

Measuring the Vocabulary of College General English Textbooks and English-medium Textbooks of Business Core Courses

Wenhua Hsu

(whh@isu.edu.tw)

I-Shou University, Taiwan

Abstract

This research aims to create two corpora, one for college General English (GE) textbooks used in Taiwan, and the other for English-medium textbooks for business core courses, to form a basis of comparison. The operational measures for analysis involved vocabulary size, vocabulary levels (distribution among the British National Corpus 1st–14th 1,000 high-frequency word families) and text coverage. Coxhead's (2000) Academic Word List containing 570 word families (AWL 570) was chosen as one of the base word lists. The two corpora were lexically compared using Nation and Heatley's (2002) RANGE software. The results show that approximately 49 to 415 interdisciplinary academic words can be learned from a GE reading textbook as opposed to 421–537 academic words from a business textbook. The business textbooks used a smaller vocabulary than the GE textbooks (the former converging at the 4,000–5,000 word levels versus the latter spreading among the 3,000–13,000 word levels). Beyond the top 2,000-word level and the AWL 570, a GE textbook can supply students with 190 to 1,327 new word families. It is hoped that the indices examined in this study will help English teachers to take into account the continuity of curriculum design while preparing General English and Business English teaching materials.

1 Background

English is not an official language in Taiwan. Taiwan's Ministry of Education referred to the English program design used by high schools in Asian countries and published a 1,000-basic-English-word list and a 2,000-English-word list in 2003. Since then, the 1,000- and the 2,000-high-frequency-word lists of English have served as a curricular standard for the English course design for junior high schools (Grades 7–9) and senior high schools (Grades 10–12). Therefore, the 2,000 lexical items are generally presumed to be the vocabulary of EFL senior high school graduates entering university.

At the tertiary level, English is a required language. English courses for general purposes (EGP), also called General English courses (GE) are offered to non-English majors two to three hours per week in the freshman and sophomore years respectively. The content of GE texts is broadly humanities-based covering topics such as culture, nature, geography, health and so on. Its teaching, provided with versatile academic content, is geared towards the general interest of students rather than to specific purposes, in order to broaden students' horizon so that they can achieve an all-encompassing development of knowledge.

In addition to the General English subject as one of the required courses, students have to take some specialist core courses based on the requirements of their departments. Many college

teachers in Taiwan hold a Ph.D. degree in a subject field from an English-speaking country. They tend to use specialist textbooks in English because English-medium textbooks often provide the most up-to-date information in various fields. As such, it is very common to see non-English course teachers use English-medium textbooks and deliver specialist knowledge in either the target language (English) or Taiwan's native language (Chinese).

After taking the required English courses in the first two years of college (namely, when formal English education is completed), maintaining non-English majors' English proficiency may rely on reading English-medium specialist textbooks and materials in content area courses offered by their departments.

In view of the above, we would like to explore the vocabulary size as well as levels involved in the GE textbooks and English-medium specialist textbooks, and compare their nature in terms of interdisciplinary academic vocabulary. We also want to know how large a college EFL student's vocabulary may be after four years of regular courses.

2 Literature Review

2.1 *The teaching context: A semi-discipline-based model in content-based instruction*

As just mentioned, college non-English course teachers in Taiwan prefer to use English-medium textbooks and deliver specialist knowledge in either English or Mandarin Chinese. From the perspective of English language teaching, the instructional format in non-language subjects is loosely equivalent to a discipline-based model in the field of content-based instruction (CBI).

The CBI approach which integrates academic content into language instruction originated in immersion programs in Canada in the 1960s (Brinton, Snow, & Wesche, 1989). The immersion programs were designed for English-speaking kindergarten pupils, who received regular half-day instruction in school subjects entirely through the medium of French. The success of this immersion project for the mastery of both French and school subjects was influential in drawing attention to the CBI methodology. By the 1980s, content-based instruction had grown by leaps and bounds. Various CBI models can be classified according to their relative focus on language and content: theme-based, sheltered, adjunct, sustained models and so on. Among a diversity of CBI formats is discipline-based language instruction where the content normally comprises discipline-based materials taught in the target language. There are two distinct contexts for this model. One involves foreign language instruction that is organized around cultural, geographic, historical and literary themes. The other embraces instruction in non-language courses that makes extensive use of informational resources in a foreign language or in content courses taught in a foreign language. As far as our context is concerned, the first happens to depict GE courses, while specialist core courses fit the second.

