การพัฒนารายการคำศัพท์วิชาการทางภาษาอังกฤษธุรกิจ
โดยใช้วิธีการด้านภาษาศาสตร์คลังข้อมูล
Constructing an Academic Word List of Business English:
A Corpus-Based Approach

อังคณา ทองพูน พัฒนศร

Angkana Tongpoon Patanasorn

คณะมนุษยศาสตร์และสังคมศาสตร์

Faculty of Humanities and Social Sciences

มหาวิทยาลัยขอนแก่น

Khon Kaen University

บทคัดย่อ

   งานวิจัยนี้เป็นงานวิจัยด้านภาษาศาสตร์คลังข้อมูล มีวัตถุประสงค์เพื่อพัฒนา
รายการคำศัพท์วิชาการด้านภาษาอังกฤษธุรกิจที่ยังไม่เคยมีการศึกษามาก่อน
โดยคอร์ปัสหรือภาษาศาสตร์คลังข้อมูลทางภาษาอังกฤษธุรกิจซึ่งในงานวิจัยนี้เรียกว่า
KKU-BE Corpus พัฒนามาจากหลักการการสร้างภาษาศาสตร์คลังข้อมูลจากการ
ทบทวนวรรณกรรมและงานวิจัยที่เกี่ยวข้อง ซึ่งจำนวนคำในภาษาศาสตร์คลังข้อมูล
KKU-BE Corpus มีจำนวนประมาณ 16 ล้านคำ โดยรวบรวมเนื้อหาจากตำรา บทความ
วิจัย บทความออนไลน์ และหนังสือพิมพ์จากสาขาวิชาทางภาษาอังกฤษธุรกิจ 8 สาขา
วิชา ประกอบด้วย การบัญชี การตลาด การโฆษณา การเงิน กฎหมายธุรกิจ การท่องเที่ยว
เศรษฐศาสตร์ และการจัดการ โดยใช้เกณฑ์ในการคัดเลือกคำและพัฒนารายการ
คำศัพท์วิชาการด้านภาษาอังกฤษธุรกิจที่สำคัญในการเรียนการสอนด้านภาษาอังกฤษ
ธุรกิจ 3 ข้อ คือ คำศัพท์ที่จะคัดเลือกต้องมีความเป็นเฉพาะทางทางศาสตร์ภาษาอังกฤษ
ธุรกิจ มีความถี่ตามเกณฑ์ที่กำหนด และมีการครอบคลุมสาขาวิชาย่อยของภาษา
อังกฤษธุรกิจ จากผลการวิจัย พบว่ารายการคำศัพท์วิชาการด้านภาษาอังกฤษธุรกิจ

หรือ BEAWL ที่พัฒนาขึ้นประกอบด้วยคำหลัก 415 คำ ซึ่งเมื่อรวมคำหลักและคำสมาชิก
ของคำหลักด้วยจะมีคำทั้งหมด 1,572 คำ รายการคำศัพท์วิชาการด้านภาษาอังกฤษ
ธุรกิจจำนวน 415 คำนี้สามารถแบ่งออกเป็นรายการคำศัพท์ย่อย 7 รายการ ในการแบ่ง
เป็นรายการย่อยของคำศัพท์วิชาการทางภาษาอังกฤษธุรกิจนี้ใช้วิธีการใหม่
โดยพิจารณาเกณฑ์ด้านการครอบคลุมของคำศัพท์เป็นเกณฑ์หลักก่อนการพิจารณา
ด้านความถี่ของคำ โดยในรายการย่อย 3 ชุดแรกเป็นคำที่เกิดขึ้นในสาขาย่อยของภาษา
อังกฤษธุรกิจทั้ง 8 สาขาวิชา ส่วนในรายการย่อยชุดที่ 4-7 เป็นคำที่เกิดขึ้นอย่างน้อย
ใน 4 สาขาวิชาขึ้นไป รายการคำศัพท์วิชาการด้านภาษาอังกฤษธุรกิจที่สร้างขึ้นนี้
ครอบคลุมคำใน KKU-BE Corpus ประมาณ 12.66 เปอร์เซ็นต์ ซึ่งรายการคำศัพท์
วิชาการด้านภาษาอังกฤษธุรกิจนี้มีประโยชน์อย่างมากในการเรียนการสอนภาษา
อังกฤษธุรกิจซึ่งเป็นการสอนภาษาอังกฤษเฉพาะทาง คำศัพท์เหล่านี้เป็นคำที่จำเป็น
และสำคัญที่จะช่วยให้การเรียนและการพัฒนาคำศัพท์ของผู้เรียนมีประสิทธิภาพ
มากยิ่งขึ้น

คำสำคัญ: คำศัพท์วิชาการ ภาษาอังกฤษธุรกิจ การศึกษาด้านคลังข้อมูล

Abstract

This corpus-based study aims to construct an academic word list for Business English (BE) which has not yet been conducted before. To do this, a corpus of Business English was compiled: the KKU-BE Corpus. The corpus was developed applying principles reviewed from previous literature and related studies. It contains approximately 16 million words, covering four text types (i.e., book, journal articles, websites, and newspaper) across the eight major disciplines of Business English: accounting, marketing, advertising, finance, business law, tourism, economics, and management. Three principles of specialization, frequency, and range were adopted to guide the construction of the Business English Academic Wordlist (BEAWL). Based on the findings of this study, the BEAWL contains 415 headwords and 1572 family members. The list was further divided into sub-lists using an innovative method. The sub-lists are arranged first by word representation across subject areas and then by word frequency. The words in the first three sub-lists occur in all eight subject areas, and those in the other four sub-lists appear in four to seven subjects areas. The BEAWL accounts for 12.66% of the total BE corpus. These findings can be applied in the teaching of Business English and English for Specific Purposes (ESP); this word list can help language learners acquire the vocabulary they need to be successful in their learning within the time constraints of their academic lives.

