Creating an Academic Business English List: a Corpus Based Study

Stella Carolina Stella

A thesis in the Department of Education

Presented in Partial Fulfillment of the Requirements

for the Degree of Masters of Arts (Applied Linguistics) at

Concordia University

Montréal, Québec, Canada

September 2, 2015

CONCORDIA UNIVERSITY
School of Graduate Studies

This is to certify that the thesis prepared

By: Stella Carolina Stella

Entitled: Creating an Academic Business English List: a Corpus Based Study

and submitted in partial fulfillment of the requirements for the degree of

Master of Arts (Applied Linguistics)

Complies with the regulations of the University and meets the accepted standards with respect to originality and quality.

Signed by the final Examining Committee:

_____ (Chair)

Teresa Hernandez-Gonzalez

_____ (Examiner)

Walcir Cardoso

_____ (Examiner)

Pavel Trofimovich

_____ (Supervisor)

Marlise Horst

_____ (Supervisor)

Joanna White

Approved by: _____

Graduate Program Director

_____

Dean of Faculty Date: _____ 2015

Abstract

Past studies have shown that second language (L2) learners need to know between 95% to 98% of the vocabulary in a text in order to understand its content (e.g. Hirsh & Nation, 1992; Laufer, 1989; Laufer & Ravenhorst-Kalovski, 2010; Nation, 2006; Schmitt et al., 2011; Stæhr, 2008). Other studies tried to identify exactly what words these students need to know in order to achieve this level of comprehension, and researchers created frequency lists that could be used to teach vocabulary (e.g., Browne, Culligan, & Phillips, 2013; Nation, 2012; West 1953). However, as General English (GE) frequency lists are not very efficient for teaching specialized vocabulary, other lists were created in order to solve this issue (Coxhead, 2000; Hsu, 2011a; Konstantakis, 2007). This thesis study investigated the vocabulary needed by undergraduate students in business in order to understand their textbooks at this 95-98% level of text coverage. A corpus based on undergraduate business core courses textbooks was compiled by the researcher in order to extract a list of the most used vocabulary excluding the BNC/COCA 3K and proper nouns (which are considered to be already known by these learners). The Academic Business English List (ABEL) was created, and it consists of 840 word families which cover 2.86% of the ABEL corpus. With knowledge of these 840 word families, the BNC/COCA 3K and proper nouns, text coverage reached 96.94%. The ABEL list can be used by materials creators and business English teachers in order to prepare students for their undergraduate business programs.

Keywords: vocabulary, business English, reading and comprehension, frequency list, English for Specific Purposes (ESP), English for Occupational Purposes (EOP), English for Academic Purposes (EAP), corpus linguistics.

Acknowledgements

I would like to thank my supervisors Marlise Horst and Joanna White for their patience, time commitment, feedback and words of encouragement throughout the creation of this thesis. I would also like to thank Tom Cobb for his invaluable help with the technical part of this research, and for all the work he put into upgrading Lextutor to accommodate my corpus. I would also like to thank Walcir Cardoso and Pavel Trofimovich for being part of this thesis committee and for investing their time to read and provide me with feedback. A special thanks goes to my family and friends who were very patient and supportive of me engaging in graduate studies; more specifically my father, who dreamed of going to graduate school but who never saw that dream materialize, and who always encouraged us to pursue our studies and become the best person we could be.

Table of Contents

List of Tables

Chapter One - Introduction

Personal Motivation

My interest in Business English (BE) arose from a professional struggle. When I was a novice teacher working in a private language school, I was asked (along with my other novice colleagues) to teach a BE course which was part of the school's curriculum. However, none of us had previous training or experience teaching it, and the language school was not planning on providing us with any training either as they judged it to be unnecessary. As novice teachers, we did not want to insist on having specific training because we did not want to seem unprepared in the eyes of our new employers. This did not stop us from being terrified of teaching BE, and this would be a major discussion topic in the teachers' room during our breaks.

As I started teaching BE, I noticed that the major challenge for me was to become familiar with the technical terms and their meanings. I also noticed that the grammar of BE or the pedagogical needs (such as basic teaching techniques and knowledge of the Communicative Approach) were not an issue. I then decided to pay more attention to my colleagues' comments, and I realized that the challenge was the same for them. I believe that most of us felt unprepared because we did not have the vocabulary needed to converse in the language of BE, not because we felt we were not good teachers.

I consequently related this feeling of lack of vocabulary to my students' feelings. In other words, I believe that BE students are mostly comfortable with the grammar of the English

language, but they look for a BE course because they want to be able to converse or write in BE, and above all, they want to have the appropriate vocabulary needed to do so.

The simplest way to solve this issue would be to become familiar with the BE terminology. BE dictionaries already exist (Dictionary, 2014; Dignen, 2000; Parkinson, 2005; just to name a few) and are excellent sources of definitions. However, BE dictionaries contain too many entries to be learned, and they usually do not tell readers which words are more frequent in the language. Therefore, I propose in this thesis to create a list with the most frequent words in BE so that students and teachers can spend their time working on the most cost-effective vocabulary.

In Chapter Two of this thesis, the literature review is presented, and the reader will find the definition of the main terms used throughout the text, as well as a detailed presentation of the research in reading and comprehension related to vocabulary size. The research questions are also presented here. Chapter Three contains the description of the materials and design used to create the Academic Business English List (ABEL). Chapter Four presents the results of this study and Chapter Five presents the discussion.

In the next chapter, I begin by explaining why BE is an important topic for English as a Second Language (ESL), English as a Foreign Language (EFL) and English for Academic Purposes (EAP) fields, and I then define the main terms surrounding this area of study.

Chapter two - Literature Review

International exchange programs and study abroad programs have become very popular in the past few decades. In 1975 approximately 800,000 students travelled abroad to seek education at the tertiary level (OECD, 2011) and by 2012, this number jumped to four million (Unesco, 2014b). Out of the top ten destination countries (United States, United Kingdom, France, Australia, Germany, Russian Federation, Japan, Canada, China and Italy), four are English-speaking nations, and together they represent 38% of this market. That is roughly 1,520,000 students in 2012 alone. The same report also shows that of the top ten countries from which the international students originate (China, India, Republic of Korea, Germany, Saudi Arabia, France, United States, Malaysia, Vietnam and Iran), in only two do people speak English as their first (United States) or second language (India). This means that most of these international students are pursuing their studies in a foreign language (FL) or in a second language (L2). In a different UNESCO report (2014a), researchers reveal that international students have a preference for programs in the Social Sciences, Business and Law (p. 355). If we average figures for the top four Anglophone countries mentioned above (United States, United Kingdom, Australia and Canada), we see that an average of 43.5% of foreign students are coming to their institutions for those specific programs. Unfortunately, the report does not tease apart these three main fields of study (Social Sciences, Business and Law). Therefore, it becomes difficult to know exactly the percentage of international students who enroll in Business programs specifically. Nevertheless, the numbers presented in this paragraph are substantial and show the popularity of Business courses among international students. For this reason I have chosen to work with this particular student population.

These international students face multiple challenges when adapting to their new universities: language issues, adapting to new cultural norms, cultural misunderstandings, financial problems, lack of social support, discrimination, among others (Sherry, Thomas, & Chui, 2010). Some researchers consider the lack of adequate language proficiency to be the major challenge for international students; more specifically, some researchers have found that lack of vocabulary and writing skills play an important role in impeding the academic success of international students (see Andrade, 2006, for a literature review).

Parry (1997) illustrated this struggle well by presenting a case study of two L2 learners who were taking an introductory Anthropology course. In her research she investigated the learning strategies the two participants used to cope with the amount of vocabulary necessary to read the materials and succeed in the course. The results showed a not so positive reality for L2 learners at the university level, as both students struggled with hundreds of words as they attempted to read textbooks designed for native speakers of English and finished the course with relatively low marks. In a similar study, Raymond and Parks (2002) investigated how international students from China transitioned from reading and writing assignments in their EAP courses to their Master of Business Administration (MBA) courses. The results showed that the major problem students had was coping with the amount of reading they had to do for their MBA courses, which was significantly larger and more complex than the reading they had to do for their EAP courses. The second major problem was lack of vocabulary needed to understand MBA texts. Students with previous knowledge of the subject matter had less difficulty learning the new terminology in English; however, students who lacked subject knowledge in business struggled to learn the new concepts and terminology at the same time. Zhang (2013) also touched on vocabulary learning in BE contexts in his genre study. In this paper he had business

practitioners evaluate the texts of five advanced BE students, and he classified their feedback in four dimensions of genre knowledge: formal, process, rhetorical and subject matter. The formal dimension is the one that involves lexical choices, and the results showed that this dimension received the greatest number of comments from the Business practitioners. He also mentioned in the discussion that the students "…were particularly weak regarding the nuances or connotations of words and expressions…" (p. 153). For the reasons presented in this section, I have chosen to narrow the scope of this study to the investigation of the vocabulary needed to cope with reading in Business undergraduate programs.

Before turning to an examination of the role of vocabulary in reading comprehension, I will introduce some key concepts that will be used in this study.

Definitions

There are many ways of defining the term *word* in linguistics, but for the purpose of this study, a word is "a string of letters bounded by spaces, or rather by some combination of space and punctuation mark" (Halliday, Matthiessen, & Matthiessen, 2014, p. 48). For example, *brother* is a word consisting of seven letters and the word *brother-in-law* is a different word consisting of twelve letters and two dashes. This definition may seem simplistic at first; however, when we consider that corpus linguistics research involves using computerized search techniques, it becomes clear why this definition is relevant. Software is not yet capable of analyzing words as humans do; therefore, a simple machine-readable definition is necessary to understand how the researchers and the software will deal with the words in a corpus.

Defining a word is important in a corpus study because it can determine how to count them, and there are many ways to count words. The simplest way is to count every word which appears in a text, and in that case we refer to them as *tokens* or *running words*. For example, the sentence *My dog is the best dog in the world!* has nine tokens. Another way to count words is by counting only the *types*. That is to say, we count only once a form that appears multiple times in a text. In this case the sentence *My dog is the best dog in the world!* has nine tokens, but only seven types because the words *dog* and *the* are repeated. A third way to count words is to count as one item the words that are closely related; that is, instead of counting *dog* and *dogs* as different words we count them as one word. When words are counted this way we call them *lemmas,* and this way of counting consists of identifying the headwords and including inflected or reduced forms as part of the same lemma. A similar but more complete way of counting words is to classify them as belonging to the same *word family*. Nation (2013) defines a word family as consisting "of a headword, its inflected forms and its closely related derived forms" (p. 11). For instance, the word *"happy"* is the headword of the word family consisting of *unhappy, (un)happiness, (un)happily, happier, (un)happiest*. The concept of word family is crucial for creating a frequency list because usually an entire word family is counted as one item in the list and represented by its headword. The reason researchers who are interested in pedagogy count word families is because it is believed that once a learner learns the headword, the learning burden of other forms in the same word family is very low.

Once the decision regarding how to count the words is made, a corpus is needed to create a word list. "A corpus is a collection of texts, written or spoken, which is stored on a computer" (O'Keeffe, McCarthy & Carter, 2007, p.1). To create a written corpus, one needs to collect electronic versions of texts in the desired field of study. An electronic corpus is readable by

software, which makes it possible to determine which lemmas or word families occur more frequently. The creation of a spoken corpus is much more complex and time consuming. One needs to first record oral production, transcribe the audio files (the most time-consuming task), and then code the transcription according to research needs. Some examples of corpora are the British National Corpus (BNC), the Brown Corpus and the Cambridge International Corpus (CIC) (see Appendix 1 in O'Keeffe et. Al, 2007, for an overview). A corpus can also have a technical or specialized feature, and the more the corpus is specialized, the better job it will do of providing the researcher with the words that are specific to that field. Specialized corpora are very powerful when it comes to investigating special uses of language (Nelson, 2010). The Cambridge and Nottingham Business English Corpus (CANBEC), the Wolverhampton Business English Corpus and the Business English Corpus (BEC) are three examples of corpora specialized in BE.

The size of a corpus can vary substantially, and O'Keeffe et. al (2007) say that for a written corpus, anything below five million tokens is considered small and for a spoken corpus, anything above one million tokens is considered large (p.4). However, Reppen (2010) notes that the size of the corpus depends on which language features are being observed and that if the scope of the research is very specialized, then a smaller corpus can suffice (p.32). Nelson (2010) gives a detailed literature overview of how specialized corpora went from being very large in the 1960s (20 millions tokens and plus) to being relatively small in the 1990s (one million tokens). He concludes his chapter in the book by saying that if the purpose of the corpus is to study general English, than a mega corpus is needed (billions of words); however, he adds that if a corpus is specialized (such as English for Specific Purposes (ESP) or English for Academic Purposes (EAP) corpora), then it can be small (around one million tokens). Bearing in mind that

the thesis study is similar to Coxhead's (2000) but with an extra layer of specificity (general academic English versus academic BE), a corpus of a similar size (3.5 million tokens) or smaller (one million tokens) should be enough to generate reliable results.