In contrast to 'classic' discipline-based language courses where subject teachers are responsible for presenting content materials in the target language in a way that will support language learning (Krueger & Ryan, 1993), the current EFL context is a situation where subject teachers (non-language specialists) focus on instruction of academic concepts mostly in the native language with little effort in syntax or language analysis of the English-medium specialist texts. The emphasis on what items of language are to be learned in a GE course turns out to be a focus on what specialist content is to be learned in such a 'semi'-discipline-based class. The nature of General English and English-medium specialist courses is characterized by the redirection of attention from 'content in favor of language' to 'language in favor of content'.

In the research domain of content-based instruction, a wealth of CBI-related studies are often associated with English for Specific Purposes (ESP) and task-based learning (TBL) in the following areas: register, discourse and genre analysis, learner interaction/communication as well as learning strategies (for details about the distinct and common features among CBI, ESP and TBL, see Hsu, 2006). However, to date there have been few studies on vocabulary levels and

lexical coverage in CBI's variety of teaching contexts. The need for lexical analysis in our EFL content area classrooms has triggered the present research motive in this regard.

2.2 Vocabulary size, levels and lexical coverage

The acquisition of vocabulary is an indispensable component in the process of learning a language. For instance, a rich vocabulary makes the reading skills easier to perform. More specifically, the ability to read depends in the first instance on lexical and then linguistic knowledge. The breadth and depth of a learner's vocabulary have a direct impact on reading comprehension. Limited vocabulary may be a major source of difficulty in reading an English text. According to Goulden, Nation and Read (1990), a well-educated adult native speaker of English has a vocabulary of around 17,000 words. This dramatically large number of English words, however, is a learning goal far beyond the reaches of foreign language learners like ours.

Hsu (2009) built a corpus of thirty-six General English textbooks used in universities in Taiwan and analyzed the vocabulary size and levels contained in the General English texts. Her research showed that beyond the 2,000-word level, a GE textbook can supply students with 162–2,001 new word families. The figures clearly demonstrate how big the gap is in the vocabulary capacities between a college EFL learner with limited hours of exposure to English and an English-native speaker. Fortunately, not all English words are equally important in different phases of language learning. A few researchers (e.g. Coxhead, 2000; Nation, 2001; Nation & Waring, 1997) proposed that for different purposes or in different stages of learning, some words deserve more attention and effort than others.

Nation (2001) divided vocabulary into four categories: (1) high-frequency or general service vocabulary, (2) academic vocabulary, (3) technical vocabulary and (4) low-frequency vocabulary. High-frequency words refer to those basic general service English words which constitute the majority of all the running words in all types of writing. The most well-known general service vocabulary is West's (1953) General Service List of English Words (GSL). The GSL containing the most frequently-occurring 2,000 word families of English (3,372 word types) accounts for approximately 75% of the running words in non-fiction texts (Hwang, 1989) and around 90% of the running words in fiction (Hirsh, 1993). Academic vocabulary, also called sub-technical vocabulary (Cowan, 1974) or semi-technical vocabulary (Farrell, 1990) is a class of words between technical and non-technical words and usually with technical and non-technical implications. Technical words are the ones used in a specialized field and are considerably different from subject to subject. About 5% of the words in an academic text are made up of technical vocabulary, with each subject containing roughly 1,000 word families (Nation, 2001).

Nation and Waring (1997) pointed out that the beginners of English learning should focus on the first 2,000 most frequently-occurring word families of English in the GSL, while for intermediate or advanced learners who usually study English for academic purposes, the command of the top 2,000 frequent words may no longer be their concern and the priority of their vocabulary learning may be shifted to the next level of vocabulary, i.e. sub-technical/academic vocabulary. In academic settings, ESP students do not see technical terms as a problem because these terms are usually the focus in the specialist textbooks. Low-frequency words are rarely used terms. Academic vocabulary with medium-frequency of occurrence across texts of various disciplines (somewhere between the high-frequency words and technical words) has some rhetorical functions. Acquiring these sub-technical words seems to be essential when learners are preparing for English for Academic Purposes. Alternatively, vocabulary based on Nation's (2001) four divisions can be learned in a systematic order. Students should learn first the 2,000 general words of English, followed by a set of academic words common to all academic disciplines. In line with Nation and Waring (1997), Coxhead (2000) compiled a corpus of around 3.5 million running words from university textbooks and materials from four different academic areas (law, arts and commerce as well as science), and identified 570 academic word families (AWL), which were claimed to cover almost 10% of the total words in a general academic text. Her research suggested that for learners

with academic goals, the academic word list contains the next set of vocabulary to learn after the top 2,000-word level. To put it concretely, greater lexical coverage is gained by moving on to learning 570 academic words (10% coverage) than by continuing to learn the next 1,000 words (“3–5%” coverage for the 3rd 1,000, Nation, 2006, p. 79) after the top 2,000 word families on a frequency list.