Keywords: academic words, Business English, corpus-based studies

Acquiring a relevant vocabulary is crucial to developing fluency in another language. Language learners must acquire an especially large and specific lexicon in order to be successful when conversing, reading, and writing in academic settings (Coady, 1997; Donley & Reppen, 2001; Grabe & Stoller, 2002; Nation, 2001, 2013). An extensive vocabulary helps learners to understand written and spoken texts, as well as to perform different skills in their second language more efficiently (Meara, 1990). With a large vocabulary size, such learners are able to produce precise, high-quality writing (Read, 2000). Without an appropriate vocabulary size, language learners may have difficulty understanding and clearly producing the target language (Koda, 2005; Laufer, 1997; Lee, 2003).

Previous studies have shown that academic words are particularly challenging for English for Specific Purposes (ESP) learners compared to other types of words in specialized texts (Chung & Nation, 2003; Cobb & Horst, 2004; Coxhead & Nation, 2001; Shaw, 1991; Thurstun & Candlin, 1998). This is because academic words are frequently found across disciplines and in different types of texts such as textbooks, research articles, and news articles. Therefore, providing ESP learners with a list of academic words will help them to better understand the core ideas of their academic specialties as well as to achieve competence in their target areas (Durrant, 2016; Hutchinson & Waters, 1996). Moreover, such a list will help learners with little time to their language-learning efforts for maximal success.

One way to create such an academic wordlist is through the use of corpus linguistics (Dudley-Evans & St. John, 2011). In the 1990s, scholars such as Dudley-Evans and St. John (1996) and Robinson (1991) called for more research to help produce BE wordlists. In the years that have followed, many scholars have answered this call. The specialized vocabulary employed in

written texts has been examined in various disciplines, including medicine (Chen & Ge, 2007; Lie & Liu, 2016; Wang, Liang, & Ge, 2008), agriculture (Martinez, Beck, & Panza, 2009), nursing (Yang, 2015), and applied linguistics (Vongpumivitch, Huang, & Chang, 2009). These studies show that academic texts contain a high percentage of academic words, approximately 9.06-12.24% of the overall word count. The lists of academic words created from these texts vary according to discipline, of course, as well as text type and text section. The academic wordlists built from specialized corpus tied to the specific areas are considerably useful for ESP learners because they provide lists of words learners actually need, while the general academic word lists may expose learners to too many general words that they may not encounter in their areas (Martinez, Beck, & Panza, 2009). Accordingly, they may be less valuable to learners to invest their limited time for learning.

Only a handful lexical research studies in the field of Business English using a corpus-based analysis, however, were found such as Nelson (2000), Chen, Hu, and Ho (2004), and Wood and Appel (2014). Nelson (2000) compiled a BE corpus and then compiled a list of high-frequency words for that corpus. He then compared that list to a list of high-frequency words compiled from textbooks in the field. Despite what one may think, the two lists were different. Importantly, Nelson's study did not examine academic words in the BE field.

Nonetheless, other studies have substantiated the notion that the words teachers and learners think are important to acquiring BE proficiency are those used most frequently in spoken and written texts. For example, Chen, Hu, and Ho (2004) compared the use of academic words in the abstracts of business and management journals to the use of those words in Taiwanese journal abstracts using the Vocabulary Profile (www.lextutor.ca). The results revealed that the Academic Wordlist (AWL) words occurred more frequently in the

business and management journal abstracts than the Taiwanese journal abstracts. A recent study by Wood and Appel (2014) examined multiword constructions (MWCs) in first-year business and engineering university textbooks and EAP textbooks. They found that EAP textbooks under-represent MWCs in their materials.

Even though these studies shed some light on what is most important to the teaching and learning of Business English, they leave many gaps in the knowledge of BE. Only one of the aforementioned studies deals specifically with academic vocabulary: Chen, Hu, and Ho's study. Further, Chen, Hu, and Ho's corpus was a compilation of business and management abstracts; it did not cover the wide range of topics and text types that ESP learners may encounter in their course of study. As we know, different specialized fields may be diverse in terms of lexical forms and use (Chen & Ge, 2007). The more specialized the field is, the more specialized word forms and meanings it employs.

Thus, it is important to make a complete list of frequent academic words for learning and teaching each specific discipline of BE (Chen & Ge, 2007; Hyland & Tse, 2007; Martínez, Beck, & Panza, 2009; Wang, Liang, & Ge, 2008). Such graded lists when conducted will be a valuable resource for teachers, students, as well as course designers (Nation, 2016). Coxhead (2016), in her own words, states that, "Much more work remains to be done in vocabulary studies and EAP/ESP" (p.183). The present paper aims to address existing gaps in the literature by constructing lists of frequent academic words from a new corpus of Business English, which will be referred to as the KKU-BE Corpus.

Types of Vocabulary and Academic words

According to Nation (1990, 2001, 2008) and Schmitt (2000), vocabulary is divided into four categories: high frequency words, academic words, technical words, and low-frequency words. High frequency words are words that commonly occur in any language contexts, in both writing and speaking, and in both formal and informal texts (Nation, 2008). Most lists of high-frequency words contain approximately 2,000 word families, which are mainly comprised of short, common words. Many of them are function words (e.g., *a*, *in*, *of*, and *this*) and content words (nouns, verbs, adjectives, and adverbs) such as *bird*, *buy*, *difficult*, and *fast*. This group of words accounts for up to 80% of the running words in any given text. The most well-known list of frequent words is *A General Service List of English Words* (GSL), which was created by Michael West in 1953. These 2,000 words are essential for those people learning English as Foreign/Second Language to acquire before they can move onto more specialized areas of study (Grabe, 2009; Nation & Hwang, 1995).