The *range* of texts is also a factor to be considered when creating a corpus. For instance, Coxhead (2000) used a range of four disciplines to create her academic corpus: arts, commerce, law, and science. And she further divided them into 28 sub-corpora (p. 220). The commerce sub-corpora were divided as follows: accounting, economics, finance, industrial relations, management, marketing, and public policy. To qualify for inclusion on her *Academic Word List* (AWL), a word family needed to appear frequently in the corpus, that is to say, it needed to appear at least 100 times in the entire corpus, at least 10 times in each of the four sections (arts, commerce, law, and science), and in at least 15 of the 28 sub-corpora. As we can see, careful attention was given to finding words with a high range inside the academic corpus, but this did not mean that the words had to appear in 100% of the sub-corpora (which would mean occurrence of a word in 28 out of the 28 sub-corpora). On the contrary, with a range of approximately 53% of the sub-corpora, Coxhead created a frequency list that has a high coverage of academic texts. The concept of range is important in assuring that a frequency list derived from a corpus is representative of the corpus as a whole and not only a small section of it.

After the corpus is compiled, the researcher is ready to use software to create a frequency list. A frequency list contains the most frequent tokens, lemmas or families in a corpus. Well-known lists of frequent English word families are the General Service List (GSL, 2,000 families), the Academic Word List (AWL, 570 families) and the lists by Nation derived from the British National Corpus and Corpus of Contemporary American English (BNC/COCA, 25,000 families).

The concept of *coverage* is relevant in making the connection between frequency lists and language pedagogy. Coverage refers to the proportions of running words in a text that are accounted for by a particular frequency list or set of lists. Research consistently shows that frequent words have a high *coverage* of written and spoken language, and therefore, it is important for learners to know these words. For example, the GSL covers 76.1% of the word families in Coxhead's (2000) academic corpus, and it is reasonable to assume that it would have a similar level of coverage of other academic texts.

Learning from frequency lists offers a better cost-benefit ratio than learning vocabulary randomly. The *cost-benefit* principle is a simple teaching strategy which proposes that teachers and learners focus their efforts on teaching/learning the most frequent word families first (Nation, 2013). The logic is that by knowing the most frequent word families in a language, there is a greater probability of understanding a text or being better able to communicate in general, due to the high coverage provided by these words. Students should then learn more vocabulary from these core word families first before they learn less frequent word families. In other words, learning the most frequent words increases the chances of understanding a written text or engaging in any kind of communication.

Frequency lists are used to investigate different kinds of vocabulary (academic, technical, BE), and it is important to differentiate between them.

Nation (2013) describes *academic vocabulary* as being "common to a wide range of academic texts, and not so common in non-academic texts" (p.291). Coxhead (2000) defines academic words as being those that "(e.g., *substitute, underlie, establish, inherent*) are not highly salient in academic texts, as they are supportive of but not central to the topics of the texts in which they occur" (p. 214). To complement Coxhead's point, Nation mentions that academic

vocabulary is "generally not as well known as technical vocabulary" (p. 292), thereby causing students more problems.

According to Nation (2013), *technical vocabulary* refers to the "words (that) are reasonably common in this topic area but are not so common elsewhere" (p. 19). He also says that when we see these technical words we easily recognize the topic area. The term is relevant to the proposed study, which sets out to identify a technical list for business for undergraduate students. The list will be called the Academic Business English List (ABEL); from this point onwards in the thesis, the list will be referred as the ABEL.

Another term used in framing this discussion is *English for Specific Purposes* (ESP). Strevens' (1988, as cited in Dudley-Evans & St John, 1998, p. 3) defines ESP using two subcategories: absolute characteristics and variable characteristics. The absolute characteristics of ESP are: the desire of practitioners to meet the learners' specific needs; the choice of specific materials focused on a discipline or occupation that will help in achieving these specific needs; the use of language that is appropriate for this kind of discourse (vocabulary, syntax, semantics, etc.); and finally, the definition of ESP as a separate construct from General English (GE). The variable characteristics include the idea that ESP may be restricted to the learning of one single skill (reading or lexical learning, for example) and that ESP may not follow any methodological trend such as Communicative Language Teaching (CLT). ESP can be sub-divided into English for Academic Purposes (EAP) and English for Occupational Purposes (EOP). One example of EOP is BE.

As mentioned in the previous paragraph, BE is a subset of ESP, but more precisely, Ellis and Johnson (1994) define BE as being much like ESP because they both rely on "needs analysis, syllabus design, and materials selection and development" (p. 3). Nevertheless, they say

that BE differs from ESP in regards to its general content or need to perform in a communicative business environment. In other words, they draw our attention to the fact that it is important to teach BE students skills that will help them perform tasks that are work related such as negotiating, delivering an effective presentation and developing meeting skills in the context of business and commerce.

Now that the key terms used in this paper were defined, I will turn to the role of vocabulary in reading comprehension.

Vocabulary in reading and comprehension

A great deal of research has been done regarding the role of vocabulary knowledge in L2 reading comprehension. One of the first studies was by Laufer (1989); she found that students needed to understand 95% of the vocabulary in a text in order to score 55% or more on a comprehension test. In a follow-up study, Hirsh and Nation (1992) concluded that learners needed to know 98% of the vocabulary in an unsimplified text in order to achieve "pleasurable reading", but that even this high level of lexical knowledge did not guarantee high levels of comprehension. These two studies were landmarks in establishing the level of known-word coverage needed to understand a text. A problem in this line of research is that it is not clear what 'adequate' comprehension is; therefore, the comprehension test results in Laufer (1989) described above could be considered relatively low.

As software technology advanced, corpora became bigger and data analyses more accurate. This allowed for updated lists of frequent vocabulary to be created and for closer studies of the relationship between vocabulary size and comprehension. For instance, using

frequency lists based on the BNC, Nation (2006) determined that, in order to achieve 98% coverage, knowledge of 8,000 to 9,000 word families was needed for comprehension of written texts and 6,000 to 7,000 word families were needed for spoken texts. Van Zeeland and Schmitt (2012) decided to challenge Nation's (2006) listening comprehension figures. They did so because they questioned whether the 98% coverage mark from reading comprehension studies could be automatically transferred to listening. The results of their study were revealing in showing that only 90%-95% vocabulary coverage, or 2,000 to 3,000 word families, could provide participants with good listening comprehension rates. In a study with lower secondary students in Denmark, Stæhr (2008) investigated the correlation between vocabulary size and listening, reading and writing skills. He found that participants' receptive vocabulary size was strongly associated with their ability to read and write, but that it was only moderately associated with their listening comprehension. Although these studies use different methodologies, they point us to similar results: L2 learners need a larger vocabulary size to read than to listen. Because reading is important and challenging to international undergraduate business students, in this study I focus on written academic texts in business with the goal of addressing the larger vocabulary requirements associated with it.

Many of the studies mentioned above concentrate on student populations studying GE, but researchers also recognized a need to do more research on Academic English (AE) in order to understand the challenges and needs of these distinct students. With that aim in mind, Laufer and Ravenhorst-Kalovski (2010) investigated the threshold for reading comprehension among college students. In this study they suggested two possible thresholds: optimal (consisting of 8,000 words for 98% text coverage) and minimal (4,000 to 5,000 for 95% text coverage). In a similar study, Schmitt, Jiang and Grabe (2011) tried to find a threshold in reading

comprehension. More specifically, they wanted to know if the relationship between comprehension and vocabulary coverage could be represented as a linear, threshold or curved graph. Their results showed that there was no threshold in reading comprehension, and that the vocabulary coverage is fairly linear, such that greater vocabulary gradually leads to more text coverage. That is to say, there is no threshold where a specific vocabulary size suddenly increases coverage and reading comprehension. Most importantly for this study, their results suggest that coverage of 98% is reasonable when reading academic texts, which is the case of our target student population. In a more specialized study, Hsu (2011b) investigated the amount of vocabulary needed in order to reach 95% and 98% coverage in academic texts used in business courses. She started by creating two corpora, one for business textbooks from undergraduate core courses (consisting of 7.2 million tokens) and another one from business research articles (consisting of 7.62 million tokens). Her results showed that in order to achieve 95% coverage, knowledge of 3,500 word families was needed for textbooks and 5,000 word families were needed for research articles (plus proper nouns). However, these numbers jumped up to 5,000 and 8,000 respectively when the desired coverage was 98%. These numbers are not very different from the ones presented in previous studies in this thesis. Nevertheless, what is interesting to notice is that the amount of vocabulary needed by undergraduate students (the ones who read core course textbooks) is significantly smaller than the vocabulary needed by graduate students (the ones reading research articles). Unfortunately Hsu's study does not provide the readers with a specific BE list. Instead, the author suggests that creating a specialized BE list would be a desirable follow-up study, and this is indeed the intention of the present study.

In an academic environment, high levels of comprehension are expected of students, and vocabulary knowledge is an important determiner of reading comprehension success but not the

only one. Other reading factors include the students' level of education, their ability to read in their first language, their previous knowledge of the topic, their grammar knowledge and the distance between their L1 and L2 (see Bernhard, 2011 for an overview). Nevertheless, research has shown that vocabulary knowledge is the foundation block that allows students to understand a text. Numerous studies point to a strong direct link between vocabulary coverage and reading comprehension (e.g. Hirsh & Nation, 1992; Laufer, 1989; Laufer & Ravenhorst-Kalovski, 2010; Nation, 2006; Schmitt et al., 2011; Stæhr, 2008).

Lists of frequent words have greatly contributed to helping us determine the lexical frequency coverage of texts and how to use them effectively in teaching and in materials creation. Most importantly, frequency lists are very useful in telling us which words should be learned first if we are to follow the cost-benefit principle. Let us now take a look at the most popular frequency lists in our field today.

Frequency lists

In a pioneer project, West (1953) created the General Service List (GSL), which consists of the 2,000 most common word families in English (henceforth, I will use the notation 1K and 2K to refer to 1000-word frequency bands such as the GSL 1000 and 2000). More recently, Coxhead (2000) created the Academic World List (AWL), which consists of 570 additional word families that recur frequently across four academic disciplines (arts, commerce, law and science). These are all-purpose academic words such as *chapter, hypothesis* and *theory*. As shown in Table 1 these two lists together cover 86.1% (71.4% from the first GSL 1K, 4.7% for the second GSL 1K, and 10% from AWL) of lexical items in academic texts in general, leaving a gap of

13.9% of word families still unidentified in a frequency list. Table 2 provides information that is specific to business. As we can see, the GSL 2K plus the AWL cover 88.8% of word families in academic texts in commerce (Coxhead, 2000, p. 224) leaving a gap of 11.2% in this specific field. These coverages are high in general terms, but they are not high enough to reach the 95-98% mark suggested by researchers to be necessary for good text comprehension (Hirsh & Nation, 1992; Laufer, 1989; Laufer & Ravenhorst-Kalovski, 2010; Nation, 2006; Schmitt et al., 2011; Stæhr, 2008). Therefore, these lists combined do not fully contribute to L2 learners' achievement of the minimum comprehension of academic texts. To fill this gap, I propose to create a new specialized list in which the lexical coverage will be increased to the 95-98% level when combined with a core frequency list and proper nouns. That is, it will fill in the gap that remains once learners have knowledge of the most frequent general words and the most frequent academic words. As mentioned above, the list to be created in this research has a technical focus, and it will be called ABEL. Its aim will be to identify the most common word families in the academic business field.

Table 1

Gap left in general academic texts

Vocabulary Lists	Coverage of Academic Texts in Coxhead's (2000) Corpus	Cumulative coverage
GSL 1K	71.4%	71.4%
GSL 2K	4.7%	76.1%
AWL	10%	86.1%
Other words (gap)	13.9%	100%

Table 2

Gap left in Coxhead's (2000) Commerce sub-corpus

Vocabulary Lists	Coverage of Texts in Coxhead's (2000) Commerce Sub-corpora	Cumulative coverage
GSL 1K	71.6%	71.6%
GSL 2K	5.2%	76.8%
AWL	12%	88.8%
Others (gap)	11.2%	100%

Many researchers believe that technical lists can be very useful in increasing lexical coverage in ESP texts (Chen & Ge, 2007; Coxhead, 2000; Hyland, 2008; Hyland & Tse, 2007;

Nation, 2013; Sutarsyah, Nation, & Kennedy, 1994), even though they have very different points of view in that regard. Coxhead (2000), Nation (2013) and Sutarsyah et al. (1994) for instance, believe that a general academic list like the AWL is a cost-effective way to deal with EAP core vocabulary since it has been shown that with them the text coverage becomes high. In addition, a general academic list allows for communication across fields, and it consists of the core vocabulary across disciplines. Furthermore, they believe that specialized frequency lists like the one proposed here can successfully complement the GSL and the AWL when it comes to increasing text coverage. On the other hand, Chen and Ge (2007), Hyland (2008) and Hyland and Tse (2007) criticized the AWL for its generic nature by showing that the list behaves very differently depending on the discipline. They suggest that instead of a general academic list, each discipline should have its own specific frequency list. In this study I took the approach defended by Coxhead (2000), Nation (2013) and Sutarsyah et al. (1994) because I hypothesize that the AWL will show great coverage of the ABEL corpus and because I believe that the knowledge of general academic vocabulary is an important foundation for university learners as they are typically required to take courses in a variety of disciplines; as I see it, the goal of higher education is to develop whole persons, not just area specialists. Building on that general academic foundation by acquiring specialist, discipline-specific vocabulary is then a logical second step and the ABEL is designed to facilitate this process. Therefore, the ABEL should be built by first excluding the words on a core GE list as the GSL and a core academic frequency list as the AWL.

