a focus on politeness can be seen in the Thai BELF corpus. In addition, in this corpus the additional avoidance to direct negative statements is also an important factor.

The fourth observation of note is the strong influence of the government on Thai business. The bundle list revealed 155 clusters containing reference to the government or one of the divisional ministries. Bundles in this search included clusters such as "NEW GOVERNMENT POLICY", "GOVERNMENT SHOULD INCREASE", GOVERNMENT SHOULD SET", and "THE MINISTRY WILL". These bundles and the ones like them convey a dependency that Thai business has on the government. From my experience, the Thai government has control over most Thai industries in the form of policy, price and the like. This is reflected in the Thai BELF language and confirms that doing business in Thailand requires knowledge of some government information such as policy, restrictions and rules of import and export.

Following are in-text examples of government in **business key words and bundles**:

> Mr Yuthasak estimates it will take around two months to implement the interest rate compensation after getting approval from the **government.**

> The Industry **Ministry** will discuss the issue next week with a view to helping SME suppliers, though officials expect normal operations in the industry to resume in the second half this year.

> The Commerce **Ministry** will seek 400 million baht from the Cane and Sugar Fund to compensate sugar producers for packaging costs, to encourage them to supply more one-kilogramme packages through modern-trade retailers.

> NESDB chairman Panas Simasathien agreed the **government should increase** R&D funding as well as support R&D for small farmers and entrepreneurs.

> In addition, **the ministry will** freeze the price of swine at Bt59-Bt60 a kilogram at the farm gate for three months in a bid to control the retail price of pork at Bt135-Bt140 a kilo.

The examples show that the government and government entities are needed as the control point in each context. In example one the government is the approving body. In example two the Industry Ministry will be in discussion to make a decision. In example three the Commerce Ministry is going to obtain money. In the two bundle examples, "GOVERNMENT SHOULD INCREASE" and "THE MINISTRY WILL", the role of the government is clearly connected to the farming and entrepreneurial industries. Specifically, the governments' direct relationship to funding R&D as well as price which are two very important parts of business.

The last observation from the Thai BELF bundles is in regards to time importance. Bundles which include time indicators for reporting events and forecasting make up a large number of bundles. There are a total of 655 bundles involving days, months, quarters and years which represents 1% of all bundles. Table 4.4 shows their ranking in terms of frequency.

Table 4.4 Time Indicator Rankings

Time Indicator	Frequency
YEAR, YEARS, YEARLY	431
MONTH, MONTHS, MONTHLY	78
DAY, DAYS, YESTERDAY, <WEEKDAYS>	76
QUARTER, QUARTERLY	70

As the chart shows, when conveying information in the Thai BELF environment, years take the highest rank. Months, days, and quarters are all around the same ranking. This shows that, according to this context, the year view holds the most functional importance. It is most commonly used to forecast as in "NEXT FIVE YEARS", historically as in "IN RECENT YEARS", or showing continuation "OVER THE YEARS".

Following are in-text examples of time indicator key words and bundles:

Somboon Advance Technology Plc (SAT) will review this **year's** revenue forecast of 7.3 billion baht to take into account the Japanese disasters and Middle East riots.

CIT is then imposed based on information and evidence received, with maximum penalties of 200% and surcharges of 1.5% per **month** on the amount of corporate tax owed.

He said AIS and its peers were spending more on network security these **days** because operators have essentially become internet service providers rather than just voice services.

For 10 of the 11 SET-listed banks (excluding LH Bank) reporting as of **yesterday**, combined second-**quarter** net profit was 35.66 billion baht, an increase of 32.68% from the same period a **year** ago.

"We want to be one of the leading brands in particular countries with the highest market share in the **next five years**," said Saran.

Mr. Terapun, who has headed the factory for nearly 20 years and is perhaps the longest-tenured executive of any state enterprise, says he is proud of the organisation and the progress it has made **over the years**, both in terms of profitability and production.

First, "YEAR" is used to sum up a large revenue creating period. Next "MONTH" is used for a narrow time frame for a more specific subject. Following is "DAY", which is used as a vague representation of a non-discrete time frame. In the last example, "DAYS" in the form of "YESTERDAY" as a definite time indicator in conjunction with "QUARTER" to set the time frame to compare a period existing a "YEAR" ago. The two bundle examples show the same type of use as the keyword use of "YEAR".

4.3.1 Business Specific Bundles

Business specific bundles were apparent in the Thai BELF corpus. In line with the definition of Business English in Chapter one, these bundles are considered business specific because they are only used in the business environment. They pertain to business functions, finance, products, and various industries in business as well as management. Table 4.5 shows some of these bundles.

Table 4.5 Business Specific Bundles

TO RAISE DEPOSITS	THEIR SALARY STRUCTURES	FOR CORPORATE ADMINISTRATION
TO QUARTERLY PAYMENTS	THEIR PRODUCTION BASES	SALES VOLUME GROWTH
TO MITIGATE RISKS	THEIR GLOBAL PEERS	SUPPLY DEMAND DYNAMICS
TO LOGISTICS COSTS	THEIR INTELLECTUAL PROPERTY	SUSTAINED DEMAND FROM
TO DIVERSIFY BUSINESS	THE WTO CUSTOMS	TELECOM CONCESSION AMENDMENTS
TO CORPORATE GOVERNANCE	THE VOLATILE CLIMATE	THE CAPITAL EXPENDITURE
TO CLOUD SERVICES	THE TRACEABILITY SYSTEM	THE PRODUCTION VOLUME
EMERGING MARKETS LIKE	THE STANDARD DEDUCTION	THE SITUATIONAL LEADERSHIP

61

Business lexical bundles are not necessarily unique to only BELF in Thailand, but are common in general business registers. The significance of these bundles is predominantly to show unity and commonality with other business registers.

Following are in-text examples of business specific bundles:

> The loan-to-deposit ratio of SCB, the country's fourth-largest-bank, dropped to 92-93% reflecting strong loan expansion, and the bank needs **to raise deposits** to deal with higher loan demand.

> Regarding the fund-raising plan in Thailand, Fernandes said Thai AirAsia needed to own 25 aircraft and also generate cash flow **to mitigate risks** from uncertainty and fluctuating oil prices.

> Siam Design Group, a leading interior design and construction firm, is branching out into property development **to diversify business** risks.

> Supan Mongkolsuthee, chairman and CEO of Synnex (Thailand), said his company was also moving **to cloud services to capitalise on** growing demand and improve its margins.

> While the unrest in the Middle East and North Africa might trigger increases in the price of crude oil, **the volatile climate** might affect agricultural production - both of which could fuel inflation in the second half.

> CAT will soon submit **the capital expenditure** budget for the consideration of the National Economic and Social Development Board (NESDB).

In summary, the Thai BELF corpus shows significant attributes of male dominance, collectivism, a focus on Asian business, the lack of use of negative bundles and the influence of government control. There are also business specific bundles in the Thai BELF corpus. These findings correspond to Kankaanranta's (2010) description of BELF as being a mix between cultural context and corporate/business culture along with national and social culture. (Kankaanranta et al) Analysis of the Thai top 100 bundles gives further insight into the Thai BELF corpus.

4.4 Analysis of Thai BELF Top 100 Bundles

A list of the top 100 most frequent bundles was made from the 65,500 3-word lexical bundles excluding bundles with newspaper section headings and published date information because that is exclusive to newspapers only. Appendix I shows the full 100 bundles. The list includes bundles regarding time, business entities such as the stock exchange, formal titles of people and businesses, and geographic regions. Bundles used for reference such as "IN TERMS OF" and "IN LINE WITH" are also found in the list. The top 100 bundles have a frequency rating from 558 to 81 which represents the most frequently occurring bundles in the corpus. This relates to the research of Tribble (2002), Hyland (2008a), and Cortes (2006) who believe there is strong research value in terms of context and register definition that can be placed on these reoccurring bundles. Also, following Sinclair's (1991) four essential characteristics of corpus based analysis, the actual lexical bundles are analyzed in text, it is corpus based, and takes advantage of quantitative analysis such as frequency and qualitative techniques such as cultural implications. There are several common features of the Top 100 bundles.

First, types of bundles in the top 100 bundles that have the most occurrences are those that denote time orientation at 31 occurrences. There are ten occurrences of bundles with some form of "PER CENT". The use of formal titles such as "MANAGING DIRECTOR OF" appeared eight times. Another common type of bundle is those with Thai related vocabulary such as "EXCHANGE OF THAILAND" and "BILLION BAHT IN".

Following are in-text examples of top 100 frequent bundles:

> He said the labour cost accounted for only 10 **per cent of** total production costs and if it were increased it would not have much affect on business.

> Kesara Manchusree, **managing director of** the Thailand Futures Exchange, said commodities are another alternative asset class for investors.

Pheu Thai will also push for the merger of the Agricultural Futures **Exchange of Thailand**, the country's main agricultural derivatives market, with the much larger Stock **Exchange of Thailand** to improve liquidity in the futures market.

However, the Commerce Ministry expects it will fetch 5 **billion baht in** export income this year, 15% growth over last year.

There are also some infrequently used bundles types. The use of pronouns only occurs once in the top 100 list with the use of "HE SAID THE". However, using "COMPANY" was used frequently in bundles such as "SAID THE COMPANY" and "THE COMPANY WILL". This type of bundle occurred seven times.

Here is an example of this type of bundle below:

Sarinporn Jivanun, managing director of Envirosell Thailand, said in its first year **the company will** focus on introducing itself to corporations with the hope of generating at least 50 million baht revenue in the second year.