The second vocabulary type is academic words. Academic words are frequent words that occur outside the GSL and that are commonly used in academic texts across different disciplines. Academic words are also referred to as semi-technical vocabulary (Farrell, 1990) and sub-technical words (Yang, 1986) in the literature. The best-known academic wordlist is the Academic Word List (AWL) created by Coxhead (2000). Coxhead developed this wordlist by performing a corpus-based analysis of a compilation of academic texts. The AWL is comprised of 570 word families and covers around 8.5-10% of running words in a given text (Nation, 2008).

The third type of vocabulary is technical words. Technical words are words whose meanings are specially tied to one specific area, such as engineering, medicine, or business. These words occur frequently in particular

academic disciplines, but not in others. Technical words may be listed in the GSL or the AWL; however, their definitions when used in a particular discipline may be specific to that field. Technical words cover approximately 5%–30% of running words in a technical text (Chung & Nation, 2003; Nation, 2001, 2008) and they are essential to those who want to develop their knowledge in a particular field of ESP.

The last type of vocabulary is low-frequency words. Examples of such rare words include *abort*, *accusation*, and *acre*. Low-frequency words can be found in *Webster's 3rd International Dictionary*, or *The Shorter Oxford English Dictionary* and on such websites ashttp://users.tinyonline.co.uk/gswithenbank/unuwords.htm, and http://phrontistery.info/ihlstart.html. According to Nation (2008), rare words cover around 5% of running words in a given text.

For practical purposes, core vocabulary should be the main focus for language learners (McCarthy, 1990). Therefore, BE wordlists should contain words that learners need most. In ESP classrooms, academic vocabulary or those words that occur frequently in a specific field should be emphasized most (Dudley-Evans & St. John, 2011).

Method

Design of the KKU BE Corpus

Based on the literature review, four major factors need to be taken into consideration when compiling a corpus. These include representativeness, sampling, balance, and size (Biber, 1993; Biber, Conrad, & Reppen, 1998; McEnery, Xiao, & Tono, 2006; Sinclair, 1991). The four factors needed to be addressed when constructing the KKU-BE Corpus are detailed below.

Representativeness: In the present study, representativeness (i.e., a selection of texts that represents the possible sub-fields and text types in BE) is achieved by adopting Biber's (1993) guidelines on boundaries and hierarchical organizations. First, the boundaries of the target population (i.e., written texts published between 2001 and 2010) and the hierarchical organizations (i.e., major text categories) were defined. To achieve this, all areas of business (e.g., marketing, tourism, and economy) were listed by aggregating business majors from various universities, as well as sub-sections of books and newspapers. A list of text types teachers and students were likely to read was also created (e.g., books, academic articles). Two experts in business who taught at the tertiary level were then asked to prioritize and select the most relevant business areas as well as the most relevant text types to be included in the corpus. The final list contains eight sub-areas of Business English—marketing, finance, management, business law, tourism, economy, and accounting—and the following text types: textbooks, academic journals, websites, and newspapers.

Sampling: Another variable that must be considered when constructing a sound, well-balanced corpus is sampling; in other words, texts that are included in the corpus must be representatives of the breadth of texts as whole (Biber, 1993; McEnery, Xiao, & Tono, 2006). To achieve proper sampling, the population boundary must be clearly defined by outlining the sampling units and the sampling time frame (i.e., a list of possible sampling units). Once the boundary has been defined, one has a choice of two possible methods for sample selection: random sampling and stratified sampling. The present corpus was constructed utilizing a combination of both methods.

The target population was first divided into different strata, but the texts in each stratum were chosen randomly. In particular, the population for the present corpus is written texts in the business field from four sources: textbooks,

academic journals, websites, and newspapers. The sampling frame is published texts (i.e., texts printed in multiple copies for distribution and/or copyright registered) (Biber, 1993). Textbooks were then randomly selected from textbooks available in KhonKaen University Library. Experts in each subfield were asked to give a list of academic journals and websites that teachers and students use in the given areas. For the newspaper stratum, business news articles from the three English newspapers in Thailand (i.e., Nation, Bangkokpost, and Business Day) were included.

Balance: Balance refers to the coverage of a range of text categories representing the corpus (Atkin, Clear, & Ostler, 1992; Biber, 1993; McEnery, Xiao, & Tono, 2006). There are no specific rules governing which text categories and text types may be included in a corpus. A corpus is acceptably balanced simply if it can answer the intended research question (McEnery, Xiao, & Tono, 2006). Yet, Coxhead (2000) argued that one should have an equal number of running words in each text category. If this guideline is followed, then the words contained in the resulting BE corpus will not be skewed more to a particular genre.

The main aim of the current study is to identify academic words that occur frequently in BE fields, so as to make them available for teachers and students of BE , as well as others who work with BE. Balance is important to this goal. Thus, three major balancing criteria were used when constructing the KKU-BE corpus: domain, medium, and time. The domain of the corpus (i.e., the content type or subject fields), is the following eight major subfields of BE: marketing, finance, management, business law, tourism, economy, and accounting. The four text types are textbooks, research articles, websites, and news. The period of text production is ten years (2001-2010), in order to cover contemporary language behaviors that will be most useful to current learners.

In addition to these criteria, following Coxhead (2000), it was aimed to include an equal number of running words from each publication type (i.e., books, research articles, newspapers, and websites). Specifically, the goal was to include approximately 500,000 running words from each of the four text types for a total of about 2million running words in each academic subfield.

However, the number of running words in some text types was constrained by their relative availability online and in the KKU library (e.g., news about business law). As a result, these text types had fewer running words than the other text types. The total number of running words in the final KKU-BE corpus is 15,936,237.