As mentioned previously, Coxhead (2000) excluded words from the GSL 1K and 2K lists (West, 1953) from the academic corpus to create her AWL. Although "the GSL has been criticized for its size (Engels, 1968), age (Richards, 1974) and need for revision (Hwang, 1989)"

(cited in Coxhead, 2000, p. 213), Coxhead used the GSL because it still covered 76% of the words in her academic corpus. For these reasons, it is important to test different frequency lists and learn more about their coverage on this specific corpus before choosing one. In addition to an updated list of frequent GE words that the target learners can be expected to know, I would also like to identify an existing list of general academic words that they would likely already know, and these too will be excluded from ABEL. The reason it will be important to exclude a core academic frequency list is that it will serve as a bridge between the general vocabulary and the more specialized BE list. In other words, it would be improbable that ABEL could help learners transition directly from a general English list without leaving a gap in their knowledge. Nation and Macalister (2009) also support this vision when describing the vocabulary stages as having three parts: stage one - 2,000 GE word families; stage two - 570 general academic word families; and stage three - approximately 1,000 specialized/technical word families (p.31). The AWL is a strong candidate to serve as this bridge, but the weaknesses mentioned above regarding the GSL are not negligible, and they certainly cascade into the AWL itself. Above we mentioned that the BNC/COCA could be a good alternative for the GSL, but there is no AWL equivalent to it which would serve as a bridge between BNC/COCA and ABEL. The recent publication of the NGSL and NAWL seems to be more promising as they solve the issue of using a small and old corpus, on top of offering a reported 92% coverage of tokens in an academic text (Browne, Culligan, & Phillips, 2013). This study will investigate the lexical coverage that would result from using each of these individual lists and their respective academic list (if applicable). The one that results in the highest coverage will officially be the core frequency list to be excluded before the creation of ABEL.

It is important to note that a study of this kind has already been attempted by Konstantakis (2007), who also aimed to create a BE frequency list for undergraduate business students. Using the Published Material Corpus (PMC), he created the Business Word List (BWL) consisting of the 560 most frequent word families in BE textbooks for ESP students. The weaknesses of this study are that the corpus used to create the frequency list is rather small (only 600,000 tokens while Coxhead had approximately 3.5 million in total and a little over 879,000 just for the commerce subsection), and it is based on ESL BE textbooks designed for L2 learners of English instead of authentic business materials. The latter issue seems to be particularly problematic because it runs counter to the logic of this kind of corpus study. In other words, L2 learners will benefit most from becoming aware of what native speakers (NS) use as authentic language and not what materials creators intuitively decide is more important to teach to L2 learners. Unfortunately, the result of the study by Konstantakis was a list of words based on simplified materials that may not correspond to the authentic language that students have to cope with while in university. The corpus used seems inconsistent with the targeted goal and shows an evident issue regarding corpus design. That is to say, Konstantakis' aim was to create a BE frequency list with the most common words used in academic business so that prospective undergraduate business students could prepare for their studies abroad. However, as he created a frequency list based on ESL business textbooks, the result was a list of the most common business terms taught to L2 students and not the most common vocabulary in authentic business textbooks.

Hsu (2011a) also created a frequency list from business materials called the Business Word List (BWL) (confusingly, the same name used by Konstantakis, 2007). However, this list was created based on a corpus consisting of research articles only, which generated a list of more

"advanced" words intended for graduate students' use. Thus it appears that neither list addresses the needs of our targeted student population. This group consists of undergraduate business majors who must read business textbooks designed for native speakers. The need for such a list becomes clear when the Konstantakis' BWL and Hsu's BWL are compared. Surprisingly, there is little overlap between both lists. Konstantakis' list has 560 word families while Hsu's list has 426; however, only 59 word families overlap; words that appear on both lists such as *asset, enterprise,* and *headquarters,* are clearly useful for business students of any level to know. However, unique to Konstantakis' list are some very basic items such as *fax, menu* and *television*, while Hsu's contains words that seem very specialized such as *exogenous, idiosyncratic* and *stochastic*. This thesis research is intended to specify closely those words that will be most useful for undergraduate business students to know. This involves gathering a large corpus of authentic undergraduate business texts and deriving a list of frequent technical families from it. Because it draws on a different kind of corpus, the ABEL is expected to differ from both Konstantakis' and Hsu's BWL.

In the following section the research questions are presented as well as the rationale and hypothesis underlying each one of them.

Research questions

Since there is clearly a need for an improved list based on authentic business materials, the aim of this study is to create an academic word list specialized in BE to help L2 undergraduate students in business cope with the vocabulary in their academic reading. Creating such a list involves gathering an appropriate corpus and identifying the frequent words in it. The

first part of the job is to determine the coverage power of core and academic (i.e., non-technical) words in the corpus. Once this has been determined, then the technical list can be compiled. This sequence is reflected in the research questions, which are as follows:

RQ 1 - Which existing core and academic frequency lists (approximately 2K-3K word families) provide the greatest text coverage of the ABEL corpus?

Considering that the desirable text coverage (proportion of words in a text that are known to the reader) is between 95% (minimal) and 98% (optimal) (Laufer & Ravenhorst-Kalovski, 2010), different general and academic frequency lists will be tested to see which one results in the highest text coverage of the ABEL corpus before the selection of words to be part of ABEL starts. In other words, it is assumed that these students already know the most frequent 2K or 3K word families (depending on the list) plus the core vocabulary in academic English (AWL or similar list). Frequency lists to be tested are the GSL and the AWL, the New General Service List (NGSL) and the New Academic Word List (NAWL) (Browne, Culligan & Phillips, 2013), and the BNC/COCA. My hypothesis is that the more recent NGSL plus the NAWL will provide a greater coverage than the old GSL combined with the AWL because they are larger and more current. I also hypothesize that the NGSL combined with the NAWL will offer greater coverage than the BNC/COCA because the second has no academic frequency list to accompany it; therefore, it is lacking a more specialized list which would cover more word families. Once core frequency list(s) has (have) been determined, words in it (them) will be put aside and I will look at the words that are left in order to create the ABEL. This leads us to the next research question:

RQ 2 - Which word families occur most frequently in authentic university level commerce texts excluding the ones in a core frequency list (or combination of core lists)?

This research question addresses the main topic of this study as answering it will unveil the actual content of the ABEL list. Once the initial version of the list is created, it will be important to analyze it against each of the sub-corpora in ABEL (outlined in the Methodology chapter) to assure that the word families have a wide range. Words which appear in only one or a few sub-corpora should be excluded because they are too specific to a particular area of study. That is to say, if a word family appears only in the accounting sub-corpus, it cannot be considered a general BE word and has to be considered an accounting term only. This range principle leads us to research question number three:

RQ 3 - Do the lexical items occur with different frequencies in different sub-corpora such as marketing or finance?

Not only will I make sure that the word families in ABEL appear with a wide range, I will also observe in which frequency they appear in each of the sub-corpora.

Considering that a frequency list with the same purpose has already been created (the BWL by Konstantakis, 2007) as well as a list with a similar purpose (Hsu, 2011a), this leads us to question how different or similar these lists are from ABEL. We can hypothesize that there will be some overlap between them since they are all based on business corpora, and this will be a good indicator of the business character of the lists. However, some other word families should be different due to our use of different corpus and methodology. Therefore, research question four is:

RQ 4 - Are the words in ABEL different from Konstantakis' (2007) and Hsu's (2001a) BWLs? If so, how?

The reason for posing this question is to validate the usefulness of ABEL and its contribution to research in vocabulary. For instance, we saw earlier that Konstantakis' BWL is

quite different from Hsu's. If we conclude that ABEL is yet substantially different from these two lists, it could indicate that as suspected, Konstantakis' list is indeed not advanced enough for undergraduate students and that Hsu's list is too advanced. This would mean that ABEL would not be a redundant list, but a useful one for this specific student population.

Another indicator that ABEL has a true business character will be its coverage of non-academic corpora compared to academic business corpora. That is to say, ABEL should cover more of the words in academic business corpora than in non-academic corpora. Therefore, this leads us to the last two research questions:

RQ 5 - How well do the words in ABEL cover non-academic texts?

and

RQ 6 - How well do the words in ABEL cover another academic business corpus?

This means that ABEL will have to be validated against two other corpora: one non-academic one consisting of books and/or magazines and another business corpus consisting of different texts. If the ABEL is valid, it will perform poorly in the non-academic corpus showing inferior coverage to the one shown with business text. On the other hand, ABEL will be expected to perform fairly similarly in a different business corpus.

In this section the research questions and hypothesis were presented as well as the rationale for each one of them. In the next chapter I will begin with a discussion of issues related to the composition of the proposed list and corpus, and I will outline solutions with reference to the research questions.

Chapter three – Methodology

In this part of the study I outline how the ABEL corpus was created. I also describe other materials and tools used to test the ABEL and to answer each research question.

Materials

Corpus design. The ABEL corpus is composed of 15 textbooks currently used within the 14 core courses of Concordia University's undergraduate Business program (see Appendix 1 for the full list of textbooks and their lengths in numbers of pages). Two major issues were considered when choosing this type of corpus design: choosing a corpus containing authentic materials representing the actual language students will come into contact with; and finding a trustworthy source of texts.

This approach is very different from Coxhead's (2000) and Hsu's (2011b) corpus designs, where great emphasis was placed on ensuring that the sub-areas of their corpora were all equivalent in size. As mentioned previously, Coxhead (2000) built a corpus of approximately 3.5 million tokens, and this total was sub-divided into four disciplines: arts, commerce, law, and science (which means approximately 875,000 running words for each section). This approach of assigning each sub-area the same weight was logical given that the goal was to create a generic, core academic frequency list. Indeed, it is not realistic to think that all students will read all texts from all disciplines during their undergraduate studies, justifying such a balanced approach. However, since the goal of the present research was to create a core frequency list in one specific field of study (business), it was decided that the corpus should contain an exact representation of

what students would actually read, despite unbalanced text length. The logic of this approach is that reality can be unbalanced in the sense that students do not necessarily read equal amounts of text in all areas of business. This characteristic was explored by using authentic materials without manipulating their original length.

The second major reason for choosing this "intact" sample of business textbooks was to address the issue of having to take responsibility for choosing the material which becomes part of the corpus, as the author is not an expert in the business field. Rather than selectively deciding which material should be included, the author adopted materials from an existing (and very reputable) program. In doing so, responsibility was transferred to a team of specialists who have already carefully selected what content they believed to be most relevant for business students.

The decision concerning the adoption of core courses only (as opposed to all courses) comes from a combination of the purpose of this study and the nature of the business programs in general. The aim of this study was to create an academic business English list that represents the core vocabulary of the business field. That is to say, the vocabulary should be relevant to undergraduate students of all specializations of business (e.g. marketing, entrepreneurship, accounting, finance, etc.). In fact, most business programs tend to have core courses that are mandatory for all business majors, which are usually taken in their first year of studies before specializing in a specific business sub-field. Students' vocabulary acquisition will thus most probably follow a funnelling process of going from GE, to AE, to academic BE, and finally to one of the business sub-fields as they further their studies. In this study the academic BE stage of this funnel is re-visited and a new BE frequency list is proposed.