These features show that technical words in the business context such as corporate information, management, marketing and finance are salient. This confirms Nelson's research that there are key words that make Business English its own register. (Nelson 2005)

The top 100 clusters reflect the frequent use of time indicators and percentage use that shows a focus on planning and finance. The use of formal titles is also a key feature of the top 100 bundles because of its business context. Thai specific bundles reflect the context of BELF in Thailand. Also, the use of the company as the representative of the company instead of an individual person shows a strong focus on company unity.

4.5 Analysis of MAS and LIM Corpora

The MAS corpus is made of 23,587 tokens from 44 texts from the mass produced coursebooks, *Intelligent Business*, *Essential Business* and *New Business Matters*. It produced a total of 651 lexical bundles out of which 135 had a frequency of three or higher. Of the 135 clusters, 108 clusters appeared in the Thai BELF corpus. This result means that 80 per cent of

the bundles from MAS can be found in the Thai BELF corpus. However, of the 135 bundles, only 12 or 8.9 per cent can be found in the top 100 most frequent bundles in the Thai BELF corpus. The top five most frequent bundles in the MAS corpus as well as the bundle's frequency in the BELF corpus can be seen in Table 4.8.

The LIM corpus is made of 11,192 tokens from 23 texts from the limited production coursebooks used at Assumption University, *BG 2000 English III* and *BG 2001 English IV*. It produced a total of 700 bundles of which 79 had a frequency of three or higher. Of the 79 bundles, 53 bundles appeared in the Thai BELF corpus. So, 67.1 per cent of the bundles from LIM can be found in the Thai BELF corpus. Of the 79 bundled, 19 or 24 per cent can be found in the top 100 bundles in the Thai BELF corpus. A comparison is shown in Table 4.6.

Table 4.6: MAS and LIM Corpus Comparison

Statistic	MAS	LIM
Tokens	23,587	11, 192
Texts	44	23
Total Bundles	651	700
Bundles with Freq. or 3+	135	79
Bundles Shared with BELF Corpus	108 (80%)	53 (67.1%)
Per cent in Top 100 BELF Bundles	8.9%	24%

This MAS bundle list in its entirety did not produce any bundles related to Thailand or Asia. The somewhat low number of bundles found in this corpus compared to the BELF corpus is similar to the finding from Biber et al (2004) that showed a lack of lexical bundles. He attributed this somewhat to "the fact that textbook authors are free from real-time production constraints of face-to-face teaching and therefore make more diverse language choices". (Biber et al 2004: 383)

Following are examples of some of the lexical bundles from the MAS corpus and their degree of commonality to the BELF corpus and the LIM corpus.

Edelman, one of the world's largest PR firms, not only has its own 24-hour newsroom, it also offers clients something it calls i-Wire - a monitoring service that allows you to trace, download and evaluate anything written about your company **on the internet** by anyone, anywhere, anytime.

The bundle "ON THE INTERNET" occurred only four times in the MAS corpus and only six times in the BELF corpus. It didn't occur at all in the LIM corpus. This bundle's low frequency throughout all three corpora is unexpected because the internet is assumed to be a major part of business. Furthermore, the texts in the coursebooks are dated no earlier than 2007 making them current enough to reflect the popularity of the internet. However, the low frequency of use suggests that business reporting is still focused on concepts of business such as companies and customers instead of different modes of technology.

> It was widely predicted to wipe **out** seventy-five per cent **of the world's** computers in the very first second of the 21st century.

> And with profit margins of up to a phenomenal fifty per cent, Philip Morris can still gross around $100 billion a year, making the makers of Marlboro cigarettes the most profitable company **in the world**.

The bundle "OF THE WORLD'S" appeared six times in the MAS corpus and 34 times in the BELF corpus. It did not appear in the LIM corpus. The second bundle "IN THE WORLD occurred five times in the MAS corpus and 76 times in the BELF corpus. It didn't appear in the LIM corpus either. The infrequent use of this type of bundle could be explained by a focus on the Asian region content in the BELF and LIM corpora.

> For a while, it seemed like every post-adolescent with a laptop and a business plan written **on the back** of a rock concert ticket could get access to unlimited venture capital.

The bundle "ON THE BACK" of has a frequency of four in the MAS corpus and a frequency of 24 in the BELF corpus. It doesn't show up in the LIM corpus. This bundle is used

specifically to describe a physical location. It is not common in many business articles as it can only be used in a limited number of situations such as when describing a store location, product, or advertisement design. It is also not a strongly business oriented bundle as described earlier in this chapter, and its use is widespread across many registers.

> In the USA, where unemployment benefit is cut after six months and staying out of work is not an option, they are creating jobs at **the cost of** decreased incomes.

The bundle "THE COST OF" occurred four times in the MAS corpus and 87 times in the BELF corpus. It did not occur in the LIM corpus. Typically it is used to denote the price of an item, or in this case, the amount given up.

> Like a democracy, **it can be** a noisy place where citizens sometimes think the people in charge have no idea what they're doing.

The bundle "IT CAN BE" occurred five times in the MAS corpus and only 10 times in the BELF corpus. This bundle is common in general English which is discussed in Chapter two, but is not generally used in business English. One reason for this may be its indefinite nature. This meaning, business articles and newspaper articles tend to be written to show more fact than assumptions because the typical readers are using the article for information purposes.

> It's the way **of the future**,' says Bradley Bonifacius, internet manager at Dean Stallings Ford, Inc.

The bundle "IN THE FUTURE" as well as other bundles involving "FUTURE" occurred infrequently at a frequency of two. This shows that the use of "FUTURE" is not favored. One possible reason could be its indecisiveness with more use of precise time indicators such as "OF THIS YEAR" being used. Also, the context of this bundle is that of prediction. The

"FUTURE" in this example is thought of as a far off point in time as opposed to a continuous state. However the bundle "WILL BE" is used 86 times and can replace "future".

'I wanted to give **the power of** the market back to individuals,' says Omidyar.

The bundle "THE POWER OF" occurred three times in the MAS corpus and 13 times in the BELF corpus. It did not occur in the LIM corpus. Typically the use of "POWER" in the BELF corpus is used to denote the strength of a formal entity such as a company or government agency. This use of "POWER" is representational of that typical use because the entity it is referring to is the market.

In the end, they nailed the deadline and here's how they did it.

The bundle "IN THE END" has a frequency of three in the MAS corpus and 24 in the BELF corpus. It showed up in the LIM corpus in other forms such as "BY THE END" and "THE END OF". Statements of "THE END" in the BELF corpus tended to be in the context of specific time frames such as "by the end of next quarter". This use of "IN THE END" is slightly more recount in its meanings than a specific time frame.

You have to be willing to say, 'I am in the wrong place in this process. **You have to** take yourself out of the system.

The bundle "YOU HAVE TO" occurred three times in the MAS corpus and a low 26 times in the BELF corpus. It didn't occur in the LIM corpus. The use of personal pronoun "you" is mostly used in articles that are trying to persuade the reader to think of a certain thing or think a certain way. An example from the BELF corpus, is a news article that gives advice to managers on how to deal with their employees.

In sum, the MAS corpus correlated to the Thai BELF corpus at a high 80 per cent but of those lexical bundles that correlated, only 8.9 per cent of them appeared in the top 100 bundles from the Thai BELF corpus. There is also no use of bundles having to do with Asia or Thailand. These bundles use vocabulary that refers to a location in Asia and/or Asian organization. This suggests that although the coursebooks used in the MAS corpus reflect a wide variety of bundles, they do not represent bundles frequently used in Thai BELF.

The top five most frequent bundles in the LIM corpus as well as the bundle's frequency in the BELF corpus can be seen in Table 4.8. This bundle list in its entirety showed some clusters related to Thailand and Asia. Table 4.7 shows these bundles. In accordance with Cortes's (2002) and Hyland's (2008a) research, the use of these Asia specific bundles could show that register-distinct bundles do exist. It should also be kept in mind that although some of the texts are from the same sources as the Thai BELF corpus texts, the texts used in the coursbooks at Assumption University are altered for student use. This alteration for a particular audience, according to Hyland (2008b), can influence the bundles that are used in the text.

Table 4.7: LIM Corpus Thailand and Asia Specific Bundles

UNILEVER THAI HOLDINGS	PROBLEM IN CHINA
CARREFOUR SUAN LUANG	SIAM COMMERCIAL BANK
ARE IN BANGKOK	SUAN LUANG AREA
BAHT PER MONTH	TELEWORK IN THAILAND
BANGKOK TO HELP	THAI GERMAN DECOR
IN BANGKOK TO	THAI HOLDINGS RECENTLY
IN CHAING MAI	THAILAND'S FIRST EXCLUSIVE
IN THE SUAN	THE SIAM COMMERCIAL
LAUNCHED THAILAND'S FIRST	THE SUAN LUANG
MANUFACTURING TO CHINA	THE THAI MARKET
MARKET IN THAILAND	WITH THE SIAM
OF UNILEVER THAI	

The sentence below is an in-text example of an Asia specific bundle in the LIM coursebooks that is also one of the news stories corpus bundles.

> The total market for laminated flooring **in Thailand is** expected to increase by 50 per cent by the end of this year thanks to the local property boom, but the market has not expanded as quickly as it did last year.

Table 4.8: Top Five Bundles

MAS Bundles	Freq. MAS	Freq. BELF	LIM Bundles	Freq. LIM	Freq. BELF
PER CENT OF	23	558	AS WELL AS	10	436
OF THE COMPANY	10	40	PER CENT OF	8	558
ONE OF THE	10	206	CREDIT CARD PROGRAM	6	4
FIVE PER CENT	8	5	NUCLEAR POWER PLANT	6	19
CENT OF THE	7	168	THE END OF	6	330
Average BELF Freq.		195.4	**Average BELF Freq.**		269.4

Following are examples of some of the lexical bundles from the LIM corpus and their degree of commonality to the BELF corpus and the MAS corpus.