Size: Corpus size is defined exclusively by the number of words in the corpus. There is no single rule of thumb concerning how many words should be included in a corpus; the size of the corpus should suit the research questions at hand. In general, however, the bigger the corpus, the better (Sinclair, 1991).With a large size of corpus, more information can be combined (Coxhead, 2000). In other words, a large corpus always contains more information for researchers to analyze.

Hunston (2002) states that the number of words in a corpus should be in the multi-millions. A corpus with less than one million words is considered small. The size of the KKU-BE corpus is approximately 16 million words. The previous studies discussed in the literature review had far smaller corpora, averaging between 40,000 and 2,000,000 running words (Alcantar, 2007; Chung, 2009; Coxhead, 2000; Para, 2004). Even the other best-known corpora, such as the Brown Corpus, contain1-million running words or fewer. The size of approximately 16-million running words is therefore considered sufficient for the purposes of this study.

A summary of the KKU-BE Corpus design appears in Table 1. Table 2 gives information concerning the four text types and the number of running words in each of the eight sub-fields. Each text type contributed about 1.6 to 2.3 million running words to the total constructed corpus, which is precisely 15,936,237 words.

Table 1

Summary of the KKU-BE Corpus Design

Issue in designing a corpus	Current study
Representativeness	- Included major text categories and text types representing the target population - Target population: Business English texts published between 2001-2010 - Text categories: eight sub-areas of Business English (marketing, finance, management, business law, tourism, economy, and accounting) - Four text types: textbooks, research articles, website articles, and newspapers
Sampling	- Sampling frame and units of the target population - The sampling frame/units of the target population: texts published between 2001 and 2010 - Techniques for sampling units - Strata: stratified sampling selection: eight BE sub-areas and four text types (textbooks, research articles, website articles, and newspapers) - Selection of texts: random sampling selection

Balanced	- Covering of a wide range of text categories and text types
	
	    - textbooks
	
	    - Journals
	
	    - Articles
	
	    - Newspapers
	
	- Equal proportion from each text category

Size	- 15,936,237 running words
	
	    - approximately 2 million words from each text category
	
	    - approximately 500,000 words from each text type

Table 2

Text Categories and Text Types of KKU BE Corpus

Text Category	Text Type	No. of Texts	No. of Running Words
Accounting	Books	20	535,944
	Journals	117	446,737
	Websites	483	510,779
	Newspapers	290	426,694
	Total	910	1,920,154
Advertising	Books	20	450,490
	Journals	115	433,083
	Websites	410	451,009
	Newspapers	550	310,700
	Total	600	1,645,282
Business Law	Books	19	605,727
	Journals	80	626,353
	Websites	500	638,866
	Newspapers	300	561,100
	Total	899	2,432,046
Economic			
	Books	20	506,394
	Journals	80	510,268
	Websites	372	491,842
	Newspapers	1,498	554,471
	Total	1,970	2,062,975

Table 2

Text Categories and Text Types of KKU BE Corpus

Text Category	Text Type	No. of Texts	No. of Running Words
Finance	Books	21	575,613
	Journals	79	548,265
	Websites	532	433,881
	Newspapers	527	447,205
	Total	1,159	2,004,964
Management	Books	21	472,923
	Journals	99	544,196
	Websites	533	426,232
	Newspapers	407	412,884
	Total	1,060	1,856,235
Tourism	Books	19	427,712
	Journals	95	590,535
	Websites	521	569,082
	Newspapers	801	422,500
	Total	1,436	2,009,829
Marketing	Books	20	493,333
	Journals	70	541,453
	Websites	521	525703
	Newspapers	1,038	444,263
	Total	1,649	2,004,752
Total			15,936,237

Research Instruments

Two research tools were used in this study: WordSmith and Range. WordSmith Version 6, a concordance program created by Mike Smith (Oxford University), was used to generate word lists of high-frequency words and the keyness indexes of those words. The Range program, developed by Paul Nation, was used to analyze how these words were loaded across the different BE disciplines.

Data Collection

Textbooks from each of the eight sub-fields were randomly selected from the KKU library, scanned, and saved as text files. To equalize the number of running words from each text type, pages of the textbooks were also randomly selected to meet the set criteria of 500,000 words. Texts of research articles, website articles, and newspaper were obtained electronically through the library's online databases. All texts were scanned and save as text files, which were then categorized according to subfield, text type, and publication year. After scanning, the text files were also revised to clean up paper titles, graphs, acknowledgements, references, typographical errors, and other potentially incorrect words or symbols that were corrupted during the scanning process.

Development of Business English Academic Word List

The KKU-BE corpus as then analyzed to create a Business English Academic Word List (BEAWL) by following Coxhead (2000) and Wang, Liang, and Ge's (2008) three primary principles: specialization, frequency, and range. In these studies, range is considered a more important criterion than the others for constructing usable word lists. As a result, only words that occur in half of the subject areas or more (four or more) were included as members of a word family in the final list (Coxhead, 2000; Wang, Liang, & Ge, 2008). The following are the criteria used to identify academic words in this present study.

1. Specialized occurrence: The words included in the BEAWL must not be in West's GLS (1953), or be one of the first 2000 most frequent words in the KKU-BE corpus as a whole.

2. Frequency: Words included as members of a word family must occur at least 100 times in the KKU-BE corpus, and at least 10 times in each discipline.