Concordia's business school website (Concordia University, 2015) was consulted to verify the relevant list of courses and their textbooks. Concordia's core courses were also

compared with their equivalents at two other major universities in Canada: McGill University (McGill University, 2015) and the University of Toronto (University of Toronto, 2015). Table 3 was created to compare the three programs, as well as Coxhead's (2000) six Commerce sub-areas. As is evident from the table, while the programs are not identical, they are very similar. Furthermore, it should be noted that Coxhead's (2000) sub-corpora selection was much smaller than the three university programs (only six items in Table 3 compared to 14-15 from the other universities), but this is understandable due to the fact that she did not specialize in BE. These differences could be considered as a limitation in this kind of corpus design; however, as the majority of subjects from the universities do overlap, I believe this "intact" approach of corpus building (as opposed to artificially creating sub-fields of business and controlling their lengths) is nevertheless suitable for the purpose of this study. In addition, in order to better understand this option of corpus design, we can take into consideration what Clear said about corpus linguistics:

> (...) corpus linguistics: it's like studying the sea. (...) both are very very large… - and difficult to define precisely, - subject to constant flux, currents, influences, never constant, - part of everyday human and social reality. Our corpus building is analogous to collecting bucketfuls of sea water and carrying them back to the lab. It is not physically possible to take measurements and make observations about all the aspects of the sea we are interested in *in vivo*, so we collect samples to study *in vitro.* (Clear, 1997, as cited in Nelson, 2010, p.57)

Table 3
Comparison of core courses from three major Business Schools in Canada (items marked with an asterisk indicate non-correspondence across the row)

Concordia University	McGill University	University of Toronto	Coxhead (2000)
Business Communication		Business Communication Skills	
Contemporary Business Thinking	Social Context of Business	Introduction to management I*	
Business Statistics	Business Statistics	Introduction to management II*	
Financial Accounting	Introduction to Financial Accounting	Introductory Financial Accounting I	Accounting
Analysis of Markets	International Business	Principals of International Marketing*	
Organizational Behaviour and Theory	Introduction to Organizational Behaviour	International Organizational Behaviour	
Marketing Management I	Marketing Management I	Principles of Marketing	Marketing
Productions/Operations Management	Operations Management	Managing in Organizations	Management
Business Technology Management	Information Systems	Managing Groups and Organizations*	
Managerial Accounting	Managerial Economics*	Introductory Management Accounting	
Introduction to Finance	Finance I	Principles of Finance	Finance
Business Law and Ethics	Organizational Policy*	International Human Resources*	
Entrepreneurship	Macroeconomic Policy*	Managing People in Organizations*	Economics
Strategy and Competition		Introductory Financial Accounting 2*	Industrial Relations*

Given the impracticality of collecting data from multiple universities across the world and comparing programs, this sample was chosen because of its convenience but also because of its potential accurate representation of core Business language. In other words, although this combination of textbooks is specific to Concordia University's Business program, it is hoped that ABEL will represent core Business language as a whole and will be useful to international Business students attending any English-medium university in the world.

As mentioned earlier, full information on the textbooks used can be found in Appendix 1. The majority of textbooks are either custom editions for Concordia University or Canadian editions of the book; however, the contents of the books are virtually identical to the American editions, except for the addition of one or two chapters. It thus seems to be the norm that business textbooks are highly standardized worldwide, but may contain localized content for any given country. Analysis of Concordia University's custom editions revealed that they were merely abridged forms of the original US versions. It seems that instead of making students pay for 25 chapters of a book that has over one thousand pages, Concordia offers a smaller version of the same book which contains only the chapters that will be used in class, consequently costing students less (as the number of pages is reduced to around 600-700). It was also observed that one of the custom textbooks was simply a combination of half of two different American editions. Due to these issues, though the materials used to create the corpus seem to be very specific to Concordia University, it is nevertheless believed that the words generated from them will be of value to students in other parts of the world.

Copyright authorization. Once the corpus design was settled and the decision to use the textbooks in their entirety was made, the legal authorization required to scan the books was obtained. It is important to note that in Canada such authorization is granted if the materials are to be used for private research.

Creating the corpus. Appendix 1 lists the business courses at Concordia University as well as their relevant textbooks. The only course with more than one textbook was Entrepreneurship, which had two: one on entrepreneurship and a very small one on plagiarism. The corpus creation procedure was pilot tested by working with only one book to start with (from the Contemporary Business Thinking course – one of the smallest ones); subsequently, the following process was undertaken for the rest of the material. All books were scanned using a flatbed Epson GT 15000 with a resolution of 300 dpi, and saved as JPEG files. After the entire textbook was scanned, all individual files corresponding to the book were converted into a single PDF document using Adobe Acrobat XI Pro. The PDF files were then individually submitted to the optical character recognition (OCR) software ABBY FineReader Express in order to transfer them from PDF to text format (.txt). Once in text format, the books were edited manually and using the Vim software to correct any errors in the ORC recognition. By the end of the process the ABEL corpus had approximately 3.5 million tokens. The files were edited to exclude headings, title pages, tables consisting largely of numbers and other features that were not considered true reading material. Full details regarding material that was excluded from the textbooks can be found in Table 4.

Table 4

Content excluded from and kept in the ABEL corpus

Excluded	Included
Cover page and introduction pages	Preface
Table of contents	Author's biography
Headings	Numbers that were part of sentences
Page numbers	Appendices
Index	Endnotes
Tables consisting mainly of numbers	Glossaries
Mathematical formulas	Exercises
Answer keys consisting of numbers only	Answer keys
Appendices consisting of numbers only	
Photo credits	
Reference lists	
Text in another language (French in the Law textbook)	

Creating the list

To create the BE frequency list per se, the list of the most common words in English language as well as core academic vocabulary needed to be identified (i.e., RQ1: Which existing core and academic frequency lists (approximately 2K-3K word families) provide the greatest text coverage of the ABEL corpus?). This was done using AntWordProfiler (Anthony, 2014), a software program which automatically categorizes all the words of an entered text according to their appearance on lists chosen by the user (i.e., the percentages of word families on each list). In the present study, the ABEL corpus was uploaded and compared to the default lists available in AntWordProfiler (GSL 1K, 2K and AWL). The output showed the coverage percentage of each of these lists individually as well and their cumulative coverage. The output also provided information about the number of word families in each frequency list as well as the total number of tokens in the ABEL corpus. This procedure was then repeated using the NGSL+NAWL and the BNC/COCA lists.

Some of the vocabulary which remained unclassified after this process was categorized using BNC/COCA list number 31, which consists of over 21 thousand proper nouns, and then excluded from the data because they are considered to be easily recognized (the capital letter being the main signal that the word is the name of a person, place, company, etc., rather than a true lexical item). Once these words were excluded, another list of proper nouns was manually created, excluding the ones that were not included in the BNC/COCA 31, but were still in the corpus. These proper nouns consisted mainly of company names such as Wal-Mart, Facebook and Twitter, name of places such as UK and Boucherville, letters of the alphabet excluding the letters “a” and “I”, and website addresses (including www, http and com).

Once the core frequency list(s) and the nouns were identified, the next step was to address RQ2: which word families occur most frequently in authentic university level commerce texts excluding the ones in a core frequency list (or combination of core lists)? This involves creating the actual list and deciding on the criterion of inclusion of word families. In other words, we first identify and exclude (or put aside) the words that are more frequent in the General English lists (RQ1) and then we look at what is left and create ABEL. The words to be included in ABEL were decided according to their frequency range in the corpus. As previously stated, Coxhead (2000) adopted the standard that each word had to appear in at least 53% of the sub-corpora in her academic corpus (at least 15 out of the 28 sub-corpora). Taking into consideration the success of the AWL, it is reasonable to require that in order to be part of ABEL each word also has a range of around 50% within the sub-corpora, meaning that a word family would have to appear in at least seven of the 15 textbooks used to create the corpus. Seven was chosen rather than eight because two books in the corpus are fairly small (No. 1 and 14, see Appendix 1); therefore, it would be relatively hard to find occurrences of many word families in them. All word families that met the range seven criterion were put into a .txt file and categorized into full families using the ones present in the BNC/COCA. Those word families which did not already exist in the BNC/COCA were manually created. For example, the headword *businessperson* and its family (*businesspersons*, *businesspeople* and *businesspeoples*) were missing and therefore needed to be added. This procedure concluded the steps required to answer RQ2.

To address RQ3 (Do the lexical items occur with different frequencies in different sub-corpora such as marketing or finance?), the same procedure used to answer RQ1 was adopted, the only difference being the application of the ABEL list and testing the coverage of each sub-corpora individually.

Validating ABEL

RQ4 asked about the validity of ABEL and its contribution to research by comparing it with the two existing BWLs (Are the words in ABEL different from Konstantakis' (2007) and Hsu's (2001a) BWLs? If so, how?). The Lextutor tool "Text Lex Compare" was used to address this research question. ABEL was first entered into the Text Lex Compare software, and then each of the two BWL lists was entered one at a time. The Lextutor output provided the percentage of words that overlapped as well as the complete list of overlapping words. The list of words which did not overlap was also available with the results.

Testing ABEL against other corpora

To answer the last two research questions, two other corpora were needed for comparison. Research Question 5 asks about ABEL's coverage of non-academic corpora (RQ 5 - How well do the words in ABEL cover non-academic texts?). To address this, a non-academic corpus of 988,523 tokens was created from six novels retrieved from Project Gutenberg (Project Gutenberg, 2015). These books were chosen because copyright restrictions do not apply to them anymore, they were already edited and in the right format (.txt), and because the texts are relatively close to everyday life or general English (as opposed to a technical corpus such as the ABEL corpus). The texts were downloaded in .txt format and processed in AntWordProfiler. The titles of the six books, classic novels by Dickens and Twain, can be found in Table 5. A corpus of other similar academic business English textbooks was required in order to answer the last

research question (RQ 6 - How well do the words in ABEL cover another academic business corpus?). Permission from the creators of the "First year textbook corpus" (Wood & Appel, 2014) was obtained to run this last test on 774,042 tokens from five textbooks (excluding five engineering books). This corpus was the perfect fit for this part of the study because it shares many characteristics with ABEL, but on a smaller scale. This corpus also consists of textbooks from business core courses scanned in their entirety from a program at a Canadian university (Carleton). Fortunately, the books from the First year textbook corpus did not overlap with the books used to create ABEL. The complete list of materials from the First year textbook corpus can be found in Table 6.

Table 5

List of books composing the non-academic corpus

A Christmas Carol – Charles Dickens
Little Dorrit – Charles Dickens
The Life And Adventures Of Nicholas Nickleby - Charles Dickens
Adventures of Huckleberry Finn Complete – Mark Twain
The Adventures of Tom Sawyer Complete – Mark Twain
The Prince and The Pauper Complete – Mark Twain

Table 6

List of books in the First year textbook corpus

Friedman, T. (2008). *The world is flat: A brief history of the twenty-first century*. New York: Farrar, Straus and Giroux.
Hilton, R. (2011). *Managerial accounting (9th ed.).* Boston: McGraw-Hill.
Kimmel, P.,Weygandt, J., Kieso, D. (2011). *Financial accounting: Tools for business decision making (6th ed.)*. Massachusetts: John Wiley & Sons.
Lay, D. (2003). *Linear algebra and its applications (3rd ed.)*. Boston: Addisone Wesley.
Mankiw, G. (2004). *Principles of macroeconomics (3rd ed.)*. Ohio: Thomson Higher Education.

In this chapter the materials and tools used to answer the six research questions proposed in this study were presented. In the next chapter, the results for each of them are presented along with a discussion.

Chapter four - Results

This chapter presents answers to the six research questions presented in this study. Research Question 1 asked which existing core and academic frequency lists (approximately 2K-3K word families) provide the greatest text coverage of the Academic Business English List (ABEL) corpus. As mentioned in the methodology section, all list combinations were entered into AntWordProfiler to test for their coverage; the results for this test are presented in Table 7.

When we look at the first set of lists, which were the two parts of the General Service List (GSL) and the Academic Word List (AWL), we can see that the GSL 1K list covers 72.63% of the ABEL corpus, which is very impressive for a list that was created in 1953. The 2K list covers 5.98% and the AWL covers 10.83%; this indicates that with a total of 2,570 word families, this set of lists has a total cumulative coverage of 89.44% of the corpus, leaving 10.55% for the creation of ABEL (less proper nouns). This result challenges the hypothesis that the GSL and AWL would not be a good fit as core lists to be excluded from ABEL.

The second set of results used the New General Service List (NGSL) 1K, 2K and 3K and the New Academic Word List (NAWL) (Browne et al., 2013). Note that the NGSL includes a third set of words that the creators call the 3K, but that actually consists of 801 words. The reason for the adoption of this extra set of word families is that the creators of the NGSL believed that that the third thousand is essential for basic communication in English. Coverage of the ABEL corpus by these lists amounts to 75.47% for the first K (higher than the old GSL1K), 8.76% for the second K (also higher than the old GSL 2K), and 3.32% for the third K (no equivalent to the old GSL). What was not so surprising was the performance of the NAWL: 2.23% as compared to 10.55% from the old AWL. Perhaps the coverage of the NAWL could be

expected to be smaller than the coverage of the AWL because

Table 7

Coverage of the ABEL corpus by three sets of frequency lists

Frequency list	Coverage per list	Cumulative coverage
GSL 1K	72.63%	72.63%
GSL 2K	5.98%	78.61%
AWL	10.83%	89.44%
NGSL 1K	75.47%	75.47%
NGSL 2K	8.76%	84.23%
NGSL 3K	3.32%	87.55%
NAWL	2.23%	89.78%
BNC/COCA 1K	67.10%	67.10%
BNC/COCA 2K	15.25%	82.35%
BNC/COCA 3K	9.31%	91.66%
BNC/COCA 4K	1.75%	93.41%

the words in the NGSL 3K could cover some of its word families. However, even when we sum up these two lists (NGSL 3K 3.32% + NAWL 2.23% = 5.55%), their coverage only achieves about half of the coverage in the old AWL (10.55%). The total coverage of the NGSL and NAWL was 89.78%, which is slightly higher (0.34% more) than its old equivalents. However, the NGSL and the NAWL are composed of 3,764 words (1K +1K + 801 + 963 for NAWL),

amounting to almost 1200 words more than the 2570 words of its older version, but with only a very small amount of additional coverage. In pedagogical terms this means that the old lists are more cost-efficient for students because they have fewer words to learn.