> He must work effectively on teams, be accepted by others as the person with the best sense of the challenge confronting the group, and **be able to** break problems into manageable, status-free tasks that others are willing to focus on.

The use of the bundle "BE ABLE TO" appeared five times in the LIM corpus and a high 144 times in the BELF corpus. It did not show up in the MAS frequency of three or more bundle list.

The toy industry, along with business, has moved so much manufacturing to China, **in order to** cut costs, that it remains exposed to problems despite laws and efforts in the ground to contain them.

The bundle "IN ORDER TO" appeared three times in the LIM corpus and a high 167 times in the BELF corpus. This bundle showed up in the MAS corpus with a frequency of four, a significantly lower frequency than the LIM corpus as the MAS corpus is bigger.

Women particularly appreciate the Whif's convenient, handbag-friendly size, **as well as** the novelty of offering Whifs to guests after parties.

The bundle "AS WELL AS" appeared ten times, making the highest frequency item of the LIM corpus, and a very high 436 times in the BELF corpus. This bundle showed up in the MAS corpus with a frequency of only six. This significant difference suggests that the bundles in BELF and MAS are more different than BELF and LIM.

Each **one of the** 2,000 or so he makes each year passes through his hand at his shop in Ogunquit, Maine, and no child he said, has ever been hurt by one of his small boats, cars, helicopters or rattles.

The bundle "ONE OF THE" appeared four times in the LIM corpus and a very high 206 times in the BELF corpus. This bundle showed up in the MAS corpus with a frequency of 10. In this case, **the MAS corpus closer represents this than the BELF corpus.**

The information gathered in the Harris survey shows the type of activities that people **in the US** do in their leisure time.

The bundle "IN THE US" appeared five times in the LIM corpus and a lower that previous bundles 59 times in the BELF corpus. This bundle showed in the MAS corpus four, however

the bundle "IN THE USA" appeared seven times. "IN THE USA" doesn't appear in the BELF corpus at all. This could mean that the MAS coursebooks are using lexical bundles not frequently found in this particular business genre.

> About 70 **per cent of** sales come from new housing projects, while 30 per cent are from renovations or flooring replacement.

The bundle "PER CENT OF" appeared eight times in the LIM corpus and an extremely high 558 times in the BELF corpus making it the most frequent bundle in the BELF corpus. This bundle showed up in the MAS corpus with a frequency of 23. In this case the MAS corpus adequately reflects this bundle and is also the most frequent bundle in the MAS corpus.

> KFC has enjoyed 16-**per-cent sales** growth due to its continuing popularity, the launch of new menu items and marketing campaigns, its policy of maintaining current prices for food and its delivery service.

The bundle "PER CENT SALES" appeared three times in the LIM corpus and an extremely low three times in the BELF corpus. This bundle did not show in the MAS corpus. These frequencies are unexpected results because the per cent of sales is a common method of describing sales figures. However, further research into the in-text use of bundles showed that the bundle "PER CENT OF", the most common bundle, is actually used frequently as "per cent of sales". This revealed that "PER CENT SALES" is actually a matter of style and not necessarily an infrequently used concept.

> **In line with** the government's policy of promoting fair competition between modern and traditional retail stores, the government recently approved the setting up of an independent trade body, as proposed by the Commerce Ministry.

The bundle "IN LINE WITH" appeared four times in the LIM corpus and a high 183 times in the BELF corpus. This bundle did not show in the MAS corpus. This type of bundle is used

frequently as a connecter in information report type genre. The frequency results suggest that

the MAS coursebooks don't offer texts that are uniformly informational. They tend to use a

more story-type genre.

> Yum Restaurants International (Thailand) will spend Bt 400 million **in the second**
> half of this year to open 30 new branches and launch new menus for its KFC and
> Pizza Hut chains.

The bundle "IN THE SECOND" appeared three times in the LIM corpus and a high 202

times in the BELF corpus. This bundle did not show up in the MAS corpus. Based on the in-

text use of this bundle, it is most commonly used as a time indicator for forecasting sales and

revenue. After considering this bundle use and reviewing texts in the coursebook, the MAS

coursebooks do not offer texts that represent sales and revenue oriented foci.

> **In terms of** pricing, instead of increasing the prices of its products, the company had
> to adjust its operations and ask for cooperation from its business partners to help
> improve their systems and maximize cost efficiency.

The bundle "IN TERMS OF" appeared four times in the LIM corpus and a high 184 times in

the BELF corpus. This bundle did not show up in the MAS corpus, however the bundle

"TERMS OF THE" with a frequency of three, was shown. This bundle shows in the BELF

corpus with a frequency of only 21. Again, this could point to a difference in stylistic choice

and not necessarily content.

> The "Unilever Credit Plus" program, run in association with the Siam Commercial
> bank, is aimed at strengthening traditional, **small, and medium sized** retailers
> **throughout the country** by improving their cash-flow capability.

This sentence contains three bundles in the BELF corpus. The first, "SMALL AND

MEDIUM" and "AND MEDIUM SIZED" has a frequency of four in the LIM corpus and 101

and 79 respectively in the BELF corpus. Both of these bundles do not appear in the MAS corpus. Further investigation shows that these bundles are very context specific to various types of businesses. The MAS coursebooks do not have texts that reflect this content. The third bundle "THROUGHOUT THE COUNTRY" has a frequency of four in the LIM corpus and a low frequency of nine in the BELF corpus. It doesn't show up in the MAS corpus at all. As this is a very general and non-specific term its use may not be appropriate when details are important.

> Unilever expects more than 10, 000 traditional retailers to join its **credit card program** over the next twelve months, according to the Theo Koenders, chairman of Unilever Thai Holdings.

The bundle "CREDIT CARD PROGRAM" is one of the only fully content oriented bundles. It occurred six times in the LIM corpus and four times in the BELF corpus. It didn't show in the MAS corpus. This bundle is a good example of content that could possibly be too narrow in scope to appropriately represent important lexical bundles for learners.

> **He added that** all Unilever's distribution partners now have hand held Palm Pilot computers for their sales staff, for recording orders and information on their visits to retail stores.

The bundle "HE ADDED THAT" occurred four times in the LIM corpus, a moderate 78 times in the BELF corpus and not at all in the MAS corpus. This type of cluster is common to information reporting genres but is not used as frequently as the two word bundles "HE SAID", "SHE SAID" and "*NAME* SAID".

Third, the MAS bundles did not show any Thailand or Asia specific clusters but the LIM bundles showed a total of 23 clusters. This shows that although the MAS coursebooks claim to have texts that reflect the global business environment, they do not have a concentration of language regarding entire Asia.

74

An example of text from the LIM corpus containing an Asia specific bundle is:

> As a result of the low priority given to agriculture, irrigation **in Thailand is** totally inadequate with the implementation of modern irrigation technology far behind that of the other major producers.

Lastly, the MAS and LIM bundle lists shared only ten common bundles shown in Table 4.9. This is only a 4.7 per cent overlap. The low overlap gives some evidence that the texts in the MAS and LIM coursebooks are significantly different despite having similar purpose. This is an important point to note because, as Bremner (2008) suggests, it is important for texts to function appropriately in a register. This means that text not representing the same lexical bundles as the register do not serve their intended purpose in the coursebook. The lack of bundles in the MAS corpus also point to a lack of content that is relevant to the Thai business environment in the sense that local regional content does not exist. This, to some extent, means much of the coursebook content is somewhat irrelevant to the learner living in Bangkok.

Table 4.9: MAS and LIM Common Bundles

AS WELL AS	CENT OF THE
PER CENT OF	IN ORDER TO
IN THE US	IN THE PAST
ONE OF THE	OF THE WORLD
A LOT OF	THERE IS NO

The analysis of the LIM corpus shows that although there is a lower parentage of common bundles to the Thai BELF corpus than the MAS corpus, more of the bundles are in the top 100 Thai BELF bundles. This is partly due to the presence of Thai and Asia specific bundles

as well as common sources of text. However, as the texts in the LIM corpus have been adapted they could have had a higher correlation if left in their original state. Following is a discussion of the findings of this research as it pertains to the research questions.

4.6 What are the most frequently used BELF lexical bundles and their attributes in the business pages of the *Bangkok Post* and *The Nation* newspapers?

As explained earlier on page 50, the completed news stories corpus is made of 810,662 tokens from 1,768 texts from the *Bangkok Post* and *The Nation* newspapers. It produced 65,500 3-word lexical bundles which can be found in Appendix IV in the CDROM. The following table shows the top 20 lexical bundles of the Thai BELF corpus.

Table 4.10: Top 20 Bundles in the Thai BELF Corpus

PER CENT OF	HE SAID THE
IN THE FIRST	ONE OF THE
IS EXPECTED TO	IN THE SECOND
AS WELL AS	THE STOCK EXCHANGE
THE FIRST QUARTER	IN TERMS OF
THE END OF	IN LINE WITH
OF THIS YEAR	MANAGING DIRECTOR OF
YEAR ON YEAR	PER CENT TO
SAID THE COMPANY	STOCK EXCHANGE OF
THE NUMBER OF	EXCHANGE OF THAILAND

As discussed earlier on page 50, five cultural based observations about the Thai BELF corpus can be made: male dominated and collectivist in nature, Asian type clusters show a focus on Asia business rather than global or western business, strong negative key words and bundles are not used frequently, the government has a strong role in business, and "YEARS" are the most used time indicator in Thai BELF. Business specific bundles also occur throughout the corpus.