Lexical/text coverage refers to “the percentage of running words in the text known by the readers” (Nation, 2006, p. 61). Technically, it is here calculated as “the number of the words known in a text, multiplied by 100 and then divided by the total number of running words, i.e. tokens in the text” (Nation, 2001, p. 145). The assumption made behind lexical coverage is that there is a lexical knowledge threshold which marks the boundary between having and not having sufficient vocabulary knowledge for adequate reading comprehension. Or how much unknown vocabulary can be tolerated in a text before it interferes with comprehension? Some researchers regard one unknown word in every twenty words, roughly in every two lines of a text, as the necessary level beneath which readers are not expected to read an authentic text successfully (Laufer, 1989; Read, 2000; Schmitt & McCarthy, 1997). That is, one needs to know sufficiently different words/types to account for 95% of the running words in a text. Laufer (1989) noted that reading comprehension at an academic level requires 95% lexical coverage at the minimum. A lack of familiarity with more than 5% of the running words in a text (one unknown word in less than 20 words) can make reading a formidable task.

Applying the 95% threshold comprehension of input to the present research, we expect to see that the 2,000 word families (having been learned during the primary and secondary education) and the AWL 570 word families (to be learned in the ensuing phase) would cover 75% and about 10% of tokens in a text respectively, as analyzed by Hirsh (1993) and Coxhead (2000) for non-fiction texts. Taking together, the 2,570 word families would lead to 85% lexical coverage. An increase of 10% lexical coverage is therefore a goal to work towards for 95% text comprehension. In other words, apart from the 2,570 word families, how much larger a vocabulary is needed to obtain an addition of 10% lexical coverage? From another angle, this can be formulated as “beyond the 2,570-word level, to what extent does a college English-medium textbook contain higher-level vocabulary?”

3 Method

To examine the vocabulary of English-medium textbooks our students may read during college, two kinds of textbook corpora were compiled to form a basis of comparison: General English textbooks and textbooks of business and management core courses. The selection of business textbooks as a research focus was simply a random choice in that business represents one of the college majors. By lexically comparing the two corpora of textbooks, three specific questions were thus formulated as follows:

- (1) What percentage of the words in a General English textbook and in an English-medium business textbook does Coxhead’s (2000) Academic Word List cover? How many interdisciplinary academic words may one learn from a GE textbook and a business textbook in English?
- (2) What are the vocabulary levels of a General English textbook and a business textbook in English?
- (3) If a college freshman has a vocabulary size of 2,000 of the most frequent English words and the academic words across disciplines, how many new words may one learn from a General English textbook?

These three questions are closely related, being investigated from different angles. The first question was more concerned about the nature of two types of college English-medium textbooks in terms of academic vocabulary. The second question focused on the vocabulary size used in these two kinds of textbooks. The answer to Research Question 3, derived from the first and the

second, may provide some pedagogical implications for the potential role GE courses play in the current EFL context.

3.1 Textbook selection criteria

The specialist English corpus contained eight business core textbooks that were widely used among the twelve¹ departments in the college of business and management at I-Shou University in southern Taiwan, totaling 1,087,723 tokens (running words) after excluding illustrations, tables, references and proper nouns. Each business specialist textbook was scanned and the scanned chapters each contained an approximate equal number of running words, around 136,000 tokens per book on average, which took up more than one third of the book in length. Table 1 shows the composition of a corpus of business core textbooks (also see Appendix A).