3. Range: Words included as member of a word family must occur at least 4 or more of the 8 subject areas.

The words that met all three criteria were then ordered according to their range (i.e., four to eight disciplines covered) and frequency. Words belonging to the same word family were grouped and counted as one word, using the base word as the head of the family. According to Bauer and Nation (1993), "a word family consists of a base word and all its derived and inflected forms that can be understood by a learner without having to learn each form separately" (p.253). For example, *approach*, *approaches*, *approached*, *approaching*, and *approachable* are counted as one word and the head, which is the base form, is *approach*. Similar to Coxhead (2000), both American and British spellings were kept in the list (e.g., *individualize* and *individualise*); both forms are useful for BE learners to know. Proper nouns (e.g., *Germany*, *Japan*, *Korea*, and *Facebook*), abbreviations (e.g., *IMF*, *CEO*, and *GDP*), and Latin forms (e.g., *etc*., *e.g*., and *i.e*.) were excluded.

Finally, two experienced university English teachers were asked to categorize potential words into word families. When there were any uncertainties, dictionaries and experts were consulted. A complete agreement rate between the two teachers was eventually achieved. Unlike previous studies, which have mainly used frequency as the sole criterion for sub-list development, the sub-lists of the BEAWL were created first by range and then by frequency. t is hoped this method will make the resulting list maximally useful to BE learners and teachers.

Results

The complete KKU-BE contains 15,936,237 running words (or 212,576 types). However, after the words in the GSL were eliminated as described above, there were 4,009,825 running words (or 205,083 types) that were potential to be included in the BEAWL. Of those 4,009,825 words (205,083 types), 1864 words satisfied both the frequency and range requirements; that is, 1864 words occurred over 100 times in the KKU-BE corpus and in four or more disciplines of Business English. These words were categorized into 415 word families with the base form as the headword. The total occurrences of academic words and their family members in the BEAWL are 2,107,550, which is12.66% of the total running words in the KKU-BE corpus. In the BEAWL, the most frequently used headword in all eight disciplines is *lien* (frequency = 44679), while the least frequently used headword in all eight disciplines is *implicate* (frequency = 124).

As discussed, the current wordlist differs from Coxhead (2000) and Wang, Liang, and Ge (2008) in that those lists were listed based on word frequency alone, whereas this list sorted first by range (four to eight) and second by frequency. As such, the BEAWL is comprised of seven sub-lists: Sub-lists 1 to 3 cover all eight subject areas (166 headwords, 100 percent of range coverage); Sub-list 4 covers seven subject areas (43 headwords, 87.5 percent of range coverage); sub-list 5 covers six subject areas (65 headwords, 75 percent of coverage); sub-list 6 covers five subject areas (63 headwords, 62 percent of range coverage); and sub-list 7 covers four subject areas (78 headwords, 50 percent of range coverage) (see Table 3). Of the total 415 word families in the list, 166 headwords occurred in all subject areas for a total of 1,215,119 occurrences (7.62% coverage); word families that covered four or more subject areas occurred 802,432 times (5.04% coverage). The BE academic words are shown in Table 4.

Table 3

Coverage of the BEAWL in the KKU-BE corpus

Sub-list	Number of word families	Family members	Range (i.e., subjects areas in which they occurred)		Frequency of headwords and their word families	
			Occurrence	Coverage (%)	Occurrence	Coverage (%)
1	60	259	8	100	727,210	4.56
2	60	233	8	100	298,450	1.87
3	46	171	8	100	189,459	1.19
4	43	164	7	87.50	235,612	1.48
5	65	239	6	75.00	223,364	1.40
6	63	225	5	62.50	171,726	1.08
7	78	281	4	50.00	171,729	1.08
Total	415	1572	N/A	N/A	2,017,550	12.66

Table 4

Description of Academic Words and Sub-lists in the KKU-BE Corpus

Sub-list	Number of Words	Headwords
Sub-list 1	60	*lien ratio sum research data income cash process vice norm edit journal percent cant period credit policy legal corporate chapter site economy finance project significant goods media opt strategy approach impact text major section structure role issue available positive individual source internet potential global specific web focus method technology revenue job mission theory similar environment evidence negative range factor create*
Sub-list 2	60	*design area team core target response overall internal software community via access client prior function consistent relevant context affect computer alternative appropriate empirical network benefit require concept previous identify external survey fined option ensure item executive initial volume hypothesis hence dividend category commission framework final estimate perspective goal complex series achieve indicate primary conduct objective despite contrast unique resource link*

Sub-list 3	46	*summary majority instance versus status maintain whereas liable generate trend author challenge occur respond labour theoretical establish furthermore column conclusion construct procedure seek outcome feature define vary comprehensive analyse select involve professor element locate purchase statistic incorporate sued consist participant tradition invest consume distribute species implicate*
Sub-list 4	43	*port mental equity cult sage rely budget federal fund annual index domestic crisis region normal capacity retail senior obtain medium enterprise electronic physical administration aware evaluate criteria schedule component promote principle assess usage contribute expert enhance institution aspect communicate academy interact integrate regulate*
Sub-list 5	65	*brand contract fee inventory asset culture portion plus entity output graph equipment error exhibit merge assume lease authority partner discount equation cycle governance forecast minimum style transfer shift energy sufficient contact input career assumption incentive medical fundamental maximum expansion significance bias huge emphasis dynamic magazine appendix implement investigate considerable facility assist automobile consequence perceive sequence emerge publish paragraph infra consequent compensate innovate correlate transact regress*
Sub-list 6	63	*format stake bond imply currency file image panel venture consumption cope code micro mobile principal scope gross ministry channel decline prime database temporary acquisition premium constant conflict infrastructure expenditure oriented subsequent mutual institute proportion accurate valid expand expertise boost score generation turnover qualitative proposition engage phase interview household commit obvious monitor convention publication participate release recover entrepreneur comment ultimate prospect assign approximate circumstance*
Sub-list 7	78	*spa vision super portfolio lever mart campaign monetary estate segment television coefficient margin exit compliance export utility administrative deposit document attitude subsidiary reform chart mortgage appeal traffic equivalent linear overseas conversion capture enable flexible topic personnel analyst perception thereby eventually mechanism vehicle questionnaire contemporary personality transport scheme notion scenario decade ongoing launch demonstrate acquire dimension attribute predict distinct diverse reveal differential logic digit motive derive sustain liberal restrict coup practitioner correspond announce initiate expose guideline foundation constrain inflate*
Total	415	

In addition, a further analysis was conducted to examine the coverage of academic words in each text type. Academic words account for 13.79% of the total word count in journal articles and about 11.34% of the total word count in textbooks. Further, the coverage of academic words in newspapers and websites was quite high: 8.81% and 8.93%, respectively (see Table 5).