Male dominance can be seen through a nine to one ratio of the use of male words to female. There are two gender bundles that represent both male and female genders at once that show gender neutral inclinations. The use of "WE" opposed to "I" shows a collectivist nature that can possibly be based in the need to represent a company or business entity instead of an individual. Similarly, the use of "YOU" related to internal reflection and "THEY" is connected to a business entity. The Asian clusters that appeared in the Thai BELF corpus and the LIM corpus are a result of similar content sources. These sources come from the local region and were not in the MAS corpus as it was produced internationally by western companies. Negative key words and bundles were not used frequently to avoid companies or people from losing face. Alternatively, positive bundles were used to recover from a negative condition. The second to last attribute, government influence, showed that much of business in this region is dependent on actions or policies of the government. Lastly, the use of the time indicator "YEAR" over other time indicators such as "DAYS" "MONTHS" and QUARTERS" showed a focus on forecasting and broad views regarding financial and product information. The existence of business related bundles throughout the corpus act as register markers that show this corpus as being business specific.

4.7 To what extent are the most frequently used words in the corpus represented within the *Intelligent Business, Essential Business and New Business Matters* coursebooks available in Thailand and the *BG 2000 English III and BG 2001 English IV* coursebooks used at Assumption University (ABAC) in Thailand?

The data analysis of the *Intelligent Business, Essential Business and New Business Matters* coursebooks (MAS) and the *BG 2000 English III and BG 2001 English IV* coursebooks (LIM) shows a number of things about the extent to which the news stories corpus is represented. First, the MAS bundles showed more bundles that appear in the BELF corpus (80%) than the LIM bundles (67.1%).

Table: 4.11: MAS and LIM Common Bundles within the 100 Top Bundles

MAS Corpus	Freq.	Freq. BELF	LIM Corpus	Freq.	Freq. BELF
PER CENT OF	23	558	AS WELL AS	10	436
ONE OF THE	10	206	PER CENT OF	8	558
CENT OF THE	7	168	THE END OF	6	330
AS WELL AS	6	436	BE ABLE TO	5	144
THE NUMBER OF	5	228	IN LINE WITH	4	183
A LOT OF	4	88	IN TERMS OF	4	184
IN ORDER TO	4	167	ONE OF THE	4	206
IN THE PAST	4	133	SMALL AND MEDIUM	4	101
THE COST OF	4	87	A LOT OF	3	88
ACCORDING TO THE	3	140	BY THE END OF	3	162
THE COMPANY IS	3	153	CENT OF THE	3	168
THE UNITED STATES	3	133	IN ORDER TO	3	167
			IN THE PAST	3	133
			IN THE SECOND	3	202
			LINE WITH THE	3	82
			OVER THE NEXT	3	149
			OVER THE PAST	3	85
			PER CENT AND	3	121
			THE COMPANY HAS	3	176

Second, although MAS has more tokens and texts, the LIM coursebooks created more bundles and more high frequency bundles that are found in the BELF corpus. This suggests that more frequent and possibly more important bundles are concentrated in the LIM coursebooks. It can also show that a more uniform writing genre is used in the texts. The top five bundles from LIM represent more of the most frequent BELF corpus clusters at an average of 269.4 than the MAS coursebooks with an average of 195.4 even with 50 per cent less tokens. Table 4.11 shows the top frequent bundles between the LIM and MAS corpus and the Thai BELF corpus. This also shows the LIM bundles may have a closer relationship

to the BELF corpus than the MAS bundles. As discussed earlier, this can be attributed to the LIM corpus texts coming from the same sources as texts in the Thai BELF corpus. These sources would have had an even higher correlation had they not been altered by the coursebook's authors. The MAS corpus did not share any similar data resources with the Thai BELF corpus.

Some discrepancy in the bundles' correlation can be a result of stylistic choice and not necessarily missing content. As discussed in the analysis of the LIM corpus, some bundles were only different by one word such as "IN TERMS OF" and "TERMS OF THE".

4.8 Conclusion

The analysis of data has covered two main areas of this thesis. The analysis followed Sinclair's (1991) essential characteristics of corpus based analysis. By focusing on the actual lexical bundles in text and taking advantage of quantitative analysis in terms of frequency and qualitative techniques such as cultural implications characteristics of the corpus were easily discovered. First the attributes of the Thai BELF corpus created bundles was analyzed. Five areas have been identified: 1. Male dominated and collectivist in nature, 2. Focused on Asia business instead of Global business, 3. Abstaining from the use of negative clusters, 4. Strong government influence, and 5. "Year" is the most used time indicator. As stated earlier, these attributes correspond with Kankaanranta's (2010) description of BELF as being a mix between cultural context and corporate/business culture along with national and social culture. (Kankaanranta et al: 2010) This mix is what determines the words and bundles that will be part of the Thai BELF register. In addition, business specific bundles can be found throughout the corpus. These bundles represent the commonality between regions outside of Thailand such as the USA and Australia. According to Nelson (2005), these features show that technical words in the business context are salient and make Business English its own

register. (Nelson, 2005) The main features of the top 100 Thai BELF bundles reflect the frequent use of time indicators and percentage as well as formal titles and Thai key words. All of these reflect a Thai BELF context with a business core. In addition, the use of the company as the representative of itself shows a strong focus on company unity. In agreement with Tribble (2002), Hyland (2008a), and Cortes (2006) views, the research value in terms of context and register definition can be seen through these reoccurring bundles.

The analysis of the top 100 bundles was used to analyze the MAS corpus and LIM corpus. The MAS corpus shared a high 80 per cent of the Thai BELF corpus bundles, but of those lexical bundles, only 8.9 per cent of them appeared in the top 100 bundles from the Thai BELF corpus. Asia or Thailand bundles and key words did not exist which suggests that although the coursebooks used in the MAS corpus reflect a wide variety of bundles, they do not reflect the register of Thai BELF. The LIM corpus shows a lower parentage of common bundles to the Thai BELF corpus but, more of the bundles are in the top 100 Thai BELF bundles than in the MAS corpus. Since Thai and Asia specific bundles exist in this corpus as well as common sources of texts there was higher correlation. Cortes's (2002) and Hyland's (2008a) research shows that the use of these specific bundles further shows that register-distinct bundles do exist. If texts in the LIM corpus were not altered by the authors, the correlation could have been higher. This corresponds to Biber's (2004) notion that a lack of lexical bundles in coursebooks is partly to do real-time production constraints and the need for authors to adapt their texts and to Hyland's (2008b) research on the influence of adapted texts on bundles.

In general, the findings of this analysis have shown several things. First, BELF has lexical salience that can be assumed to change depending on the context of its use. Second, coursebooks used for mass production represent a large number of lexical bundles used in business English, however, those bundles do not necessarily reflect bundles that are specific

to a BELF register. Third, LIM coursebooks that use texts from BELF sources such as local newspapers show more register specific bundles. However, by altering the texts by combining multiple texts and changing vocabulary, this correlation can be reduced. Bremner (2008) research states that it is important for texts to function appropriately in a register but with altered texts, that functionality does not adhere to the register.

The findings showing the lack of bundles in the coursebooks has implications for coursebook and syllabus design in business English courses. One implication is the need for authentic texts from the region in coursebooks. Authentic texts are unaltered and thus maintain their original lexical bundles. Authentic bundles, similar to Cortes (2006) work, students become aware of the typical use of words and can later reproduce them. This also lessens the gap between classroom business English use and real world use. Students benefit from more accurate use of lexical knowledge and a smoother transition into the business world. In addition, because the texts are from the local region, business content exists in the context of the local business environment. Not only does this have informative benefits, but can also help to make a connection between the student and the material. A student who makes a connection with the material is more engaged and likely to be motivated.

These findings have shown that the lexical bundles in the Thai BELF corpus show patterns that describe the attributes of Thai BELF. These attributes are not only purely lexical or grammatical but also cultural. The MAS and LIM corpus are useful tools to compare lexical and contextual use of bundles and texts to the Thai BELF corpus. The results of the comparison gave valuable insight into the strengths and weaknesses of the coursebooks in relation to Thai BELF. This in turn can be used to rate and improve coursebooks. Further implications for coursebook and syllabi design are discussed in the next chapter as well as a summary of the findings, limitations of this study and suggestions for further research.

Chapter 5

Conclusion

5.1 Introduction

This chapter gives a summary of the findings followed by teaching and syllabi design recommendations. Next the limitations of this study and finally recommendations for further research are given.

5.2 Summary of Findings

The completed news stories corpus produced 65,500 3-word bundles. From these bundles five cultural based observations can be made. First, Thai BELF is male dominate due to the much higher frequency of bundles containing "HE" and "HIS". Also, a higher frequency of bundles containing "WE" showed a pattern of collectivism. Finally, "I" is generally used more for personal introspective passiveness rather than "THEY" which are used for collective assertiveness. Second, the use of Asian type bundles such as "ASEAN ECONOMIC COMMUNITY" is far greater than bundles containing reference to "GLOBAL" such as "THE GLOBAL ECONOMY". The Thai BELF corpus shows that inter-Asia regions are more the business focus than a full global or western business focus.

Third, Thai BELF bundles showed a lack of strong negative bundles such as those using "BAD" or "FAIL". This shows that effort is taken to be more sensitive to the image of others and to spare people and organizations from losing face on account of poor performance. The fourth observation of note is the strong influence of the government on Thai business. Bundles such as "GOVERNMENT SHOULD INCREASE", GOVERNMENT SHOULD SET" and "THE MINISTRY WILL" shows that Thai business is significantly dependant on the governments will. Lastly, the year view holds the most functional importance as it is the

most commonly used for forecasting. Examples of its uses are, "NEXT FIVE YEARS", "IN RECENT YEARS", and "OVER THE YEARS". Additionally, bundles used only in the business genre are apparent throughout the Thai BELF corpus.