Business core textbooks	Number of chapters	Tokens
Money, Banking and Financial Markets	8 out of 23 chapters	133,015
Microeconomics: A Modern Approach	9 out of 26 chapters	139,827
Principles of Marketing	7 out of 20 chapters	137,040
Accounting Principles	9 out of 26 chapters	134,198
Management	8 out of 22 chapters	135,701
Business Essentials	6 out of 16 chapters	137,081
Foundations of Production and Operations Management	7 out of 20 chapters	135,810
Introductory Statistics	6 out of 16 chapters	135,051
Excluding illustrations, charts, tables, references and proper nouns, in total		1,087,723

Table 1: Composition of a corpus of business core textbooks

In the last issue of the e-FLT Journal, the sample size of Hsu’s (2009) research on college English textbooks for general purposes was thirty-six. The criteria for inclusion of the books in the corpus were based on the popularity of GE textbooks according to sales data from eight major import bookstores. In this sequel to Hsu’s analysis of lexical coverage of GE textbooks, instead, the selection of GE textbooks was limited to the use in the school researched, i.e. I-Shou University, in order to inspect the real condition in the present school. In total, twenty GE textbooks were chosen: three low-intermediate, six intermediate, seven upper-intermediate and four advanced (see Appendix B). After removing exercises, illustrations and proper nouns, the resulting corpus contained in total 355,958 running words, as Table 2 shows.

	GE Textbook	Book Level	Number of Texts	Tokens
1	What a World 2	low-intermediate	18	11,337
2	What a World 3	intermediate	16	16,616
3	Reading for Success 2	low-intermediate	32	11,505
4	Reading for Success 3	intermediate	32	18,566
5	Interactions 1 for reading	low-intermediate	20	14,110
6	Interactions 2 for reading	intermediate	10	9,574
7	Select Readings-intermediate	intermediate	14	10,204
8	Select Readings-upper-intermediate	high-intermediate	12	15,277
9	Reading for the Real World 2	intermediate	24	15,587
10	Reading for the Real World 3	high-intermediate	24	16,114
11	Hot Topics 2	intermediate	53	16,384
12	Hot Topics 3	high-intermediate	43	22,403
13	Mosaic Reading 1	high-intermediate	20	17,618
14	Mosaic Reading 2	advanced	34	27,512
15	NorthStar –high intermediate	high-intermediate	20	16,221
16	NorthStar focus on reading-advanced	advanced	20	17,323
17	Reading for Real-high-intermediate	high-intermediate	12	8,985
18	Reading for Real-advanced	advanced	12	10,014
19	Reading Matters 3	high-intermediate	51	44,199
20	Reading Matters 4	advanced	31	36,409
Excluding proper nouns, total tokens				355,958

Table 2 is arranged in ascending order according to the book level within the same series and the book level as a whole.

Table 2: Composition of a corpus of General English reading textbooks

3.2 The instrument

Software tools for measuring the vocabulary levels by comparing the word lists made from the target text with the GSL 1,000 words and 2,000 words are available on the Internet (e.g. Compleat Lexical Tutor at <http://www.lex tutor.ca/>) but are confined to the use at or below the 2,000-word level. College English-medium textbooks involve various levels of English vocabulary, ranging from low-intermediate GE texts to more advanced professional articles. Much of the English vocabulary at the tertiary education is over the 3,000-word level. In order to measure vocabulary levels of higher-level textbooks, a large scale of word lists based on occurring frequency is needed. One key concept for understanding vocabulary levels is rank. Frequency lists by rank are a useful benchmark because words are ranked in terms of how frequently/commonly they are used. For example, the 3rd 1,000 words on a frequency list mean they are less frequent than the 2nd 1,000 words and more frequent than the 4th 1,000 words. By analogy, the 3,000-word level refers to the vocabulary of a text reaching a level that embraces the 1st, 2nd and 3rd 1,000 most frequent words of English. Considering vocabulary learning sequencing in relation to frequency, we need a software program that is installed into a series of ranked word lists (e.g. 1,000-word, 2,000-word and 3,000-word rankings, etc.) to count the number of the most frequent word lists needed until an accumulation of lexical coverage approaches 95%.

The present two corpora were run on the computer software RANGE to compare the vocabulary levels and lexical coverage of texts. RANGE is a windows-based program (at <http://www.victoria.ac.nz/lals/staff/paul-nation.aspx>) developed by Paul Nation and Alex Heatley (2002) of the Victoria University of Wellington and is freely downloadable. The program incorporated the General Service List of English Words (GSL 2000), Academic Word List (AWL 570) and British National Corpus High Frequency Word List (BNC HFWL 1st–14th 1,000) based on English words' occurring frequency, range and dispersion figures. The RANGE software can be used to compare a text against certain base word lists to see what words in the text are and are not in the word lists, and to see what percentage of the vocabulary items in the text are covered by the lists, namely lexical coverage. In particular, it can also be used to compare the vocabulary of as many as 32 text files at a time to see how much of the same vocabulary they use (i.e. range) and the frequency of occurrence of the words in total and in each file. The word lists, named base word lists by Nation and Heatley (2002) involve word families. The first base word list consists of the BNC first 1,000 high-frequency word families of English. By the same token, the second base word list comprises the second 1,000 word families (see Figure 1).