The third set of results show the performance of the British National Corpus and the Corpus of Contemporary American English (BNC/COCA) lists. The first thousand words (1K) covered 67.10% of the corpus (less than both previous lists), the second thousand covered 15.25% (more than both previous lists), and the third covered 9.31% of the corpus –a cumulative coverage of 91.66% for 1K + 2K + 3K words, which is in between both previous frequency lists in terms of number of word families, but higher in terms of text coverage. Going further one extra K-level provides a very small amount of additional coverage (1.75%), for a cumulative coverage of 93.41%. For this reason, this fourth thousand list was disregarded during subsequent analyses.

As the cumulative coverage of these three lists showed very similar results (89.44%-GSL+AWL, 89.78%-NGSL+NAWL and 91.66%-BNC/COCA) their coverage was explored further by testing the ABEL textbooks individually. Each of the 15 textbooks was evaluated in terms of the coverage of the three sets of lists. This resulted in three coverage 'scores' for each of the 15 books. Means of these coverage figures (shown in Table 8) were then tested for differences using a one-way ANOVA. Results showed that there were significant differences in the data ($df = 2,14$, $F = 35.73$, $p < .0001$). Post hoc t-tests ($t = 6.00$, $p < .0001$) showed that the BNC/COCA coverage mean differed significantly from the mean for GSL+AWL coverage. The difference between the BNC/COCA and NGSL+ NAWL means was also significant ($t = 7.83$, $p < .0001$). However, the GSL+AWL and NGSL + NAWL means did not differ significantly from each other ($t = 0.44$, $p > 0.66$). These findings suggest that the BNC/COCA is indeed a more

powerful set of lists than the other two, though with all three coverage levels near the 90% mark, it is clear that the advantage is small.

Table 8

Means coverage of the 15 textbooks

	GSL/AWL	NGSL/NAWL	BNC/COCA
Means	89.63	89.72	91.50
Standard Deviation	1.59	1.55	1.24

Nonetheless, the small advantage may make a difference in terms of the reading experience. In practice, if there are 400 words on a typical page of a textbook, and the reader has 90% coverage (as is the case with the GSL+AWL and NGSL+NAWL), 10 words in every 100 are unknown, meaning that 40 problem words per page have to be guessed from context or looked up in a dictionary. With 91.5% coverage, only 8.5 words per 100 are unknown, which works out to 34 problem words per textbook page instead of 40. It seems reasonable to think that the reader would experience this as a real advantage. This led the author to opt to work with the BNC/COCA scheme and proceed to RQ 2: Which word families occur most frequently in authentic university level commerce texts excluding the ones in a core frequency list(s)? To answer this question, attention was given to the words that were left after the BNC/COCA 1K, 2K and 3K were excluded, as well as BNC/COCA baseword list 31 (proper nouns) and the manually created proper noun list. That is to say, words that we exclude because we suppose students already know them. As previously mentioned, the words were organized by range of occurrence, from range 15 (words that appeared in all 15 books) all the way down to range 1

(words that occurred in only one book). Words from range seven up to range 15 were selected to be part of ABEL. The result was a list of 840 word families that had coverage of 2.83% of the ABEL corpus (see Appendix 2 for the headwords of ABEL). Some of the headwords included in the ABEL were *accountant, browse* and *merchant.* The details of the coverages of all frequency lists can be found in Table 9.

Table 9

Coverage of core lists, nouns and ABEL

Frequency list	Coverage	Cumulative coverage
BNC/COCA 1K	67.1%	67.1%
BNC/COCA 2K	15.25%	82.35%
BNC/COCA 3K	9.31%	91.66%
Baseword list 31	1.36%	93.02%
Extras	1.09%	94.11%
ABEL	2.83%	96.94%
Other	3.06%	100%

Research Question 3 asked: Do the lexical items occur with different frequencies in different sub-corpora such as marketing or finance? To answer this question, each of the sub-corpora were individually tested against ABEL using AntWordProfiler. The results are shown in Table 10. The coverages varied from a low 1.83% for the first textbook in critical thinking to a high of 3.62% in finance. Only this first book had coverage below 2%, while eight textbooks were in the 2% range, and six had coverage greater than 3%.

To answer Research Question 4 (Are the words in ABEL different from Konstantakis' (2007) and Hsu's (2001a) BWLs? How?), Lextutor's Text Lex Compare software was used to examine the differences and similarities between ABEL and the two BWLs. Konstantakis' BWL

Table 10

Coverage of ABEL in each one of the sub-corpora (note that to keep this table short and easy to follow, the course names were used instead of the name of the books)

Contemporary Business Thinking	1.83%
Business Communication	2.91%
Business Statistics	2.61%
Financial Accounting	3.62%
Analysis of Markets	2.58%
Organizational Behaviour and Theory	2.22%
Marketing Management 1	2.26%
Productions/Operations Management	3.20%
Business Technology Management	2.73%
Managerial Accounting	3.18%
Introduction to Finance	3.00%
Business Law and Ethics	2.10%
Entrepreneurship 1	3.36%
Entrepreneurship 2 (plagiarism)	2.07%
Strategy and Competition	3.04%

presented an overlap of only 106 word families with ABEL. While 703 were unique to ABEL, 387 were unique to Konstantakis' BWL. Examples of words common to both lists were *theft*, *lucrative* and *workforce*. Hsu's BWL showed an overlap of just 161 word families with ABEL. Unique to Hsu's BWL were 262 word families and unique to ABEL were 675. Some of the word families in common were *attorney*, *download* and *equity*. These results confirm the hypothesis that ABEL is indeed different from the two BWLs, and that there is room for another BE vocabulary list that is more suitable for undergraduate business students.

The last two research questions inquired about the coverage of ABEL against two other different corpora:

RQ 5: How well do the words in ABEL cover non-academic texts?

RQ 6: How well do the words in ABEL cover another academic business corpus?

AntWordProfiler was used to identify ABEL's coverage of the other two corpora. Results showed that ABEL covered only 0.69% of the non-academic corpora (classic English and American literature), while it covered 2.61% of the word families in the First year textbook corpus. These results confirm the hypothesis that ABEL would perform poorly in a non-academic corpus (0.69% is only roughly a quarter of the 2.83% coverage of ABEL corpus) and similarly when tested against a similar corpus (only 0.22% coverage difference between the two business corpora).

In this chapter the results related to the six research questions proposed in this study were presented. In the next section, these findings will be discussed in detail and interpretations will be provided.

Chapter five – Discussion

This study proposed the creation of an Academic Business English List (ABEL) in order to facilitate vocabulary learning among L2 speakers wishing to pursue their undergraduate studies in business in an English speaking university. In order to create such a list, a specialized business academic corpus was required. After the creation of the corpus, it was tested against a few frequency lists for coverage, then the remaining words were organized by range of appearance and those with more range were selected to be part of the ABEL list. The ABEL was then compared to two other BWLs and tested on two different corpora. Let us now discuss the results of each one of these steps in more detail.

Core frequency list

The first test done to examine the ABEL corpus was related to the identification of the core frequency list that would offer the highest coverage of ABEL. The results showed very similar results across the three sets of frequency lists tested: GSL+AWL, NGSL+NAWL and BNC/COCA. Surprisingly, the first set of lists (GSL+AWL) showed a very high coverage of the corpus (89.44%), even though it was still the lowest of the scores overall. This goes against the hypothesis that this set of lists would perform poorly. The reason for predicting that the GSL+AWL would not have a great coverage was because the GSL is a very old list (created in the 1920s-30s and published in 1953), and it does not contain modern words related to technology such as *computer*, *website* or *download,* which were fairly common in the ABEL corpus. The 15-year-old AWL also showed a good coverage of 10.83%, which is over the 10%

coverage of the general academic texts and below the 12% coverage of commerce texts in Coxhead (2000). The hypothesis was that the NGSL+NAWL would be the set of lists that would provide the greatest coverage of the ABEL corpus because it is a newer version of the old set and because it contains more word families (3,764 versus 2,570). Unfortunately the difference between both coverages was not great (only 0.34% more coverage for the newer list) and the NAWL showed a very poor performance of just 2.23% coverage. Another reason why it was hypothesized that the NGSL+NAWL set would be the best choice was because it included the NAWL, which has an academic focus, but the results showed that the NBC/COCA first 3K were more powerful in terms of coverage than the previous two lists, and the statistical tests confirmed that this difference was significant. The fourth thousand word family in the BNC/COCA was also considered; however, its coverage dropped so drastically compared to its previous thousand (from 9.31% for 3K to 1.75% for 4K), that this was seen as an indication that it would be better to have a specialized list from that point on; a list such as ABEL, which would offer more text coverage. The take away message here is that the GSL and AWL are still powerful lists that teachers should continue to use with general academic learners, and that materials and textbooks created based on them are still relevant to today's academic reality. In addition, even though a lot of time and effort was put into creating a new version of the old set, the NGSL and NAWL are not statistically better than its old counterpart. To conclude, it is also interesting to see how powerful the BNC/COCA 1k+2k+3k lists are, even though they are not specifically academic.

Range

Once the core English frequency list was determined to be the BNC/COCA, then the ABEL could be created. Special attention was given to the range of words to be included in the ABEL list and all words that occurred with a range of seven or more (which means they were present in seven or more of the 15 textbooks) were included in ABEL. Interesting to notice is that only three word families had a range of 15: *verify, minimize* and *identical*. Range 14 had 16 word families, range 13 had 43, and range 12 had 61. It was only from range 11 down that the number of word families increased substantially to be greater than 100 per range. This confirmed that the decision of choosing to include words for a range of about 50% of the sub-corpora (7 of the 15 books) was a good one because if a higher range (e.g., 11 of 15 books) was chosen, then there would be fewer words on the ABEL list, consequently offering a lower coverage of business texts. Exploring range revealed that very few words reoccur across all sub-corpora in a corpus as big as this one (excluding the core vocabulary). This means that in order to create a specialized frequency list we cannot set the range criterion too strictly; otherwise many useful word families would be left aside and coverage would be small.

ABEL's characteristics

Once the headwords were identified and the word families were created, it became possible to have a better idea of what an academic business English list looked like. ABEL is composed of 840 word families which cover 2.86% of the ABEL corpus. It is interesting to note that some word families have a very strong business flavour to them (e.g., *equity*, *GDP* and

marketplace) while others do not (e.g., *bedroom*, *grammar* and *soccer*). We also find non-business words like *television*, *lemonade* and *beer* in Konstantakis' (2007) BWL; and it happened to some extent with Hsu's (2011a) BWL, which has words such as *ambiguity*, *contemporary* and *hence* (such words seem to be more general academic than graduate business academic in character). This probably happened due to the methodology applied to creating these lists, which are based on range and/or number of occurrences of words in the corpora. That is to say, this technique of frequency list creation is made to include all word families that satisfy the researcher's methodology criteria, which usually involves including all words that reach a certain range of appearances in the sub-corpora and a minimum number of occurrences in the corpus as a whole.

A possible explanation for this non-business flavour phenomenon could be that Konstantakis' BWL included words with a very low range, as word families with a range of five and up out of a total of 33 sub-corpora were included in the list. This means that a range of approximately 14.28% of the sub-corpora was acceptable, while in the ABEL the minimum was a range of 7 out of 15 sub-corpora (or 46.67%). In order to increase coverage to 95%, Konstantakis also included all word families that had at least 10 occurrences in his corpus, regardless of their range. This certainly increased the coverage percentage, but it probably contributed to the inclusion of words which do not sound business related.

In order to be included in Hsu's BWL, a word had to occur at least 270 times in the corpus as a whole and in at least half of the 2,200 research articles included in her corpus. The high number of occurrences (270 times) was possible due to the size of Hsu's corpus (7.62 million tokens), while the 50% range in the sub-corpora was also inspired by Coxhead (2000). Hsu arrived at the 270-occurrence number after a few tests on her corpus that showed that at that

level of instances, the word families included in her list provided a coverage of 95% of the corpus (together with the BNC 3K and proper nouns). The stricter standard in the control for range and number of occurrences probably contributed to her list having a more business or academic character.