Another analysis of the top 100 most frequent bundles was made from the 65,500 3-word bundles list excluding bundles with newspaper section headings and published date information. This list revealed that time orientation, "PER CENT", formal titles such as "MANAGING DIRECTOR OF" and Thai related vocabulary such as "EXCHANGE OF THAILAND" and "BILLION BAHT IN" are the most frequent types of bundles. There is also a lack of pronoun use with "THE COMPANY" being used more often. These factors helped analyze the coursebook bundles.

The Chulalongkorn coursebooks result showed that 80 per cent of the bundles from the Chulalongkorn coursebooks can be found in the Thai BELF corpus. However, of the 135 bundles, only 12 or 8.9 per cent can be found in the top 100 most frequent bundles in the Thai BELF corpus. The ABAC coursebooks showed that of the 79 bundles, 53 bundles appeared in the Thai BELF corpus showing that 67.1 per cent of the bundles can be found in the Thai BELF corpus. 24 per cent of those bundles can be found in the top 100 most frequent bundles in the Thai BELF corpus. The ABAC coursebook corpus was the only coursebook corpus that showed Asia type bundles.

As a result of these findings, it can be assumed that the LIM corpus created from the *BG 2000 English III* and *BG 2001 English IV* coursebooks used at Assumption University better correlated in terms of content and high frequency bundles from the Thai BELF corpus. This can be attributed to similar text sources. The MAS corpus coursebooks, *Intelligent Business*, *Essential Business* and *New Business Matters* courseboooks available in Bangkok had more

common bundles to the Thai BELF corpus, but they were not generally from the highest frequency bundles and did not represent Thai or Asian content.

5.3 Recommendations for Syllabus Design and Teaching

As discussed in the previous chapter, the lack of bundles in the coursebooks has implications for coursebook and syllabi design in business English courses. There is a need for authentic texts from the region to be used in coursebooks and classroom teaching. Cortes (2006) work, showed that students became aware of the typical use of words and can later reproduce them when exposed to bundles. This lessens the gap between classroom business English use and real world use. Students benefit from more accurate lexical knowledge and a smoother transition into the business world. The texts are from the local region and hold business content that is local and current and thus creates a better connection between it and the learners.

Another way to use the cluster data for teaching is to create supplemental activities and worksheets to give learners additional resources while using the current coursebooks. As Wei and Lei (2011) noted in their study of Chinese EFL learners, the common bundles in academic writing were not easily noticed by learners and should be taught specifically to help improve learners' use of them. The Chinese learners showed only a limited number of types of bundles compared to the professional writers in their field. When learners are made aware of bundles common to the genre they are focused in they can improve their writing and speaking more rapidly.

The findings of this study also point out the importance for evaluating coursebooks based on corpus data. Often times lexical content designed by native speakers or other professional authors is simply assumed as being appropriate. However, as this study shows, the lexical use doesn't necessarily correlate to the actual lexical environment in the real-world. As Hsu

(2009) and Partington (1998, cited in Camiciottoli 2010) discovered, corpus data can greatly help to evaluate materials especially when considering that a student's vocabulary size has reached a certain level and needs to be improved.

5.4 Limitations of the Study

1. As this research is based on a newspaper articles from a specific period in time, the data itself can become outdated over time. In order for the most current data to be available for cluster research, the corpus itself needs to be continuously updated with new articles.

2. In line with the size and scope this study the smaller corpus size for the MAS and LIM corpora are adequate. However, comparing similar sized coursebook corpora from a higher number of different books could yield different results such as different ELF varieties.

3. The supplemental material that teachers use while using the coursebooks in this study is not known. Therefore, the actual lexical bundles that learners are exposed to are also unknown and cannot be considered in this research.

5.5 Recommendations for Further Study

Since this thesis is based on corpus research, the corpus itself can be used for a wide range of study. A more sensitive look into four or five word bundles could reveal more attributes of Thai BELF. Word lists with lemmatized words can be used to study frequently used vocabulary as a means to reveal patterns of use. These methods of research can be used to better understand Thai BELF and in turn give solid data to be used for further syllabus design and other teaching materials such as worksheets and study sheets. An extended look at the corpus data over time can also show how BELF changes over time. Expanding the corpus and the coursebook corpora can also improve the existing study or be used for further study in the area if Thai BELF. This study, to the researchers understanding, is the first of its kind in

Thailand and acts as only a starting point for further research in the area of Thai BELF. In the future, Thai BELF can be expanded to include spoken language and compared to BELF in other East Asian countries.

References

Ammon, U. D., N. Mattheier and J. Klaus (2006) *Handbooks of Linguistics and Communication Science (HSK): Sociolinguistics*, Part 3 (2nd Edition). Berlin: Mounton de Gruyter A Division of Walter de Guyter & Co. KG Publishers. Available from www.ebrary.com

Baker, P. (2009). *Contemporary Corpus Linguistics*. London: Continuum International Publishing Available from www.ebrary.com

Bhatia, A. (2004) *Worlds of Written Discourse: A Genre-Based View*. London. Continuum.

Biber, D. and E. Finegan (1994) *Sociolinguistic Perspectives on Register* Oxford: Oxford University Press. Available from www.ebrary.com

Biber, D. and S. Conrad and R. Reppen (1998). *Corpus Linguistics: Investigating Language Structure and Use*. Cambridge: Cambridge University Press.

Biber, D. and F. Barbieri (2007). Lexical bundles in university spoken and written registers. *English for Specific Purposes* 26 263-286 Available from www.sciencedirect.com

Biber, D. and S. Conrad and V. Cortes (2004). If you look at...:lexical bundles in university teaching and textbooks. *Applied Linguistics* 25(3) 371-405 Available from www.ebrary.com

Bodin, E. (2002) Classroom concordancing options, in M. Tan (ed.), *Corpus Studies in Langauge Education* pp. 170-182. Bangkok: Assumption University Press.

Boxer, D. (2002) *Applying Sociolinguistics: Domains and Face-To-Face Interaction*. Amsterdam: John Benjamins Publishing Company Available at www.ebrary.com

Bremner, S. (2008) Intertextuality and business communication textbooks: why students need more textual support. *English for Specific Purposes* 27: 306-321 Available from www.sciencedirect.com

Butler, J. (1988) Performative acts and gender constitution: an essay in phenomenology and feminist theory. *Theatre Journal* 40: 519-531 The John Hopkins Univeristy Press. http://people.su.se/nsnce/texter/butlerPerformance.pdf (Last Accessed October 2011)

Camiciottoli, B. (2010) A corpus informed approach to teaching lecture comprehension skills for business studies, in M. Campoy, B. Belles-Fortuno and M. Lluisa Gea-Valor (eds) *Corpus-Based Approaches to English Language Teaching* pp. 95-106. London: Continuum International Publishing Available at www.ebrary.com

Campoy, M., B. Belles-Fortuno and M. Luisa Gea-Valor (2010) Introduction to corpus linguistics and ELT, in M. Campoy, B. Belles-Fortuno and M. Lluisa Gea-Valor (eds) *Corpus-Based Approaches to English Language Teaching* pp., 3-18. London: Continuum International Publishing Available at www.ebrary.com

Canagarajah, S. (2006) Negotiating the local in English and lingua franca. *Annual Review of Applied Linguistics* 26, 197–218. Cambridge: Cambridge University Press. *http://www.personal.psu.edu/asc16/pdf/ARALpaper.pdf (Last Accessed 10 August 2011)*

Chakorn, O. (2006) Persuasive and politeness strategies in cross-cultural letters of request in the Thai business context. *Journal of Asian Pacific Communication*, 16(1), 103-146 Available at www.ingentaconnect.com

Charles, M (2009). *Academic Writing: At the Interface of Corpus and Discourse.* London: Continuum International Publishing Available at www.ebrary.com

Cheng, W. C. (2005) The Hong Kong corpus of spoken English: the creation of a prosodically transcribed intercultural corpus: http://icame.uib.no/ij29/ij29-page47-68.pdf. (Last Accessed 18 February 2011)

Chujo, K. and M. Genung (2004) Comparing the three specialized vocabualries used in "Business English" TOEIC, and the British National Corpus of Spoken Business Communication: http://www5d.biglobe.ne.jp/~chujo/data/Practical_English_Studies_2004.pdf (Last Accessed 18 February 2011)

Conrad, S. (2002) Corpus linguistic approaches for discourse analysis. *Annual Review of Applied Linguistics.* 22: 75-95 Cambridge: Cambridge University Press

Cortes, V. (2006). Teaching lexical bundles in the disciplines: an example from a writing intensive history class. *Linguistics and Education* 17 391 - 406 Available from www.ebrary.com

Cortes, V. (2002). Lexical bundles in freshman composition, in R. Reppen (ed) *Using Corpora to Explore Language Variation* pp. 131-145 Philadelphia: John Benjamin Publishing Company

Day, R. and C. Omura (1991) Incedental EFL vocabulary learning and reading: http://nflrc.hawaii.edu/rfl/PastIssues/rfl72day.pdf. (Last Accessed 10 March 2011)

Devitt, A. (2004) *Writing Genres.* Carbondale: Southern Illinois University Press.

Ellis, M. and C. Johnson (1994) *Teaching Business English.* Oxford: Oxford University Press.

Emmerson, P. (2011) *Essential Business Vocabulary Builder.* Oxford: Macmillan.

Fuentes, A. A. and P.E. Rokowski (2002) Use of corpus in the business english classroom: from lesser to more specific: http://eric.ed.gov/PDFS/ED477551.pdf. (Last Accessed 18 February 2011)

Fuentes, A. A. (2002) Exploitation and assessment of a Business English corpus through language learning tasks: http://icame.uib.no/ij26/curadofuen.pdf (Last Accessed 18 February 2011)

Gotti, M. (2006) Creating a corpus for the analysis of identity traits in English specialized discourse. *The European English Messenger* 15 (2), 44– 47. Available at www.sjlibrary.org

Guest, M. (1998) Spoken grammar: easing the transitions. *The Language Teacher Online. The Japan Association for Language Teaching* 22 (6): 35-46 http://langue.hyper.chubu.ac.jp/jalt/pub/tlt/98/jun/guest.html (Last Accessed 10 March 2011)

Harmer, J. (1990) *The Practice of English Language Teaching.* London: Longman.