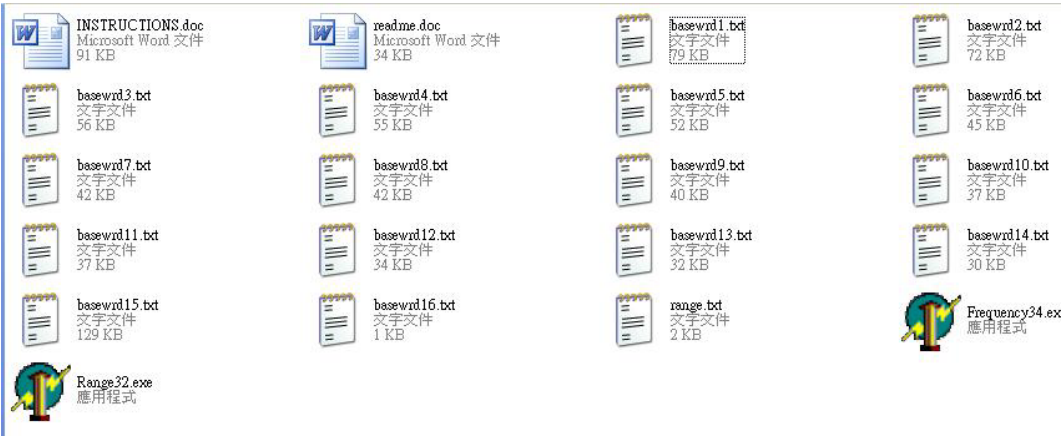


Figure 1: Base word Lists in the RANGE program

The criteria used in the RANGE program to make word families were based on Bauer and Nation's (1993) six-level basic word building processes, which include all the affixes, inflected and derived forms. Word families are regarded as an important counting unit in terms of the learning load (Nagy, Anderson, Schommer, Scott, & Stallman, 1989). The concept of a word family is used to represent a group of words whose meanings can be inferred when the meaning of the base form in the group is known to a learner. Therefore, comprehending regularly inflected or derived members of a word family does not require much effort, namely, not having to learn each form separately. For instance, the headword *ache* is grouped with its members, *aches*, *ached*, *aching* and *achy* to form a word family. Thus, the four family members are counted as the same word *ache*. The word *ache* is listed in the BNC 3rd 1,000 high-frequency word list (the Basewrd 3 in the RANGE program).

Below are some other examples of word families from the BNC HFWL 3rd 1,000.

ABSENCE
 ABSENCES
 ACCELERATE
 ACCELERATED
 ACCELERATES
 ACCELERATING
 ACCELERATOR
 ACCELERATORS

ACCELERATION
ACCELERATIONS
ACCORDINGLY
ACCOUNTANT
ACCOUNTANTS
ACCOUNTANCY
ACCUSE 3
ACCUSING
ACCUSINGLY
ACCUSES
ACCUSED
ACCUSATION
ACCUSATIONS
ACCUSER
ACCUSERS

For easy comparison, some examples of the word families in the base word lists 1 and 2 (i.e. BNC HFWL 1st 1,000 and 2nd 1,000) in the RANGE program are shown in Figure 2 and Figure 3.

The ranking of 1,000-word base word lists tells us that vocabulary in the base word list 1 appears more frequently than that in the base word list 2 in the BNC corpus.

```

A 1894
  AN 1
ABLE 4
  ABILITY 1
  ABLER 0
  ABLEST 0
  ABLY 0
  ABILITIES 0
  UNABLE 1
  INABILITY 1
ABOUT 1
ABSOLUTE 2
  ABSOLUTELY 1
  ABSOLUTIST 0
  ABSOLUTISTS 0
ACCEPT 4
  ACCEPTABILITY 0
  ACCEPTABLE 1
  ACCEPTABLY 0
  UNACCEPTABLE 1
  ACCEPTANCE 1
  ACCEPTED 0
  ACCEPTING 0
  ACCEPTS 0
  UNACCEPTABLY 0
ACCOUNT 2
  ACCOUNTED 0
  ACCOUNTING 1
  ACCOUNTS 0
ACHIEVE 2
  ACHIEVABLE 0
  UNACHIEVABLE 0
  ACHIEVED 0
  ACHIEVEMENT 1
  ACHIEVEMENTS 0
  ACHIEVER 0
  ACHIEVERS 0
  ACHIEVES 0
  ACHIEVING 0
ACROSS 1
ACT 4
  ACTED 0
  ACTING 0
  ACTION 1