While range was also controlled for in the creation of ABEL (seven out of 15 sub-corpora, or 46.67%), the minimum number of occurrences was not. That means that word families with occurrences as low as eight in the entire corpus were also included in the list. This could be the reason why some words in the ABEL seem to be general English and not BE. Important to notice is that, even though some words do not seem related to business, they are also part of the corpus and students have to know them in order to reach the 95-98% comprehension level.

Coverage

Regarding the coverage of ABEL, it may seem at first that 2.86% is a rather low number, but it is actually fairly high. If we look at the coverage of the ABEL corpus as compared to the regular BNC/COCA up to the 25th K-level (Table 11), we can see that each 1K list after the first 3K offers a very low coverage of the corpus. More specifically, after the 5K, coverage drops to below 1% and keeps decreasing until K 12, where it increases slightly by 0.02%, but keeps decreasing again until the 0.00% level as of K 22. This means that in order to achieve additional coverage of 2.86% after the first 3K and using only the BNC/COCA, students would have to know between two and three thousand word families. The calculation goes as follows: one

Table 11

Coverage of the BNC/COCA from 1K to 25K

BNC/COCA lists	Coverage	Cumulative coverage
1K	67.10%	67.10%
2K	15.25%	82.35%
3K	9.31%	91.66%
4K	1.75%	93.41%
5K	0.84%	94.25%
6K	0.44%	94.69%
7K	0.30%	94.99%
8K	0.25%	95.24%
9K	0.11%	95.35%
10K	0.10%	95.45%
11K	0.07%	95.52%
12K	0.09%	95.61%
13K	0.04%	95.65%
14K	0.04%	95.69%
15K	0.02%	95.71%
16K	0.02%	95.73%
17K	0.01%	95.74%
18K	0.01%	95.75%
19K	0.01%	95.76%
20K	0.01%	95.77%

21K	0.01%	95.78%
22K	0.00%	95.78%
23K	0.00%	95.78%
24K	0.00%	95.78%
25K	0.00%	95.78%

thousand word families for 4K providing 1.75% coverage, plus one thousand word families for 5K (0.84%) providing a cumulative coverage of 2.59%, plus fewer than one thousand words from 6K (0.44%) which provides a cumulative coverage of 3.03%. But instead of learning approximately 2.5 thousand word families, students can learn 840 to achieve the same coverage. This shows how challenging it is to have high vocabulary coverages in specialized texts based on GE frequency lists only, and it also shows how a specialized list can be much more cost-efficient.

It was mentioned previously that the BNC/COCA 4K was not considered as part of the core list to be excluded before the creation of the ABEL due to its low coverage of the ABEL corpus compared to the first 3Ks. This was a somewhat arbitrary decision that proved to be correct. This becomes clear when we compare the coverage of ABEL and the BNC/COCA 4K. Pedagogically speaking, this means that instead of asking students to learn the one thousand words included in the BNC/COCA 4K to gain 1.75% coverage, it is suggested that they learn the 840 word families in ABEL to gain 2.86% text coverage. This logic refers back to the cost-effect principle mentioned in the literature review of this thesis.

Proper nouns

Another interesting finding in this study is the percentage of proper nouns found in the ABEL corpus. List 31 (the baseword list of proper nouns that accompanies the BNC/COCA) was used in order to exclude proper nouns from the counting. This list is composed of 21,662 words and was created by the BNC/COCA team. It includes personal names such as Mary and Bruce and geographical names such as Budapest and Norway. However, as described previously, this list alone was not enough to exclude all proper nouns from the corpus; therefore, another list of extras consisting of 291 words was created to exclude the remaining proper nouns (e.g. Apple, Google and Zappos). The results showed that the BNC/COCA list 31 covered 1.36% of the ABEL corpus and the list of extras covered 1.09%. Together they covered 2.45% of the corpus, which is more than the BNC/COCA 4K and almost as much as the ABEL. This finding indicates that dealing with proper nous while reading undergraduate business textbooks is common practice among business students, and indeed we can easily see this by glancing at the corpus and noticing the high number of company names, names of entrepreneurs and politicians, as well as names of places and geographical locations. These are all assumed to be lexically transparent and it is reasonable to expect that learners already know words like *Washington* and *London*. However, it is not clear if the knowledge of company names such as *Zappos* and *Wal-Mart* is also transparent, or if knowing something about the nature of these companies may be important for full comprehension of a text. How various kinds of proper nouns function in understanding a business text is an interesting question for future research.

Coverage per sub-corpus

When we look at the coverage of the ABEL per sub-corpus (Table 10), we can see that its coverage varies significantly from the smallest amount of coverage (1.83% for Contemporary Business Thinking) to the largest (3.62% for the Financial Accounting). We can also see that the coverage is somewhat proportionate to the size of the books, which means that smaller textbooks present less coverage of ABEL than larger ones. This could be seen as a limitation of this list because it seems more representative of some business sub-areas than others. However, it is important to remember that the goal of ABEL is to be a core list for academic BE in general, not a specialized list that deals with a single business sub-field. That is to say, its goal is to serve BE teachers and students that are interested in an introduction to BE, and ABEL provides such an introduction because of its manageable size (840 word families) and because of its 50% range.

ABEL and the BWLs

As mentioned in the result section, the word families in ABEL differ from the word families in both BWLs, showing a relatively low rate of overlap. This confirms the hypothesis that ABEL would be different from the two BWLs due to the different corpus design used to create the ABEL corpus. Also worth noticing is the fact that these three frequency lists were built excluding different sets of core frequency lists. Konstantakis' BWL excluded the word families in the GSL and the AWL, while Hsu's BWL excluded words from the BNC 3K, and the ABEL excluded words from the BNC/COCA 3K. These two factors alone could indicate that the lists would be different; however, there was yet another factor that made it even more evident that the

lists would differ: their sizes. Konstantakis' BWL has 560 word families, Hsu's has 426, and the ABEL has 840. The difference among these lists could indicate that the ABEL can contribute positively to the vocabulary research in BE as it may be more suited for undergraduate business students than the BWLs.

ABEL and other corpora

Another factor that showed evidence of the quality of ABEL is its performance when tested against other corpora. The ABEL coverage was very low when tested against the non-academic corpora (only 0.69%), while it showed an equivalent coverage when tested on a similar corpus (2.61% with the First year textbook corpus versus 2.86% with ABEL). This confirmed the hypothesis that ABEL would perform poorly in the first case and similarly in the second. This constitutes further evidence for the contribution of ABEL to BE vocabulary research.

Chapter six – Implications and Conclusions

In summary, this study proposed six research questions that explored the main characteristics of an intact undergraduate core business corpus with the aim of extracting from it a frequency list. In the process, other secondary research questions were proposed and answered, and those results provide a better understanding of the academic BE vocabulary field that is not limited to the frequency list per se. For instance, in RQ 2 three different sets of core English frequency list were tested for coverage, which allowed the researcher to explore the behaviour of basic vocabulary inside the academic BE field. In addition, the exclusion of proper nouns allowed us to realize how extensively these words are used in the field. It was also possible to observe the behaviour of the ABEL in two different corpora (the non-academic and the business corpora). Furthermore, it was interesting to notice that not all word families in ABEL are clearly related to business. All this information helps us understand the ABEL, its coverage and its content.

To summarize the main finding of this study, we can say that the ABEL was successfully created, that it is different from the two existing BWLs, and that when tested against other corpora it showed statistically positive results. Moreover, ABEL was shown to have more coverage than the BNC/COCA 4K and 5K combined, even though it has fewer than half the number of word families, which indicates that learning those words is more cost effective for learners. In addition, the coverage of the ABEL, together with the BNC/COCA 3K and proper nouns, reached 96.94%, which is between the 95% minimal coverage and the 98% optimal coverage suggested by Laufer and Ravenhorst-Kalovski (2010). This coverage is also higher than the two other BWLs, which set their coverage goal to be 95% (together with their relative

core frequency list and excluding nouns). These results indicate that the ABEL makes a positive contribution to the academic BE vocabulary research.

Research implications

One of the most important research contributions in this study is the establishment of the BNC/COCA 3K as the most powerful core English frequency list in terms of having the highest coverage of GE words in the ABEL. This simple exercise revealed interesting results which favoured the newer list as opposed to the more traditional combination GSL 2K plus AWL or its newer version. This is counter intuitive to most ESL practice, which tends to favour the traditional combination GSL + AWL. Therefore, even though the coverages of the GSL 2 K + AWL and the NGSL 2K + NAWL were high, the coverage power of the BNC/COCA proved to be even better. Future research is needed to determine whether this advantage is also found in textbook corpora for other academic disciplines.

Another interesting contribution regards the coverage of ABEL throughout the 15 sub-corpora. Results indicate that Chen and Ge (2007), Hyland (2008) and Hyland and Tse's (2007) hypothesis regarding general academic lists versus specialized lists was incorrect. It was mentioned in the literature review that these researchers criticized Coxhead's (2000) AWL because of its generic nature. They did so by showing that the coverage of this frequency list behaves very differently depending on the discipline, and they suggested that a specialized list should be proposed right after the core 2K or 3K of GE. As mentioned in the discussion section, the coverage of ABEL is also unstable and varies from one sub-corpus to another, even though the ABEL is a specialized list of the type Chen and Ge (2007), Hyland (2008) and Hyland and

Tse (2007) expected to be more stable. That is, their hypothesis regarding the more stable coverage of specialized lists does not materialize. This finding indicates that further research is needed to explore the behaviour of specialized lists.

Another interesting finding is the fact that not all words in the ABEL have a business connection. This suggests that even though the BNC/COCA 3Ks were excluded, other words with a general characteristic (such as *alike, DVD* and *photocopy*) were not. In addition, many words which do have a business flavour can be found in the BNC/COCA 3Ks. For instance, the words *company, market* and *manage* can be found in the BNC/COCA 1K; the words *product, customer* and *finance* can be found in the BNC/COCA 2K; and finally the words *strategy, invest* and *manufacture,* among others, can be found in the BNC/COCA 3K. This could indicate that in order to have a list consisting exclusively of words with business characteristics, another frequency list creation technique should be favoured. Another approach could involve emphasizing the business meanings of general words in learning materials for BE students.

Pedagogical implications

The main pedagogical implication of this study is that teachers of business students would do well to focus their vocabulary teaching to the combination of the BNC/COCA 3Ks and ABEL, as this combination has shown to provide high coverage of business texts. When we look at the coverage as a whole, we can see that the BNC/COCA 3K lists, plus both extra nouns lists, plus ABEL cover 96.94% of the corpus. As mentioned previously, this is more than the 95% minimal coverage suggested by Laufer and Ravenhorst-Kalovski (2010), but less than the 98% optimal coverage. Pedagogically speaking, this means that with this vocabulary students can read

a text and have good understanding of it, but to read it without the assistance of a dictionary (being able to guess by context), they still need to know word families that would increase coverage by 1.06%. This leads us to the conclusion that teachers should do all they can to increase students' knowledge of such words, and this includes using all vocabulary teaching technique possible (e.g. paper or electronic flash cards, vocabulary games, extensive reading programs, etc.). Regarding the vocabulary left to reach 98%-100% comprehension of words in a text, teachers can teach students strategies to become autonomous learners. A few examples of this include, but are not limited to, teaching them to use dictionaries, encouraging them to keep vocabulary diaries, and showing them how to guess the meaning by the context.

ABEL can also have a few other pedagogical uses. First, the list could be used by ESL material creators when designing BE textbooks to concentrate their efforts on more cost-effective words instead of intuitively adding vocabulary in their books. The same logic can be applied when developing testing services such as the Business Language Test (BULATS). Second, the list could be used by teacher trainers interested in training BE teachers or by novice teachers who are studying on their own and eager to learn this specialized lexicon. Third, Computer Assisted Language Learning (CALL) creators could also use the list to feed their programs and have a special section dedicated to BE vocabulary. Fourth, universities could use ABEL in the admissions process to measure the BE vocabulary size of prospective students. Fifth, companies will be able to use it as part of the hiring process (language test) if they are concerned about the language level of their future employees. Lastly, L2 students in a business program at the undergraduate level will be able to use the list to improve their vocabulary size.

Limitations

This study has two main limitations: a pedagogical and a methodological one. Let us take a look at them in details.

Pedagogical limitations

This study concentrated on investigating which individual word families are more frequent in authentic undergraduate BE materials. Therefore, it did not take into consideration whether the words were part of a phrasal verb, a collocation or a chunk. For instance, the word *"meeting"* was analyzed individually and the words surrounding it to create collocations such as *"run a meeting"* or *"call a meeting"* were disregarded. Even though these items are not negligible in a corpus, the scope of the research would be too broad if they were to be included. However, they are definitely a great theme for future research.

Another pedagogical limitation is the fact that even with knowledge of the BNC/COCA 1-3k, the ABEL and the proper nouns, we still don't achieve the full 98% coverage necessary for optimal comprehension. Therefore, we can say that with ABEL students have good conditions for comprehension but not perfect conditions.