Hedge, T. (2000) *Teaching and Learning in the Langugae Classroom.* Oxford: Oxford Univeristy Press.

Hsu, W. (2009)) *2009./*College English textbooks for general purposes: a corpus-based analysis of lexical coverage *Electronic Journal of Foreign Language Teaching 6(1) 42-62*: http://e-flt.nus.edu.sg/v6n12009/hsu.pdf. (Last Accessed 10 March 2011)

Hyland, K. (2008a) As can be seen: lexical bundles and disciplinary variation. *English for Specific Purposes* 27 4-21 Available from www.ebrary.com

Hyland, K. (2008b) Academic clusters: text patterning in published and postgraduate writing. *International Journal of Applied Linguistics* 18(1) 41-62 Available from www.ebrary.com

Jenkins, J. (2006) Current english perspectives on teaching world Englishes and English as a lingua franca. *TESOL Quarterly 40 (1)* , 157-175.

Jenkins, J. (2007) *English as a Lingua Franca: Attitude and Identity.* Oxford: Oxford University Press.

Jenkins, J. (2008) English as a Lingua Franca: Keynote at JACET 47th Annual Convention: http://jacet.org/ (Last Accessed 18 February 2011)

Kachru, B. (1985) Standards, codification and sociolinguistic realism: The english language in the outer circle, in R. a. Quirk (ed.), *English in the World:Teaching and Learning the Language and Literatures* pp. 11-30 Cambridge: Camridge University Press.

Kankaanranta, A. and L. Salminen (2010) "English? - oh, it's just work!": a study of BELF user's perceptions. *English for Specific Purposes.* 29 (1): 204-209 Available from www.sciencedirect.com

Kumar, R. and T. Bunyakarte (2010a) *BG 2001 English IV* 4th Edition. Bangkok: Assumption University Press

Kumar, R. and R. Dugdale and R. Miles (2010b) *BG2000 English III* 4th Edition. Bangkok: Assumption University Press

McArthur, T. (2005) English as an Asian Language: http://www.accu.or.jp/appreb/09/pdf33-2/33-2P003-004.pdf. (Last Accessed 18 February 2011)

McEnery, T. and A. Wilson (1996) *Corpus Linguistics.* Edinburgh: Edinburgh Univeristy Press.

McEnery, T. and A. Wilson (2001) *Corpus Linguistis: An Introduction 2nd Edition.* Edinbugh: Edinbugh Univeristy Press.

Merriam-Webster online Dictionary, An Encyclopedia Britannica Company, http://www.merriam-webster.com/ (Last Accessed 11 May 2011)

Mollin, S. (2006) English as a lingua franca: a new variety in the new expanding circle: http://gupea.ub.gu.se/dspace/bitstream/2077/3149/1/5-2-Mollin.pdf. (Last Accessed 18 February 2011)

Nation, P. and R. Waring (1997) Vocabulary size, text, coverage and word lists, in N. A. Schmitt (ed.), *Vocabulary Description, Acquisition and Pedagogy* pp. 6-19. Cambridge: Cambridge University Press.

Nelson, M. (2005) Semantic associations in business English: a corpus-based analysis: http://www.sciencedirect.com (Last Accessed 18 February 2011)

Olivera, P. and A.A. Bano (2003) Jose Coca Prados 2002: Ingles para quimica e ingenieria quimica. *The Journal of the Spanish Association for Anglo-American Studies 25 (2),* 143-148.

Pakir, A. (2009) Lexical variations in "Singapore English": linguistic description and language education, in Y. Kawaguchi M. Minegishi J. Durand (eds) *Corpus Analysis and Variation in Linguistics.* Amsterdam: John Benjamins Publishing Company. Available from www.ebrary.com

Partington, A. (1998) *Patterns and Meanings: Using Corpora for English Language Research and Teaching.* Amsterdam: John Benjamins Publishing Company. Available at www.ebrary.com

Planken, B. (2005) Managing rapport in lingua franca sales negotiations: a comparison of professional and aspiring negotiators. *English for Specific Purposes.* 24, 381-400

Powell, M., R. Martinez and R. Jillett (2004). *New Business Matters.* London: Thomson.

Ranalli, J. (2003) ELT coursebooks in the age of corpus linguistics:constraints and possibilities: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.132.6585&rep=rep1&type= pdf (Last Accessed 18 February 2011)

Schmitt, N. (2002) Using corpora to teach and assess vocabualry, in M. Tan (ed.), *Corpus Studies in Language Education* pp. 31-42. Bangkok: Assumption University Press.

Schmitt, N. (2000) *Vocabulary in Language Teaching.* Cambridge: Cambridge University Press.

Scott, M. (2010) WordSmith Tools V. 5.0. Liverpool: Lexical Analysis Software

Seidlhofer, B (2002) Habeas corpus and divide et impera: "Global English" and applied linguistics', in K. S. Miller and P. Thompson (eds) *Unity and Diversity of Language Use.* pp. 198– 215 London: Continuum International Publishing.

Seidlhofer, B. (2005) English as a lingua franca. *ELT Journal 49 (4)* , 339-341.

Simpson-Vlach, R.C. and S. Leicher (2006) *The MICASE Handbook. A Resource for Users of the Michigan Corpus of Academic Spoken English.* Ann Arbor: University of Michigan Press

Sinclair, J. (1991) Introduction, in J. Sinclair and R. Carter (eds.) *Corpus Concordance Collocation.* pp. 2-9 Oxford: Oxford University Press

Someya, Y. (1999) Thesis: A corpus-based study of lexical and grammatical features of written business english: http://someya-net.com/index2.html (Last Accessed 18 February 2011)

Talja, S., K. Tuominen, and R. Savolainen (2005) "Isms" in information science: constructivism, collectivism and constructionism. *Journal of Documentation.* 61: 79-101 Emerald Group Publishing Limited. www.psu.edu (Last Accessed October 2011)

Tan, M. (2002) Introduction, in M. Tan (ed.), *Corpus Studies in Language Education* pp. 2-5. Bangkok: IELE Press.

Trappe, T. and G. Tullis (2005) *Intelligent Business.* Essex: Pearson Longman.

Tribble, C. (2002) Looking at the news: using newspaper corpus in English language education, in M. Tan (ed.), *Corpus Studies in Language Education* pp. 145-150. Bangkok: Assumption University Press.

Ur, J. (1999) *A Course in Language Teaching.* Cambridge: Cambridge University Press.

Wala, D. A. (2003) A coursebook is what it is because of what it has to do: an editor's perspective, in B. Tomlinson (ed.), *Developing Materials for Language Teaching* pp. 58-61. London: Continuum.

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

Appendix I: Top 100 Bundles

N	Word	Freq.	%	Texts	%
1	PER CENT OF	558	0.065553971	326	18.43891335
2	IN THE FIRST	496	0.058270197	314	17.76018143
3	IS EXPECTED TO	447	0.052513666	345	19.5135746
4	AS WELL AS	436	0.051221382	331	18.72171974
5	THE FIRST QUARTER	341	0.040060762	186	10.5203619
6	THE END OF	330	0.038768478	257	14.53619957
7	OF THIS YEAR	327	0.038416039	243	13.74434376
8	YEAR ON YEAR	294	0.034539189	164	9.276018143
9	SAID THE COMPANY	273	0.032072105	192	10.85972881
10	THE NUMBER OF	228	0.026785495	176	9.954751015
11	HE SAID THE	217	0.025493212	181	10.23755646
12	ONE OF THE	206	0.024200929	174	9.841629028
13	IN THE SECOND	202	0.023731008	152	8.597285271
14	THE STOCK EXCHANGE	185	0.021733845	170	9.615385056
15	IN TERMS OF	184	0.021616364	154	8.710407257
16	IN LINE WITH	183	0.021498883	164	9.276018143
17	MANAGING DIRECTOR OF	182	0.021381402	155	8.766968727
18	PER CENT TO	182	0.021381402	108	6.108597279
19	STOCK EXCHANGE OF	182	0.021381402	169	9.558823586
20	EXCHANGE OF THAILAND	180	0.021146443	165	9.332579613
21	THE COMPANY HAS	176	0.020676522	141	7.975112915
22	ARE EXPECTED TO	173	0.020324081	147	8.314479828
23	CENT OF THE	168	0.019736679	124	7.0135746
24	IN ORDER TO	167	0.0196192	146	8.257918358
25	OF THE THAI	166	0.01950172	138	7.805429935
26	PER CENT FROM	166	0.01950172	104	5.882352829
27	THE MIDDLE EAST	165	0.019384239	123	6.957013607
28	OF THE YEAR	163	0.019149279	131	7.409502029
29	BY THE END	162	0.019031798	133	7.522624493
30	PER CENT IN	162	0.019031798	107	6.052036285
31	IN THE NEXT	156	0.018326918	129	7.296380043
32	THE NEW GOVERNMENT	156	0.018326918	93	5.26018095
33	DUE TO THE	155	0.018209437	141	7.975112915
34	THE COMPANY IS	153	0.017974475	126	7.126697063
35	THE COMPANY WILL	152	0.017856997	120	6.787330151
36	OVER THE NEXT	149	0.017504556	126	7.126697063
37	PRESIDENT OF THE	149	0.017504556	127	7.183258057
38	EXPECTED TO BE	146	0.017152114	134	7.579185486
39	BE ABLE TO	144	0.016917154	131	7.409502029
40	THE PRICE OF	141	0.016564714	95	5.373302937
41	ACCORDING TO THE	140	0.016447233	132	7.466063499
42	BANK OF THAILAND	140	0.016447233	121	6.843891621
43	BILLION BAHT IN	138	0.016212273	106	5.995475292
44	EXECUTIVE VICE PRESIDENT	135	0.015859831	114	6.447963715
45	THE SAME PERIOD	135	0.015859831	104	5.882352829
46	IN THE PAST	133	0.015624871	110	6.221719265
47	THE UNITED STATES	133	0.015624871	110	6.221719265
48	END OF THIS	128	0.01503747	103	5.825791836
49	THE BANK OF	124	0.014567549	107	6.052036285
50	BECAUSE OF THE	123	0.01445007	108	6.108597279
51	FOR THE FIRST	123	0.01445007	98	5.542986393
52	AS A RESULT	121	0.014215109	111	6.278280735
53	PER CENT AND	121	0.014215109	97	5.4864254
54	VICE PRESIDENT OF	121	0.014215109	108	6.108597279