```

Figure 2: Base word list 1 (BNC HFWL 1st 1,000) in the RANGE program

```

ABOVE 1
ABUSE 1
  ABUSED 0
  ABUSIVE 0
  ABUSING 0
  ABUSES 0
  ABUSER 0
  ABUSERS 0
ACCENT 1
  ACCENTED 0
  UNACCENTED 0
  ACCENTING 0
  ACCENTS 0
ACCESS 3
  ACCESSED 0
  ACCESSES 0
  ACCESSIBILITY 0
  ACCESSIBLE 1
  ACCESSING 0
  INACCESSIBLE 1
ACCIDENT 3
  ACCIDENTS 0
  ACCIDENTAL 1
  ACCIDENTALLY 1
ACCOMMODATE 2
  ACCOMMODATED 0
  ACCOMMODATES 0
  ACCOMMODATING 0
  ACCOMMODATION 1
  ACCOMMODATIONS 0
ACCORDING 1
ACCURATE 4
  ACCURACY 1
  ACCURACIES 0
  INACCURACY 0
  INACCURACIES 0
  ACCURATELY 1
  INACCURATE 1
  INACCURATELY 0
ACID 1
  ACIDS 0
  ACIDLY 0
  ACIDIC 0
  ACIDITY 0

```

Figure 3: Base word list 2 (BNC HFWL 2nd 1,000) in the RANGE program

3.3 *Choosing base word lists*

Two categories of word lists were downloaded from the RANGE program (see below for the first two). To tackle Research Question 3, one mixed word-family list was edited for this research. They were:

1. Coxhead's (2000) Academic Word List (AWL) containing 570 word families (AWL 570).
2. The BNC 1st–14th 1,000 high-frequency English word lists (BNC HFWL 14,000).
3. A mixed word list of AWL 570 and BNC HFWL 2,000

Coxhead's (2000) 570 academic word-family list has been included in the RANGE program. This list was adopted for Research Question 1 to measure how frequently the academic words across disciplinary domains occur in a GE textbook and in a business textbook, and to examine the extent to which a GE textbook can prepare a learner for reading professional texts containing such sub-technical vocabulary.

The GSL 2,000-word families available in RANGE as a base word list for this research did not seem to suffice in measuring vocabulary size when academic articles include a higher-level vocabulary. As a result, 14,000 high-frequency word families, which were made from the British National Corpus and already built into the RANGE program, were adopted. The British National Corpus (BNC) with more than 100 million words is considered one of the largest corpora of present-day English usage in speech and in publications in the United Kingdom (Leech, Rayson, & Wilson, 2001). The 14,000 high-frequency word families were divided into fourteen base word lists, each containing exactly 1,000 word families. Apart from the BNC 1st–14th 1,000 word lists, some proper nouns and Roman numerals as well as spoken interjections and exclamations were also incorporated in the RANGE program, appearing as base word lists 15 and 16. Base word list 15 (a proper noun list) and base word list 16 (an interjection and exclamation list) were beyond the research focus and were hence not factored in.

Figure 4 is a screenshot, which demonstrates the fourteen base word lists, i.e. fourteen 1,000 high-frequency English word families made from the British National Corpus. For example, Base word list 1 includes 1,000 base forms, their inflected forms and derivatives, thereby making a total of 6,348 different words (types).