Table 5

Proportion of the BEAWL in the KKU-BE Corpus by Text Types

Text types	Token	Coverage (Percent)
Textbooks	237,976	11.34
Journal articles	278,067	13.79
Newspapers	163246	8.81
Websites	186860	8.93

A passage from an economy text, Temple and Williams (2002), was randomly selected to exemplify how academic words of the BEAWL are used in running text. As was mentioned above, BEAWL comprises 12.66% of the total words in the entire constructed KKU-BE Corpus. The academic words listed in the BEAWL are underlined in the text excerpt. The total number of the words in the passage is 365 words, and the number of BEAWL words within it is 47. The text coverage of the BEAWL in the passage, then, is 12.88%; the passage is indeed exemplary.

Proprietary measurement techniques can of course perform as proprietary assets, providing a lead advantage for the firms that develop them. Competitors have to spend time and money replicating the techniques or may have to purchase them under licence. Public provision potentially allows techniques to diffuse quickly by removing the need for replication or by providing the technologies cheaply. Without public provision, monopolistic advantage would lead to general under provision of technologies and may hold back growth.

In a similar vein, Swann (1999) argues that there are likely to be considerable positive externalities associated with measurement activity, which can be maximised if measurement is made widely available. This is a generally under researched area but NMSPU (1999) provides some tentative estimates of the lower bounds of some externalities identified from the set of case studies used there.

Public support for measurement infra-technologies also allows the regulation of other forms of market failure such as the negative externalities of pollutants. Almost all countries have a legal metrology infrastructure to enforce weights and measures standards and for wider consumer protection and health and safety. A national measurement infrastructure therefore reduces risk for consumers and industrial users by establishing common standards and compatibilities, by requiring minimum quality levels, by reducing variety and by defining common information sets. In this regard they are likely to generate advantages similar to other forms of industrial standards (see Swann et al., 1996, for a discussion of the latter's impact on UK trade flows).

Of course, as discussed earlier, the measurement activity that takes place within the publicly funded sector accounts for only part, probably a

small part, of the overall use of measurement techniques in the UK and other industrialized economies. For this to have a significant impact on growth there must be significant spillovers, externalities and other forms of social increasing returns as predicted by the so-called 'endogenous growth' literature surveyed by inter alia Barro and Sala-i-Martin (1998), Aghion and Howitt (1999).

There is of course a large academic literature on the rationale for public funding of basic research of which Mansfield (1991), Pavitt (1991), Kealy (1996), Swann et al. (1996), Tassey (1997) and Georghiou and Roessner (2000) are a small selection.

Of the 415 word families, 306 word families (73.74%) in the BEAWL are related to Coxhead's AWL words. 93 word families (26.26%) fall outside the AWL (see Table 6 for a list of those headwords that occur in the BEAWL but are not listed in the AWL). This means that the AWL does not account for over a quarter of the most frequent words that occur in over half of BE disciplines' texts. Clearly, the AWL is not enough for BE learners. This finding is in agreement with previous studies that suggest discipline-specific academic word lists are more beneficial to students than general wordlists (Chen & Ge, 2007; Hyland & Tse, 2007; Martinez, Beck, & Panza, 2009; Wang, Liang, & Ge, 2008).

Table 6

Ninety Three Words listed in the BEAWL but not in the AWL by Sub-lists

Sub list	Headword
Sub-list 1	cant cash goods internet lien mission opt vice web
Sub-list 2	client dividend executive fined software
Sub-list 3	column liable professor species sued versus
Sub-list 4	budget crisis cult electronic enterprise equity port retail sage senior usage
Sub-list 5	asset automobile brand career correlate discount forecast governance graph huge infra inventory lease magazine merge regress transact
Sub-list 6	boost cope database engage entrepreneur expenditure gross household interview micro mobile premium proposition score stake turnover venture
Sub-list 7	announce appeal campaign capture coefficient compliance coup deposit differential digit exit inflate launch lever linear mart monetary mortgage personality personnel portfolio questionnaire reform segment spa super television traffic

Discussion and Conclusion

The present study presents a list of academic words (the BEAWL) extracted from a BE corpus (the KKU-BE) containing 16 million words. The BEAWL contains 415 headwords and 1572 family members. The BEAWL was then divided into seven sub-lists arranged by subject-area coverage and then by frequency. The first three sub-lists contain words that occur in all eight subject areas, and the other four sub-lists contain words that appear in four to seven subjects areas. The BEAWL words account for 12.66% of all words in the BE corpus. Journal articles and textbooks use more of the BEAWL words than other text types: 13.79% and 11.34%. Newspapers and websites included more like 9% BEAWL words.

The BEAWL and its sub-lists will be useful to BE learners, who are reported to have the trouble acquiring academic-type words in their chosen disciplines. These lists can be used to set learning goals in Business English

programs and specialized areas of Business English. The fact that these academic words occur less frequently in non-academic text types does not indicate that they are not useful to students. Similar to Coxhead (2000), this study found that non-academic text types (i.e., newspapers and websites) still contained some academic words (here, 9%, as compared to Coxhead's 8.5%).