55	IN THE FUTURE	117	0.013745188	104	5.882352829
56	IN TRADE WORTH	114	0.013392747	92	5.203619957
57	PART OF THE	113	0.013275267	101	5.712669849
58	THE SECOND QUARTER	113	0.013275267	77	4.355203629
59	THE SECOND HALF	111	0.013040306	90	5.090497971
60	GENERAL OF THE	110	0.012922826	103	5.825791836
61	PER CENT THIS	108	0.012687866	79	4.468325615
62	VICE PRESIDENT FOR	108	0.012687866	96	5.429864407
63	PER CENT YEAR	106	0.012452905	67	3.789592743
64	CENT THIS YEAR	105	0.012335425	77	4.355203629
65	PERIOD LAST YEAR	105	0.012335425	81	4.581448078
66	CENT YEAR ON	104	0.012217944	65	3.676470518
67	CHAIRMAN OF THE	104	0.012217944	84	4.751131058
68	SAME PERIOD LAST	104	0.012217944	81	4.581448078
69	IN THE REGION	103	0.012100465	81	4.581448078
70	OF THE TOTAL	103	0.012100465	90	5.090497971
71	THE CENTRAL BANK	103	0.012100465	62	3.5067873
72	THE FIRST HALF	103	0.012100465	67	3.789592743
73	SMALL AND MEDIUM	101	0.011865504	89	5.033936501
74	TO INCREASE THE	101	0.011865504	90	5.090497971
75	IN THE MIDDLE	100	0.011748023	84	4.751131058
76	CHIEF EXECUTIVE OFFICER	99	0.011630544	90	5.090497971
77	THE GOVERNMENT SHOULD	99	0.011630544	64	3.619909525
78	EARTHQUAKE AND TSUNAMI	98	0.011513063	90	5.090497971
79	IN THE COUNTRY	98	0.011513063	88	4.977375507
80	IN THE SAME	98	0.011513063	81	4.581448078
81	THE COMPANY WOULD	98	0.011513063	83	4.694570065
82	THIS YEAR TO	98	0.011513063	90	5.090497971
83	THE COMPANY HAD	95	0.011160622	79	4.468325615
84	THIS YEAR AND	95	0.011160622	84	4.751131058
85	THE COMPANY EXPECTS	92	0.010808182	83	4.694570065
86	SAID YESTERDAY THAT	91	0.010690701	84	4.751131058
87	A LOT OF	88	0.010338261	73	4.128959179
88	END OF THE	88	0.010338261	81	4.581448078
89	FIRST QUARTER OF	88	0.010338261	73	4.128959179
90	THE COST OF	87	0.010220781	70	3.959275961
91	THE PRIVATE SECTOR	87	0.010220781	69	3.902714968
92	OVER THE PAST	85		80	4.524887085
93	TO INVEST IN	85		72	4.072398186
94	THE NEXT FIVE	84		68	3.846153736
95	ADDED THAT THE	82		76	4.298642635
96	LINE WITH THE	82		76	4.298642635
97	DIRECTOR OF THE	81		74	4.185520172
98	GROWTH IN THE	81		67	3.789592743
99	QUARTER OF THIS	81		70	3.959275961
100	THE THAI MARKET	81		68	3.846153736

Appendix II: MAS Bundles

N	Word	Freq.	%	Texts	%
1	PER CENT OF	23	0.097511344	10	22.72727203
2	OF THE COMPANY	10	0.042396236	7	15.909091
3	ONE OF THE	10	0.042396236	7	15.909091
4	FIVE PER CENT	8	0.033916987	6	13.63636398
5	CENT OF THE	7	0.029677365	5	11.36363602
6	IN THE USA	7	0.029677365	6	13.63636398
7	OF THE MARKET	7	0.029677365	6	13.63636398
8	SOME OF THE	7	0.029677365	6	13.63636398
9	THERE IS A	7	0.029677365	4	9.090909004
10	AS WELL AS	6	0.025437741	6	13.63636398
11	HAVE TO BE	6	0.025437741	6	13.63636398
12	OF THE WORLD'S	6	0.025437741	5	11.36363602
13	THERE ARE MANY	6	0.025437741	6	13.63636398
14	IN THE FUTURE	5	0.021198118	4	9.090909004
15	IN THE WORLD	5	0.021198118	5	11.36363602
16	IT CAN BE	5	0.021198118	5	11.36363602
17	OF A COMPANY	5	0.021198118	4	9.090909004
18	OF THE WORLD	5	0.021198118	5	11.36363602
19	ON THE OTHER	5	0.021198118	4	9.090909004
20	THE AGE OF	5	0.021198118	5	11.36363602
21	THE COMPANY CAN	5	0.021198118	1	2.272727251
22	THE NUMBER OF	5	0.021198118	4	9.090909004
23	THE WORLD S	5	0.021198118	4	9.090909004
24	A HANDFUL OF	4	0.016958494	3	6.818181992
25	A LONG TERM	4	0.016958494	4	9.090909004
26	A LOT OF	4	0.016958494	3	6.818181992
27	AND A HALF	4	0.016958494	3	6.818181992
28	BUSINESS TO BUSINESS	4	0.016958494	3	6.818181992
29	HAS NEVER BEEN	4	0.016958494	4	9.090909004
30	IF THEY ARE	4	0.016958494	4	9.090909004
31	IN ORDER TO	4	0.016958494	4	9.090909004
32	IN THE MARKET	4	0.016958494	2	4.545454502
33	IN THE PAST	4	0.016958494	4	9.090909004
34	IN THE US	4	0.016958494	3	6.818181992
35	IS MORE COMMON	4	0.016958494	3	6.818181992
36	IS NO LONGER	4	0.016958494	4	9.090909004
37	NEED TO TAKE	4	0.016958494	3	6.818181992
38	OF THE MOST	4	0.016958494	4	9.090909004
39	ON THE BACK	4	0.016958494	3	6.818181992
40	ON THE INTERNET	4	0.016958494	3	6.818181992
41	POWER OF THE	4	0.016958494	3	6.818181992
42	TEND TO BE	4	0.016958494	3	6.818181992
43	THE BUSINESS TO	4	0.016958494	2	4.545454502
44	THE COST OF	4	0.016958494	4	9.090909004
45	THE FORM OF	4	0.016958494	3	6.818181992
46	THE VALUE OF	4	0.016958494	3	6.818181992
47	TO LOOK AT	4	0.016958494	4	9.090909004
48	USED IN THE	4	0.016958494	3	6.818181992
49	A CHANCE TO	3	0.01271887	3	6.818181992
50	A HIGH PRICE	3	0.01271887	2	4.545454502
51	A NEW COMPANY	3	0.01271887	3	6.818181992
52	ACCORDING TO THE	3	0.01271887	3	6.818181992
53	AND ONE OF	3	0.01271887	3	6.818181992
54	AS MUCH AS	3	0.01271887	3	6.818181992