In addition, even when learners do know all the words, research shows that they still do not necessarily comprehend everything perfectly. For instance, in the Schmitt et al (2011) paper, even for those participants who knew 100% of the vocabulary in a text, the mean comprehension test score was still only 75%. This is most probably linked to the first limitation in this section, which mentioned more complex uses of words (phrasal verbs, collocations and chunks) and the impact that has on their meaning.

Furthermore, it is important to mention that this approach to vocabulary learning (which involves learning vocabulary in an orderly manner), only works if students actually know all of the 1-3k word families in the BNC/COCA and all of the ABEL words. However, it is unclear if L2 undergraduate business students actually master this vocabulary; therefore, this may be a very ambitious goal taking into consideration that many ESL classes do not focus much on vocabulary.

Methodological limitations

As in any other corpus study that builds on a previous one, the credibility of this research depends directly on the accuracy of the work that was done by others. In other words, as ABEL was built upon the BNC/COCA 3Ks, its accuracy also depends on the quality of the work done when creating these lists. The same logic applies to all software used to create and test ABEL.

Another limitation of this study regards the typos in the ABEL corpus. As mentioned in the methodology section, the textbooks were manually scanned and the OCR ABBY FineReader Express was used to transform the PDFs into .txt files. Despite the fact that the researcher spent a great deal of time manually editing the typos from the OCR, it was finally not possible to eliminate 100% of them. Instances such as tecnnology (instead of *technology)* and *offinancial* (instead of *of financial)* were not picked up by standard spellchecking software. This certainly affects the results of the coverages as the words with typos were classified as not belonging to any list. If it had been possible to correct all of the several thousand problem words and classify them, the combined coverage percentage of the core lists and ABEL would have almost certainly been higher than the 96.94% figure that was found.

To conclude, the ABEL can contribute to business vocabulary knowledge and to BE as a whole. It is also a first step towards helping (novice) teachers (as the ones described in the beginning of this thesis) master this field of study instead of fear it.

References

Andrade, M. S. (2006). International students in English-speaking universities adjustment factors. *Journal of Research in International Education, 5*(2), 131-154.

Anthony, L. (2014). AntWordProfiler (Version 1.4.1) [Computer Software]. Tokyo, Japan: Waseda University. Available from http://www.laurenceanthony.net/

Bernhardt, E. B. (2011). *Understanding advanced second language reading*. New York: Routledge.

Browne, C., Culligan, B., & Phillips, J. (2013). A New General Service List. Retrieved from http://www.newgeneralservicelist.org/

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.

Cobb, T. (2011). The compleat lexical tutor (version 6.2) [Software]. Retrieved from http://www.lextutor.ca/

Concordia University. (2015). John Molson School of Business core courses. Retrieved from: (https://www.concordia.ca/jmsb/programs/undergraduate/bachelor/program-structure/core-courses.html)

Coxhead, A. (2000). A new academic word list. *TESOL Quarterly, 34*, 213-239.

Dictionary. (2014). In *The Free Dictionary*. Retrieved from http://www.thefreedictionary.com/

Dignen, S. (Ed.). (2000). *Longman business English dictionary*. Harlow: Addison Wesley Longman.

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

Ellis, M., & Johnson, C. (1994). *Teaching business English: An introduction to business English for language teachers, trainers, and course organizers.* Oxford: Oxford University Press.

Halliday, M., Matthiessen, C. M., & Matthiessen, C. (2014). *An introduction to functional grammar.* New York: Routledge.

Hirsh, D., & Nation, P. (1992). What vocabulary size is needed to read unsimplified texts for pleasure? *Reading in a Foreign Language, 8*(2) 689-96.

Hsu, W. (2011a). A business world list for prospective EFL business postgraduates. *The Asian ESP Journal, 7*(4) 63-99.

Hsu, W. (2011b). The vocabulary thresholds of business textbooks and business research articles for EFL learners. *English for Specific Purposes*, *30*(4), 247-257.

Hyland, K. (2008). The author replies. *TESOL Quarterly, 42*(1), 113-114.

Hyland, K., & Tse, P. (2007). Is there an "Academic Vocabulary"? *TESOL Quarterly, 41*(2), 235-253.

Konstantakis, N. (2007). Creating a business word list for teaching business English. *Estudios De Linguistica Inglesa Aplicada (ELIA)*, 7, 79-102.

Laufer, B. (1989). What percentage of text-lexis is essential for comprehension? In C. Lauren & M. Nordmann (Eds.), *Special language: from humans thinking to thinking machines* (pp. 316-323). Clevedon: Multilingual Matters.

Laufer, B., & Ravenhorst-Kalovski, G.C. (2010). Lexical threshold revisited: lexical text coverage, learners' vocabulary size and reading comprehension. *Reading in a Foreign Language, 22*(1), 15-30.

McGill University. (2015). Faculty of Management, required core courses. Retrieved from: http://www.mcgill.ca/desautels/programs/bcom/academics/programstructure/90-credit-program

Nation, I. S. P. (2006). How large a vocabulary is needed for reading and listening? *Canadian Modern Language Review, 63*(1), 59-82.

Nation, I. S. P. (2012). Notes on the BNC/COCA lists. Wellington: New Zealand: Victoria University of Wellington. Retrieved from http:// www.victoria.ac.nz/lals/staff/paul-nation.aspx

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

Nation, I. S. P. & Heatley, A. (2002). RANGE [Computer software]. Retrieved March 19th, 2015, from http://www.vuw.ac.nz/lals/staff/paul-nation/nation.aspx

Nation, I. S. P., & Macalister, J. (2009). *Language curriculum design.* New York: Routledge.

Nelson, M. (2010). Building a written corpus: what are the basics? In O'Keeffe, A. & McCarthy, M. (Eds.), *The Routledge Handbook of Corpus Linguistics* (p.53-65). New York, NY: Routledge.

OECD. (2011). Education at a Glance 2011, Paris: OECD. Retrieved from http://www.oecd.org/edu/skills-beyond-school/48631079.pdf

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

Parkinson, D. (Ed.). (2005). *Oxford business English dictionary: for learners of English*. Oxford: Oxford University Press.

Parry, K. (1997). Vocabulary and comprehension: Two portraits. In J. Coady & T. Huckin (Eds.), *Second language vocabulary* (pp. 55-68). Cambridge: Cambridge University Press.

Project Gutenberg. (2015). Free e-books by Project Gutenberg. Retrieved from: https://www.gutenberg.org/

Raymond, P. M., & Parks, S. (2002). Transitions: Orienting to reading and writing assignments in EAP and MBA contexts. *Canadian Modern Language Review/La Revue canadienne des langues vivantes, 59*(1), 152-180.

Reppen, R. (2010). Building a corpus: what are the key considerations? In O'Keeffe, A. & McCarthy, M. (Eds.), *The Routledge Handbook of Corpus Linguistics* (p.31-37). New York, NY: Routledge.

Schmitt, N., Jiang, X., & Grabe, W. (2011). The percentage of words known in a text and reading comprehension. *Modern Language Journal. 95*(1), 26-43.

Sherry, M., Thomas, P., & Chui, W. H. (2010). International students: A vulnerable student population. *Higher Education, 60*(1), 33-46.

Stæhr, L. S. (2008). Vocabulary size and the skills of listening, reading and writing. *Language Learning Journal, 36*, 139-152.

Sutarsyah, C., Nation, P., & Kennedy, G. (1994). How useful is EAP vocabulary for ESP? *RELC Journal, 25*(2), 34-50.

UNESCO. (2014a). Education at a Glance 2014. OECD indicators. Retrieved from http://www.oecd.org/edu/Education-at-a-Glance-2014.pdf

UNESCO. (2014b). Institute for Statistics. Global Flow of Tertiary-Level Students. Retrieved from http://www.uis.unesco.org/Education/Pages/international-student-flow-viz.aspx

University of Toronto. (2015). Management home page. Retrieved from: http://www.utsc.utoronto.ca/~registrar/calendars/calendar/Management.html

Van Zeeland, H., & Schmitt, N. (2012). Lexical coverage and L2 listening comprehension: How much does vocabulary knowledge contribute to understanding spoken language? *Applied Linguistics, 34*(4) 457–479.

VassarStats. (2015). Website for statistical computation. Retrieved from http://vassarstats.net/

West, M. (1953). *A General Service List of English Words*. London: Longman, Green & Co.

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

Zhang, Z. (2013). Business English students learning to write for international business: What do international business practitioners have to say about their texts. *English for Specific Purposes*, 32(3), 144-156.

Appendix 1

List of textbooks used to create the ABEL corpus

COMM 210 Contemporary Business Thinking	Dyer, L. (2011). *Critical Thinking for Business Students (2nd ed.)*. Concord, Ontario: Captus Press. 119 pages
COMM 212 Business Communication	Meyer, C. M. (2014). *Communicating for results: A Canadian student's guide (3rd ed.)*. Don Mills, Ontario: Oxford University Press. 540 pages
COMM 215 Business Statistics	Groebner, D. F., Shannon, P. W., Fry, P. C., & Smith, K. D. (2014). *Business statistics: a decision making approach: Custom edition for Concordia University*. Boston, MA: Prentice Hall/Pearson. 884 pages
COMM 217 Financial Accounting	Libby, R. & Libby, P.A. (2014). *Financial Accounting (5th Canadian ed.).* Whitby, Ontario: McGraw-Hill Ryerson. 746 pages
COMM 220 Analysis of Markets	Riley, J. (2014). *Analysis of Markets (course pack). Montreal: Concordia University Bookstore.* 251 pages
COMM 222 Organizational Behaviour and Theory	Johns, G. & Sacks, A.M. (2014). *Organizational behaviour: Understanding and managing life at work (9th ed.)*. Toronto: Pearson Canada.

	659 PAGES
COMM 223	Kotler, P. & Armstrong, G., (2012). *Principles of Marketing*
Marketing	*(14th ed.).* New Jersey: Prentice Hall.
Management I	613 pages
COMM 225	Stevenson, W. J., Hojati, M. & Jacobs, F.R. (2013). *COMM*
Productions/Operations	*225 production and operations management: custom*
Management	*publication for Concordia University.* Toronto: McGraw-Hill
	Education Custom.
	719 pages
COMM 226	Kroenke, D.M., McKinney, E. H., Gemino, A. & Tingling,
Business Technology	P.M. (2015). *COMM 226 business technology management:*
Management	*second custom edition for Concordia University.* Boston, MA:
	Pearson Learning Solutions.
	460 pages
COMM 305	Weygandt, J. J., Kimmel, P. D., Kieso, D. E. & Aly, I.M.
Managerial Accounting	(2012). *Managerial accounting: Tools for business decision*
	making (3rd Canadian ed.). Etobicoke, Ont: John Wiley &
	Sons.
	624 pages
COMM 308	Ross, S.A., Westerfield, R.W., Jordan, B.D. & Roberts, G.S.
Introduction to Finance	(2013). *Fundamentals of corporate finance (8th Canadian ed.).*
	Whitby, Ontario: McGraw-Hill Ryerson
	780 pages

COMM 315 Business Law and Ethics	Crooks, F. (2013). *The fundamentals of Quebec business law and ethics.* Montreal: Paladin Publications. 386 pages
COMM 320 Entrepreneurship	Barringer, B.R. (2012). *Entrepreneurship: Successfully Launching New Ventures*. Boston : Pearson/Prentice Hall. 555 pages and Stern, L. (2009). *What every student should know about avoiding plagiarism*. New York : Pearson/Longman. 74 pages
COMM 401 Strategy and Competition	Hitt, M., Ireland, R. D., & Hoskisson, R. (2012). *Strategic management cases: competitiveness and globalization (9th ed.)*. Australia: Cengage Learning. 402 pages

Appendix 2 – The ABEL

AB
ABBREVIATE
ABC
ABIDE
ABSENT
ABUNDANT
AC
ACCESSORY
ACCORD
ACCOUNTANT
ACCRUE
ACCUSTOM
ADJACENT
ADMINISTER
ADVENT
ADVERSE
AEROSPACE
AGGREGATE
AIL
AIRBUS
AIRPLANE
AIRPORT
AL
ALIGN
ALIKE
ALTERNATE
ALUMINIUM
AMBIGUITY
AMBIGUOUS
AMBITIOUS
AMENITY
AMID
AMPLE
ANALOGOUS
ANALOGY
ANALYTIC
ANCILLARY
ANIMATE
ANONYMOUS
ANTI
AP
APPAREL
APPENDIX
APPLIANCE
APPLICABLE
APPLICANT
APPRAISE
APT
ARENA
ARRAY
ARROW
ARTICULATE
ARTIFICIAL
ASPIRE
ASPX
ATM
ATTAIN
ATTORNEY
AUCTION
AUDIO
AUGMENT
AUTOMOBILE
AUTOMAKER
AUTOMATE
AUTOMOTIVE
AUTONOMY
AVERSE
AVIATION
AWAIT
AXIS
B4
BACHELOR
BACKUP
BANDWAGON
BANKRUPT
BANNER
BARREL
BASEBALL
BASKETBALL
BEDROOM
BEHALF
BENCHMARK
BENEFICIAL
BETA
BEVERAGE
BEWARE
BICYCLE
BIRTHDAY
BLACKBOARD
BLOG
BOLD
BOLT
BONUS
BOOKING
BOOKSTORE
BOOMER
BOXED
BRACKET
BRAINSTORM
BREACH
BREADTH
BREAKDOWN
BREAKTHROUGH
BRIBE
BROCHURE
BROKER
BROWSE
BUBBLE
BULK
BULLET
BUREAUCRAT
BUSINESSPERSON
BUSINESSMAN
BUSINESSWOMAN
CALENDAR
CAMPUS
CANDY