The BEAWL and its sub-lists can be applied in the teaching of Business English and English for Specific Purposes because these words are necessary for second language learners to achieve fluency and to achieve competence in their target disciplines (Dudley-Evans & St. John, 2011; Durrant, 2016; Hutchinson & Waters, 1996). The findings of this study also have implications for Business English/English for Specific Purposes pedagogy and future research in these fields.

Recommendations for pedagogy

The BEAWL may be used to set learning goals in BE classrooms and may be one major focus in the development of BE materials, dictionaries, classroom activities, textbooks, and tests. The list is divided into seven sub-lists, which are arranged based on the number of subject areas in which they occurred in the BE. As such, the early sub-lists should receive the most attention; they are the words BE students are most likely to encounter across different BE texts. To achieve optimal results, the words in the BEAWL should be taught and learned through both explicit and implicit methods. Direct vocabulary instruction has been found to yield positive effects on vocabulary development (such as Coady, Mgoto, Hubbard, Graney, & Mokhtari, 1993; Lee, 2003; Lee & Liou, 2003; Yamamoto, 2014). Yet, learners need to be given opportunities to learn words implicitly, in natural speech and everyday readings, in order for them to learn the nuances of their meanings, derivative forms, etc. in context. In other words, it is crucial to enhance both learners' receptive and productive knowledge.

Recommendations for future research

The KKU-BE Corpus is compiled from written texts, not from transcripts of spoken English, due to access limitations and financial constraints. The BEAWL is therefore most valuable for teaching the reading and writing of BE, not conversation or listening comprehension. Future studies should compile a corpus focusing on spoken language. A BEAWL derived from a spoken corpus would help us to better understand what students of BE need to succeed in verbal business interactions, and how similar the required vocabulary may be to those words used in written texts.

References

Alcantar, C. (2007) Construction of a special corpus to extract the most frequent words in a particular field. **Essex Graduate Papers in Language and Linguistics**.*9*, 1-17.

Atkin, S., Clear, J., & Ostler, N. (1992). Corpus design criteria. In McEnery, T., Xiao, R., & Tono, X. (eds.), **Corpus-based language studies** (pp.128-130). New York: Routledge.

Bauer, L., & Nation, .I. S. P. (1993).Word families. **International Journal of Lexicography**, *6*, 253-279.

Biber, D. (1993). Representativeness in corpus design. **Literary and Linguistic Computing, 8**, 243-257.

Biber, D., Conrad, S., & Reppen, R. (1998) **Corpus Linguistics: Investigating language structure and use (Cambridge Approaches to Linguistics)**. Cambridge: Cambridge University Press.

Biber, D., & Reppen, R. (2002).What does frequency have to do with grammar teaching? **Studies in Second Language Acquisition**, *24*, 199–208.

Chen, P., Hu, K., & Ho, J. (2009). A study of academic vocabulary used in the abstracts of business and management journals. **Taiwan International-al ESP Journal**, *1*, 51-76.

Chen, Q., & Ge, G. (2007).A corpus-based lexical study on frequency and distribution of Coxhead's AWL word families in medical research articles. **English for Specific Purposes**, *26*, 502-514

Chung, T. M. (2009). The newspaper word list: A specialised vocabulary for reading newspapers. **JALT Journal**, *31*, 159-182.

Chung, T., & Nation, P. (2003).Technical vocabulary in specialized texts. **Reading in a Foreign Language**, *5*, 103-116.

Coady, J. (1997). L2 vocabulary acquisition through extensive reading. In J. Coady & T. Huckin (Eds.), **Second language vocabulary acquisition** (pp. 225-237). Cambridge, UK: Cambridge University Press.

Coady, J., Mgoto, J., Hubbard, P., Graney, J., & Mokhtari, K. (1993).High frequency vocabulary and reading proficiency in ESL readers. In T. Huckin, M. Haynes, & J. Coady (Eds.), **Second language reading and vocabulary learning** (pp.217-22). Norwood, NJ: Ablex.

Cobb, T., & Horst, M. (2004). Is there room for an academic word list in French? In P.

Bogaards & B. Laufer (Eds.) **Vocabulary in a second language** (pp. 15–38). Amsterdam and Philadelphia: John Benjamins Publishing Company

Coxhead, A. (2000). A new academic word list. **TESOL Quarterly**, *34*, 213-238.

Coxhead, A. (2016). Reflecting on Coxhead (2000), "A New Academic Word List." **TESOL Quarterly**, *50*, 181-184.

Coxhead, A., & Nation, P. (2001).The specialised vocabulary of English for academic purposes. In J. Flowerdew & M. Peacock (Eds.), **Research perspectives on English for academic purposes** (pp. 252–267). Cambridge University Press.

Donley, K. & Reppen, R. (2001).Using corpus tools to highlight academic vocabulary in sustained content language teaching. **TESOL Journal,** 10, 7-12.

Dudley-Evans, T. & St. John, M. (1996). **Report on Business English: A review of research and published teaching materials**. TOEIC Research Report No.2. Princeton, NJ: The Chauncey Group International.

Dudley-Evans, T. & St. John, M. (2011). **Developments in English for specific purposes: A multi-disciplinary approach**. Cambridge: Cambridge University Press.

Durrant, P. (2016).To what extent is the Academic Vocabulary List relevant to university student writing? **English for Specific Purposes,43**, 49-61.

Farrell, P. (1990). Vocabulary in ESP: A lexical analysis of the English for electronics and a study of semi-technical vocabulary. **CLCS Occasional Paper No. 25 Trinity College**.