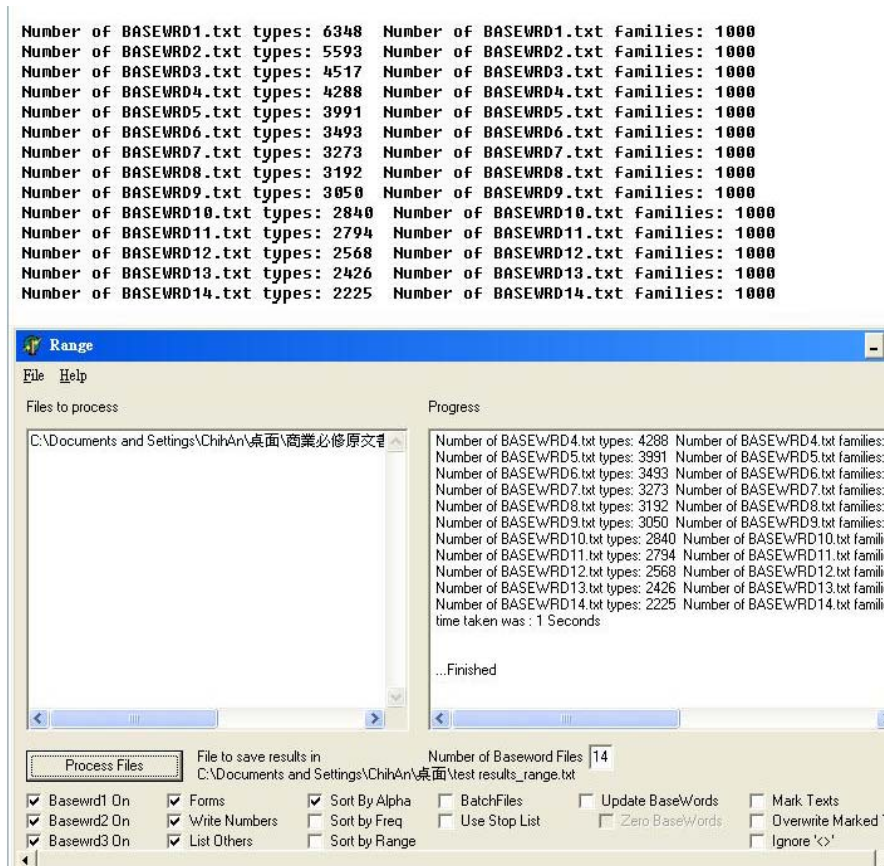


Figure 4: Number of word families and word types in the BNC high-frequency word lists

To answer Research Question 3, the word-family list utilized was a combination of two word lists, AWL 570 and BNC HFWL 2000, having a total of 2,293 word families after excluding an overlap of 277 word families occurring both in the AWL and BNC HFWL 2000. As aforementioned, matriculating freshmen who have passed the college entrance exam were presumed to have a vocabulary size of around 2,000 words. The mixed word list was hence used to measure how many new words at the minimum a college student may possibly learn from a GE course, in addition to academic words that are frequently encountered in specialist core textbooks in English.

Subsequently, to calculate lexical coverage and to assess the vocabulary levels of the textbooks contained in the two corpora, the RANGE computing program was run each time against one of the following three base word lists, using the same counting unit, word families: (1) AWL 570, (2) BNC HFWL 1st–14th 1,000, and (3) a mixed word list of AWL 570 and BNC HFWL 2,000.

4 Results and Discussion

4.1 The coverage and the number of interdisciplinary academic words in a GE textbook and an English-medium business textbook

Table 3 and Table 4 address the first research question. According to Coxhead (2000), the 570 academic word families account for approximately 10% of the tokens in academic texts but only

1.4% of the total words in a fiction collection of the same size. The most frequent word families in the AWL are, for example, *approach*, *constitute*, *identify*, *indicate*, *interpret*, *specific*, *vary* and so on. Some of the least frequent academic vocabulary in Coxhead's (2000) corpus is *adjacent*, *conceive*, *collapse*, *incline*, *persist*, *whereby* and *notwithstanding* etc. The second column from left in Table 3 shows the AWL occurrence in tokens and in percentage in the present corpus. For instance, the textbook *Reading Matter 4* contained 36,409 running words after deleting proper nouns, with 785 different types of academic words appearing in the texts 1,980 times. The 1,980 occurrences arising out of 415 academic word families made up 5.44% of the total words in the textbook. In this corpus, *Reading Matters 4* included the most interdisciplinary academic words, having 415 word families out of the AWL 570. None of the GE textbooks in the corpus seemed to meet the figure in the literature, i.e. 10% coverage counted in tokens with regard to the commonly used academic words that students may come across in reading professional articles. The top four textbooks² of higher AWL coverage were *Reading for the Real World 3* (6.54% in tokens), *Reading for the Real World 2* (5.93% in tokens), *Hot Topics 3* (5.71% in tokens) and *Reading Matter 4* (5.44% in tokens). These four books were labeled by the publishers as suitable for intermediate, high-intermediate and advanced EFL learners (see Table 2 for book levels). *What a World 2* for low-intermediate learners had the lowest AWL coverage, with 1.3% coverage counted in tokens and 3.76% in types. It is not surprising to see that the higher AWL coverage and the higher-level GE textbooks bore some relationship to each other, since advanced textbooks are meant to teach advanced vocabulary and to expand students' vocabulary breadth. In a similar vein, it is not disappointing to find that basic GE textbook contained little academic vocabulary since lower-level books serve other purposes in language learning.