CAPITA
CAPITALISE
CARDBOARD
CAROL
CASCADE
CATALYST
CATER
CAUSAL
CAUTION
CAUTIOUS
CD
CELEBRITY
CELLULAR
CENSUS
CEO
CERTIFICATE
CERTIFY
CF
CFO
CHAOTIC
CHARITABLE
CHECKOUT
CHEF
CHEQUE
CI
CIVIC
CLASSIFICATION
CLASSIFY
CLASSMATE
CLASSROOM
CLERICAL
CLERK
CLICK
CLIENTELE
CLOSURE
COGNITIVE
COLA
COLLAR
COLLATERAL
COMMODITY
COMMONPLACE
COMPACT
COMPARATIVE
COMPATIBLE
COMPEL
COMPETENCE
COMPILE
COMPLEMENT
COMPLIANCE
COMPLY
COMPOSITE
CONCEAL
CONCESSION
CONCISE
CONFIDENTIAL
CONFORM
CONGLOMERATE
CONJUNCTION
CON
CONSECUTIVE
CONSENSUS
CONSISTENCY
CONSOLIDATE
CONTEMPLATE
CONTINGENT
CONTRACTUAL
CONTRARY
CONVENIENCE
CONVENIENT
CONVERSELY
CONVERSION
COO
COOKIE
COPPER
COPYRIGHT
CORN
COSMETIC
COUNTERFEIT
COUNTERPART
COUPON
COURTESY
CPA
CREDENTIAL
CREDIBLE
CRITIQUE
CRUDE
CRUNCH
CS
CUMULATIVE
CUSHION
CUSTOMIZE
DAIRY
DC
DEADLINE
DEAN
DEBIT
DECENTRALIZE
DECIMAL
DEDUCT
DEEM
DEFAULT
DEFECT
DEFER
DEFINITIVE
DELETE
DELTA
DEMOGRAPHY
DENOMINATOR
DENTAL
DEPARTURE
DEPENDENCE
DEPLOY
DEPOT
DEPRECIATE
DEREGULATE
DESKTOP
DESTINATION
DETRIMENT
DEVIATE
DEVISE
DIAGNOSIS
DIAGNOSTIC
DIAGRAM
DIFFERENTIAL
DIFFERENTIATE
DIGIT
DIGNITY
DILEMMA
DILIGENT

DIMINISH
DISADVANTAGE
DISCARD
DISCLOSE
DISCLOSURE
DISCONTINUE
DISCOURAGE
DISCREPANCY
DISGRUNTLED
DISPOSITION
DISSATISFY
DISSEMINATE
DISTORT
DISTRACT
DIVERSIFY
DIVIDEND
DOMAIN
DOMINANCE
DOMINION
DOUGHNUT
DOT
DOWNLOAD
DOWNSIDE
DOWNSIZE
DOWNTIME
DOWNTOWN
DOWNTURN
DRAWBACK
DUAL
DUPLICATE
DURABLE
DURATION
DVD
DYNAMIC
ELIGIBLE
EMBARK
EMBED
EMPIRICAL
EMPOWER
EN
ENACT
ENCOMPASS
ENRICH
ENROL
ENTAIL
ENTITY
ENTRANT
ENTREPRENEUR
ENVISION
EQUITABLE
EQUITY
ERR
ERRONEOUS
ESSENCE
ET
ETHICAL
ETHICS
EURO
EVENLY
EX
EXCEL
EXCERPT
EXCLUSION
EXIT
EXPERTISE
EXPIRE
FACTUAL
FAIRS
FAKE
FARE
FATAL
FATIGUE
FAX
FEASIBLE
FEDERATE
FEEDBACK
FICTITIOUS
FINED
FINITE
FISCAL
FLAW
FLEET
FLUCTUATE
FLUID
FM
FOOTWEAR
FORE
FORESEE
FOREVER
FORT
FORUM
FRACTION
FRANCHISE
FRAUD
FRAUDULENT
FREIGHT
FRICTION
FRONTIER
FURNISH
GAMBLE
GARBAGE
GARMENT
GASOLINE
GATEWAY
GAUGE
GDP
GENERATOR
GENRE
GENIUS
GLOBE
GLOSS
GOODWILL
GOURMET
GOVERNANCE
GPS
GRAMMAR
GRAPH
GRID
GRILL
GURU
HALFWAY
HAMBURGER
HANDBOOK
HANDHELD
HANDWRITE
HARDWARE
HARNESS
HARVEST
HAUL

HEADACHE
HEADLINE
HEALTHCARE
HEIGHTEN
HIERARCHY
HINDSIGHT
HO
HOC
HOCKEY
HOMEPAGE
HOMEWORK
HOMOGENEOUS
HORIZON
HORIZONTAL
HTM
HTML
HTTP
HUB
HURDLE
HYBRID
HYPOTHETICAL
IC
ICON
ID
IDENTICAL
IDLE
IMPERATIVE
IMPERSONAL
IMPLICIT
INADVERTENT
INCEPTION
INCLINE
INCLUSION
INCOMING
INCREMENT
INCUR
INDIFFERENT
INDIRECT
INDUCE
INFER
INFERIOR
INFORMAL
INFORMATIVE
INFRASTRUCTURE
INHERENT
INK
INNOVATIVE
INS
INTEGRAL
INTEGRITY
INTELLIGENT
INTER
INTERCHANGE
INTERDEPENDENT
INTERFACE
INTERMEDIARY
INTERMEDIATE
INTERN
INTERSECT
INTERVAL
INTUITION
INTUITIVE
INVENTORY
INVOICE
IPAD
IPO
IPOD
IQ
IR
ITALIC
JARGON
JAY
JERSEY
JUNK
KEYBOARD
KEYWORD
KIT
LAG
LAPTOP
LASER
LAWSUIT
LAYOFF
LAYOUT
LB
LEAF
LEGACY
LEISURE
LENS
LEVER
LIFESTYLE
LIFETIME
LIKEWISE
LIME
LINEAR
LITIGATE
LLC
LOGISTICS
LOTTERY
LOUNGE
LUCRATIVE
LUGGAGE
LUMBER
MA
MAGNITUDE
MAINSTREAM
MALL
MANDATORY
MAPLE
MAR
MARKETPLACE
MARKUP
MARSHAL
MATRIX
MAXIMISE
MBA
MEDIAN
MEDICATION
MEMO
MEMORABLE
MENTOR
MERCHANDISE
MERCHANT
MERIT
METRO
METROPOLITAN
MICRO
MID
MINERAL
MINI

MINIMAL
ML
MM
MOMENTUM
MONETARY
MONOPOLY
MORALE
MOTORCYCLE
MS
MULTI
MULTINATIONAL
MULTIPLY
MUNICIPAL
NASDAQ
NAVIGATE
NECESSITY
NEWSLETTER
NICHE
NICK
NOMINAL
NONETHELESS
NORM
NOTATE
NOTEBOOK
NOTEWORTHY
NOTIFY
NOVELTY
NUMERIC
NUTRITION
OBSOLESCENCE
OBSOLETE
OBSTACLE
OFFSET
OMISSION
OMIT
ONGOING
ONLINE
OPTIC
OPTIMAL
OPTIMIST
OPTIMISE
ORE
ORG
OUNCE
OUTDATED
OUTGOING
OUTLET
OUTLOOK
OUTRIGHT
OUTSET
OUTSOURCE
OUTSTANDING
OUTWEIGH
OVERESTIMATE
OVERHAUL
OVERHEAD
OVERNIGHT
OVERSEE
OVERSIGHT
OVERTIME
OVERVIEW
PA
PAPERWORK
PAR
PASSIVE
PASSWORD
PATENT
PATIENCE
PAYOFF
PAYROLL
PC
PDF
PENDING
PERIPHERY
PERK
PERPETUAL
PERSONALISE
PERSUASION
PERTAIN
PERTINENT
PERVASIVE
PESSIMISTIC
PETROLEUM
PH
PHARMACEUTICAL
PHARMACY
PHOTOCOPY
PHYSICIAN
PI
PICKUP
PIONEER
PIPELINE
PITFALL
PIZZA
PLEDGE
PLUMBING
PM
PORTABLE
PORTFOLIO
POSTER
POSTPONE
POWERPOINT
PR
PRE
PRECIOUS
PRECLUDE
PREDETERMINE
PREDOMINANT
PRELIMINARY
PREMIER
PREMIUM
PRESCRIBE
PRESCRIPTION
PRESTIGE
PREVAIL
PREVALENT
PREVIEW
PROACTIVE
PROCURE
PROPONENT
PROPOSITION
PROPRIETOR
PROTOCOL
PROTOTYPE
PROXY
PRUDENT
PUBLICITY
PURSUIT
QUALITATIVE

QUIZ
QUO
QUOTA
RACK
RAILWAY
RAND
RATIONALE
RECEIPT
RECEPTIVE
RECIPIENT
RECIPROCAL
RECONCILE
RECOURSE
RECREATION
RECUR
RECYCLE
REDEEM
REDESIGN
REED
REFINE
REFRAIN
REFUND
REGISTRAR
REGRESS
REIMBURSE
RELEVANCE
REMAINDER
REPAY
REPETITIVE
REPLICATE
REPUTABLE
REQUISITION
RESIDUE
RETENTION
RETRIEVE
RIGID
RIGOUR
ROBIN
ROBOT
ROBUST
ROM
RUBBER
SAFEGUARD
SALESPERSON
SALEMAN
SALEWOMAN
SALON
SATISFACTORY
SAVVY
SCARCE
SCENARIO
SCRAP
SCRUTINY
SEAM
SEASONING
SELDOM
SEMI
SEMINAR
SERIAL
SETUP
SHAREHOLDER
SHEER
SHORTFALL
SHRINK
SHUTDOWN
SHUTTLE
SIC
SIGNIFY
SILICON
SIMPLICITY
SINCERE
SCEPTIC
SKIM
SKIP
SLOGAN
SLOWDOWN
SMARTPHONE
SNACK
SNAPSHOT
SOCCER
SOLICIT
SOPHISTICATION
SOUTHWEST
SPAM
SPAN
SPECTRUM
SPITE
SPOUSE
SPREADSHEET
STACK
STAKEHOLDER
STANDPOINT
STARK
STATIC
STATIONER
STEEP
STEREO
STOCKHOLDER
STRAIGHTFORWARD
STRIVE
SUB
SUBJECTED
SUBJECTIVE
SUBMISSION
SUBSCRIBE
SUBSIDIARY
SUBSTANTIVE
SUBTRACT
SUED
SUNGLASS
SUPERMARKET
SURGE
SURPLUS
SUSCEPTIBLE
SUV
SWIFT
SYNERGY
TABLET
TAG
TAILOR
TANGIBLE
TECHNICIAN
TELECOM
TELLER
TEMPLATE
TENTATIVE
TENURE
TERMINATE
TERMINOLOGY

TEXTBOOK
TEXTILE
TEXTING
THEFT
THEREAFTER
THEREOF
THRESHOLD
THRIVE
THUMB
TIER
TOBACCO
TOLERANCE
TOLL
TRACTOR
TRADEMARK
TRAILER
TRANS
TRANSIT
TRANSPARENCY
TRANSPARENT
TRASH
TRIE
TRIPLE
TRIVIAL
TRUSTWORTHY
WORKSHOP
WORLDWIDE
WORTHWHILE
TUITION
TURNAROUND
TURNOVER
TYPED
UN
UNDERESTIMATE
UNDERGRADUATE
UNDERGROUND
UNDERWAY
UNIFY
UPCOMING
UPGRADE
UPSCALE
UPWARDS
USAGE
UTILISE
VACANT
VACATION
VACUUM
VALVE
VEND
VENDOR
VENUE
VERBAL
VERIFY
VERSA
VERTICAL
VIABLE
VIEWPOINT
VISA
VOLATILE
WARD
WAREHOUSE
WARRANT
WARRANTY
WARREN
WARY
WEBSITE
WEEKEND
WELLS
WHEAT
WHEREBY
WHOLESALE
WIDESPREAD
WIRELESS
WORKFLOW
WORKFORCE
WORKLOAD
WORKPLACE
WORKSHEET