Ghapanchi, Z., Eskandari, Z. & Tabasi, E. (2012). The effect of text-based direct vocabulary instruction on vocabulary acquisition. **World Journal of English Language.2**, 74-80.

Grabe, W. (2009). **Reading in a second language: Moving from theory to practice**. New York: Cambridge University Press

Grabe, W.& Stoller, F. (2002).Comparing L1 and L2 reading. In **Teaching and researching reading** (pp. 30-41). London: Longman/Pearson Education.

Hunston, S. (2002). **Corpora in Applied Linguistics**. Cambridge, England: Cambridge University Press.

Hutchinson T.,& Waters A. (1996). **English for Specific Purposes**. Cambridge: Cambridge University Press.

Hyland, K., & Tse, P. (2007). Is there an "academic vocabulary"? **TESOL Quarterly, 41**, 235-253.

Koda, K. (2005). **Insights into second language reading**. NY: Cambridge University Press.

Laufer, B. (1997). The lexical plight in second language reading: Words you don't know, words you think you know, and words you can't guess. In J. Coady & T. Huckin (Eds.), **Second language vocabulary acquisition: A rationale for pedagogy** (pp. 20–34). Cambridge: Cambridge University Press.

Lee, S. H. (2003). ESL learners' vocabulary use in writing and the effects of explicit vocabulary instruction. **System, 31**, 537-561

Lee, C.Y., & Liou, H.C. (2003). A study of using web concordancing for English vocabulary learning in a Taiwan high school contexts. **English Teaching & Learning, 27**, 26-56.

Lie, L. & Liu, D. (2016). A new medical academic word list: A corpus-based study with enhanced methodology. **Journal of English for Academic Purposes, 22**, 42-53.

Martínez, I. A., Beck, S. C., & Panza, C. B. (2009). Academic vocabulary in agriculture research articles: A corpus-based study. **English for Specific Purposes, 28**, 183-198.

McCarthy, M. (1990). **Vocabulary**. East Kilbride: Oxford University Press.

McEnery, T., Xiao, R., & Tono, Y. (2006). **Corpus-based language studies: An advanced resource book**. New York: Routledge.

Meara, P. (1996). The dimensions of lexical competence. In G.Brown, K. Malmkjaer, & J.

Williams (Eds.), **Performance and competence in Second Language Acquisition** (pp. 35-53). Cambridge: Cambridge University Press

Nation, I. S. P. (1990). **Teaching and learning vocabulary**. New York, NY: Newbury House.

Nation, I. S. P. (2001). **Learning vocabulary in another language**. Cambridge, England: Cambridge University Press.

Nation, I.S.P. (2008). **Teaching vocabulary: Strategies and techniques**. Boston: Heinle Cengage Learning

Nation, I. S. P. (2013). **Learning vocabulary in another language**. (2nd ed.) Cambridge, England: Cambridge University Press.

Nation, I.S.P. (2016). **Making and using word lists for language learning and testing**. Amsterdam: John Benjamins.

Nation, P., & Hwang, K. (1995). Where would general service vocabulary stop and special purpose vocabulary begin? **System**, *23*, 35-41.

Nation, I. S. P., & Waring, R. (1997). Vocabulary size, text coverage, and word lists. In N. Schmitt & M. McCarthy (Eds.), **Vocabulary: Description, acquisition and pedagogy** (pp. 6–19). Cambridge: Cambridge University Press

Nelson, M. (2000) **A corpus-based study of Business English and Business English teaching materials**. PhD thesis, the University of Manchester, Manchester. Accessed on 15 July 2012 at the URL http://users.utu.fi/micnel/thesis.html.

Para, C. (2004). **A corpus analysis of high-frequency words in civil engineering research articles: Sub-disciplinary differences between structure and transportation**. Unpublished Doctoral Dissertation. Graduate School, Mahidol University.

Paribakht, T. S. & Wesche, M. (1997). Vocabulary enhancement activities and reading for meaning in second language vocabulary development. In J. Coad y& T. Huckin (Eds.), **Second language vocabulary acquisition: A rationale for pedagogy** (pp.174-200). New York: Cambridge University Press.

Read, J. (2000). **Assessing vocabulary**. Cambridge: Cambridge University Press.

Robinson, P. (1991). **ESP Today: A practitioner's guide**. Hemel Hempstead: Prentice Hall International.

Shaw, P. (1991). Science research students' composing processes. **English for Specific Purposes**, 10, 189–206.

Schmitt, N. (2000). **Vocabulary in language teaching**. Cambridge: Cambridge University Press.

Sinclair, J. (1991). **Corpus, concordance, collocation: Describing English language**. Oxford University Press, Oxford.

Temple P, Williams G. (2002). Infra-technology and economic performance: Evidence from the United Kingdom measurement infrastructure. **ELSEVIER Science BV Information Economics and Policy**.14, 435-452.

Thurstun, J., & Candlin, C. (1998). Concordancing and the teaching of the vocabulary of Academic English. **English for Specific Purposes, 17**, 267-280.

Vongpumivitch, V. Huang, J & Chang, Y. (2009). Frequency analysis of the words in the Academic Word List (AWL) and n on-AWL content words in applied linguistics research papers. **English for Specific Purposes, 28**, 33-41.

Wang, J., Liang, S.,&Ge, G. (2008). Establishment of a medical word list. **English for Specific Purposes, 27**, 442-458.

West, M. (1953) **A general service list of English words**. London: Longman.

Yamamoto, Y. (2014). Mutidimensional vocabulary acquisition through deliberate vocabulary list learning. **System, 42**, 232-243.

Yang, H. (1986). A new technique for identifying scientific/technical terms and describing science texts. **Literary and Linguistic Computing, 1**, 93-103.

Yang, M. (2015). A nursing academic word list. **English for Specific Purposes, 37**, 27-38.