55	BANK OF AMERICA	3	0.01271887	1	2.272727251
56	BILLION A YEAR	3	0.01271887	3	6.818181992
57	BUT IF YOU	3	0.01271887	2	4.545454502
58	CAN BE A	3	0.01271887	3	6.818181992
59	CAN BE SOLD	3	0.01271887	2	4.545454502
60	COMPANY FOR EXAMPLE	3	0.01271887	1	2.272727251
61	FROM THE BANK	3	0.01271887	2	4.545454502
62	GET A JOB	3	0.01271887	3	6.818181992
63	GOING TO BE	3	0.01271887	2	4.545454502
64	HAVE MORE CHOICE	3	0.01271887	1	2.272727251
65	HOW DO YOU	3	0.01271887	2	4.545454502
66	IF THEY WERE	3	0.01271887	3	6.818181992
67	IN A NEW	3	0.01271887	3	6.818181992
68	IN OTHER WORDS	3	0.01271887	3	6.818181992
69	IN THE BUSINESS	3	0.01271887	2	4.545454502
70	IN THE COMPANY	3	0.01271887	3	6.818181992
71	IN THE END	3	0.01271887	2	4.545454502
72	IN THE FORM	3	0.01271887	3	6.818181992
73	IS GOING TO	3	0.01271887	3	6.818181992
74	LITTLE MORE THAN	3	0.01271887	3	6.818181992
75	LOOK AT THE	3	0.01271887	2	4.545454502
76	MORE AND MORE	3	0.01271887	3	6.818181992
77	MORE CHOICE OVER	3	0.01271887	1	2.272727251
78	NEED TO BE	3	0.01271887	3	6.818181992
79	NEEDS TO BE	3	0.01271887	3	6.818181992
80	OF THE BEST	3	0.01271887	2	4.545454502
81	OF THE FUTURE	3	0.01271887	3	6.818181992
82	OF THE NEW	3	0.01271887	3	6.818181992
83	OF THE PRODUCTS	3	0.01271887	2	4.545454502
84	OUT OF THE	3	0.01271887	2	4.545454502
85	OUT OF WORK	3	0.01271887	2	4.545454502
86	OVER THE LAST	3	0.01271887	3	6.818181992
87	PART OF THE	3	0.01271887	3	6.818181992
88	PEOPLE TRY TO	3	0.01271887	3	6.818181992
89	PREPARED TO PAY	3	0.01271887	3	6.818181992
90	SEGMENTS OF THE	3	0.01271887	3	6.818181992
91	SEVEN PER CENT	3	0.01271887	2	4.545454502
92	SIR JAMES GOLDSMITH	3	0.01271887	2	4.545454502
93	TERMS OF THE	3	0.01271887	3	6.818181992
94	THAT THE COMPANY	3	0.01271887	3	6.818181992
95	THAT THE CUSTOMER	3	0.01271887	2	4.545454502
96	THAT WOULD BE	3	0.01271887	3	6.818181992
97	THE ANSWER IS	3	0.01271887	3	6.818181992
98	THE AREA OF	3	0.01271887	2	4.545454502
99	THE BACK OF	3	0.01271887	2	4.545454502
100	THE BOARD OF	3	0.01271887	3	6.818181992
101	THE COMPANY IS	3	0.01271887	3	6.818181992
102	THE COMPANY S	3	0.01271887	3	6.818181992
103	THE CUSTOMER PAYS	3	0.01271887	1	2.272727251
104	THE CUSTOMER WILL	3	0.01271887	3	6.818181992
105	THE DEATH OF	3	0.01271887	2	4.545454502
106	THE FOUNDER OF	3	0.01271887	3	6.818181992
107	THE LAUNCH OF	3	0.01271887	3	6.818181992
108	THE MAJORITY OF	3	0.01271887	3	6.818181992
109	THE OTHER HAND	3	0.01271887	3	6.818181992
110	THE OTHER PERSON	3	0.01271887	1	2.272727251
111	THE POWER OF	3	0.01271887	3	6.818181992
112	THE SAME THING	3	0.01271887	2	4.545454502

113	THE STOCK MARKET	3	0.01271887	2	4.545454502
114	THE TRUTH IS	3	0.01271887	3	6.818181992
115	THE UNITED STATES	3	0.01271887	3	6.818181992
116	THERE ARE VARIOUS	3	0.01271887	2	4.545454502
117	THERE IS NO	3	0.01271887	3	6.818181992
118	THEY ARE MORE	3	0.01271887	3	6.818181992
119	THEY WANT TO	3	0.01271887	3	6.818181992
120	THIS IS A	3	0.01271887	3	6.818181992
121	THIS IS THE	3	0.01271887	2	4.545454502
122	THIS LOOKS AT	3	0.01271887	1	2.272727251
123	TO BE A	3	0.01271887	3	6.818181992
124	TO BE THE	3	0.01271887	3	6.818181992
125	TO MAKE MONEY	3	0.01271887	2	4.545454502
126	TO SELL IT	3	0.01271887	3	6.818181992
127	TO STAY IN	3	0.01271887	3	6.818181992
128	TO TAKE RISKS	3	0.01271887	3	6.818181992
129	TWENTY PER CENT	3	0.01271887	2	4.545454502
130	WILL HAVE TO	3	0.01271887	3	6.818181992
131	WITH THE CUSTOMER	3	0.01271887	1	2.272727251
132	WORD OF MOUTH	3	0.01271887	3	6.818181992
133	YOU HAVE TO	3	0.01271887	2	4.545454502
134	YOU WANT TO	3	0.01271887	3	6.818181992
135	YOUR JOB BECAUSE	3	0.01271887	1	2.272727251

Appendix III: LIM Bundles

N	Word	Freq.	%	Texts	%
1	AS WELL AS	10	0.082733512	7	30.43478203
2	PER CENT OF	8	0.066186816	3	13.04347801
3	CREDIT CARD PROGRAM	6	0.049640108	1	4.347826004
4	NUCLEAR POWER PLANT	6	0.049640108	1	4.347826004
5	THE END OF	6	0.049640108	4	17.39130402
6	BE ABLE TO	5	0.041366756	4	17.39130402
7	IN THE US	5	0.041366756	2	8.695652008
8	OF LEAD PAINT	5	0.041366756	1	4.347826004
9	THE GLOBAL BUSINESS	5	0.041366756	1	4.347826004
10	A PRIVATE COMPANY	4	0.033093408	1	4.347826004
11	AND MEDIUM SIZED	4	0.033093408	1	4.347826004
12	AROUND THE COUNTRY	4	0.033093408	1	4.347826004
13	AROUND THE WORLD	4	0.033093408	3	13.04347801
14	BENEFIT FROM THE	4	0.033093408	1	4.347826004
15	BY CASH OR	4	0.033093408	1	4.347826004
16	CASH OR CHEQUE	4	0.033093408	1	4.347826004
17	CONSUMER PRODUCTS INDUSTRY	4	0.033093408	1	4.347826004
18	CREDIT PLUS PROGRAM	4	0.033093408	1	4.347826004
19	CREDIT TERM OF	4	0.033093408	1	4.347826004
20	EXCLUSIVE CREDIT CARD	4	0.033093408	1	4.347826004
21	GLOBAL BUSINESS LEADER	4	0.033093408	1	4.347826004
22	HALF OF THE	4	0.033093408	3	13.04347801
23	HE ADDED THAT	4	0.033093408	1	4.347826004
24	IN LINE WITH	4	0.033093408	2	8.695652008
25	IN TERMS OF	4	0.033093408	3	13.04347801
26	LEVELS OF LEAD	4	0.033093408	1	4.347826004
27	MEDIUM SIZED RETAILERS	4	0.033093408	1	4.347826004
28	MODERN AND TRADITIONAL	4	0.033093408	1	4.347826004
29	OF THIS YEAR	4	0.033093408	3	13.04347801
30	ONE OF THE	4	0.033093408	3	13.04347801
31	SMALL AND MEDIUM	4	0.033093408	1	4.347826004
32	THE CONSUMER PRODUCTS	4	0.033093408	1	4.347826004
33	THE CREDIT CARD	4	0.033093408	1	4.347826004
34	THE NUCLEAR POWER	4	0.033093408	1	4.347826004
35	THROUGHOUT THE COUNTRY	4	0.033093408	1	4.347826004
36	TRADITIONAL RETAIL STORES	4	0.033093408	1	4.347826004
37	UNILEVER THAI HOLDINGS	4	0.033093408	1	4.347826004
38	A LOT OF	3	0.024820054	2	8.695652008
39	ARE ALSO A	3	0.024820054	2	8.695652008
40	BE USED BY	3	0.024820054	2	8.695652008
41	BE USED TO	3	0.024820054	2	8.695652008
42	BY THE END	3	0.024820054	3	13.04347801
43	CARREFOUR MINI HYPERMARKET	3	0.024820054	1	4.347826004
44	CARREFOUR SUAN LUANG	3	0.024820054	1	4.347826004
45	CENT OF THE	3	0.024820054	2	8.695652008
46	CENT SALES GROWTH	3	0.024820054	1	4.347826004
47	COULD BE USED	3	0.024820054	2	8.695652008
48	DAY A WEEK	3	0.024820054	1	4.347826004
49	GALACTIC SUITE SPACE	3	0.024820054	1	4.347826004
50	HYPERMARKET FORMAT WITH	3	0.024820054	1	4.347826004
51	IN A PRIVATE	3	0.024820054	1	4.347826004
52	IN ORDER TO	3	0.024820054	3	13.04347801
53	IN THAILAND IS	3	0.024820054	3	13.04347801
54	IN THE PAST	3	0.024820054	3	13.04347801
55	IN THE SECOND	3	0.024820054	2	8.695652008
56	LINE WITH THE	3	0.024820054	2	8.695652008

57	MILLIONS OF PEOPLE	3	0.024820054	2	8.695652008
58	NEEDS TO BE	3	0.024820054	2	8.695652008
59	OF THE WORLD	3	0.024820054	2	8.695652008
60	ONE DAY A	3	0.024820054	1	4.347826004
61	OVER THE NEXT	3	0.024820054	2	8.695652008
62	OVER THE PAST	3	0.024820054	2	8.695652008
63	PER CENT AND	3	0.024820054	2	8.695652008
64	PER CENT SALES	3	0.024820054	1	4.347826004
65	POLICY OF PROMOTING	3	0.024820054	2	8.695652008
66	SUITE SPACE RESORT	3	0.024820054	1	4.347826004
67	THE ABILITY TO	3	0.024820054	1	4.347826004
68	THE COMPANY HAS	3	0.024820054	2	8.695652008
69	THE DEVELOPMENT OF	3	0.024820054	3	13.04347801
70	THE GALACTIC SUITE	3	0.024820054	1	4.347826004
71	THE MOST IMPORTANT	3	0.024820054	2	8.695652008
72	THE NUCLEAR PLANT	3	0.024820054	1	4.347826004
73	THE SUPPLY CHAIN	3	0.024820054	2	8.695652008
74	THE UNIVERSITY OF	3	0.024820054	2	8.695652008
75	THEIR VISITS TO	3	0.024820054	2	8.695652008
76	THERE IS NO	3	0.024820054	3	13.04347801
77	TO BUILD THE	3	0.024820054	1	4.347826004
78	WILL BE USED	3	0.024820054	2	8.695652008
79	WITH THE GOVERNMENT'S	3	0.024820054	2	8.695652008