Table 3 demonstrates that the academic lexical items that can be learned by using one of the twenty GE textbooks range from as little as 49 to as much as 415 word families. This result produces an overall picture of the amount of academic vocabulary a student will be equipped with after taking a GE course for one year.

General English Textbooks	Number of words	AWL occurrence in tokens/ %	AWL occurrence in types/ %	AWL occurrence in families
Reading Matters 4	36,409	1,980/ 5.44%	785/ 12.82%	415
Reading Matters 3	44,199	2,110/ 4.77%	776/ 11.09%	408
Mosaic Reading 2	27,512	1,247/ 4.53%	605/ 10.41%	361
Reading for the Real World 3	16,114	1,054/ 6.54%	519/ 15.25%	313
Hot Topics 3	22,403	1,279/ 5.71%	507/ 12.48%	312
Reading for the Real World 2	15,587	924/ 5.93%	477/ 13.55%	310
Reading for Success 3	18,566	877/ 4.72%	431/ 11.22%	292
Mosaic Reading 1	17,618	660/ 3.75%	398/ 10.45%	275
Select Readings-hi-intermediate	15,277	609/ 3.99%	371/ 11.00%	266
NorthStar reading-advanced	17,323	581/ 3.35%	360/ 9.75%	254
NorthStar –high intermediate	16,221	508/ 3.13%	292/ 8.73%	211
Select Readings-intermediate	10,204	472/ 4.63%	280/ 12.29%	211
Hot Topics 2	16,384	701/ 4.28%	311/ 10.63%	198
Reading for Success 2	11,505	374/ 3.25%	232/ 8.49%	174
Reading for Real-advanced	10,014	305/ 3.05%	220/ 8.40%	166
What a World 3	16,616	416/ 2.50%	208/ 7.17%	158
Interactions 1 for reading	14,110	632/ 4.48%	221/ 8.45%	151
Interactions 2 for reading	9,574	474/ 4.95%	219/ 10.51%	149
Reading for Real hi-intermediate	8,985	235/ 2.62%	183/ 8.01%	149
What a World 2	11,337	147/ 1.30%	72/ 3.76%	49

Twenty textbooks are tabulated in descending order according to the number of occurrences of academic word families.

Table 3: The AWL coverage of GE textbooks

Business textbooks	Number of words	AWL occurrence in tokens/ %	AWL occurrence in types/ %	AWL occurrence in families
Management	135,701	16,474/ 12.14%	1,932/ 18.94%	537
Money, Banking and Financial Markets	133,015	15,469/ 11.63%	1,480/ 18.41%	520
Business Essentials	137,081	14,517/ 10.59%	1,555/ 15.11%	519
Introductory Statistics	135,051	14,626/ 10.83%	1,989/ 18.16%	497
Principles of Marketing	137,040	14,417/ 10.52%	1,545/ 14.85%	485
Foundations of Production and Operations Management	135,810	19,719/ 14.52%	2,358/ 20.66 %	481
Microeconomics: A Modern Approach	139,827	15,012/ 10.74%	1,078/ 19.83%	431
Accounting Principles	134,198	15,080/ 11.24%	1,138/ 20.24%	421

Table 4: The AWL coverage of English-medium textbooks of business core courses

Moving from Table 3 to Table 4, we can find that all of the eight business textbooks in the corpus reached the AWL 10% coverage counted in tokens, with two books going over 12%. The present data results verify Coxhead's (2000) findings that the AWL's 12% coverage of the commerce sub-corpus appears to be slightly advantageous for business students, compared with the average 10% coverage of the tokens in the academic corpus including twenty-eight subject areas in arts, commerce, law and science. In addition, Table 4 shows that by taking one of the required business core courses, students can learn as low as 421 to as high as 537 academic word families.

The reader may be interested to see a short section of text from the two corpora each. The following two passages are randomly selected from the GE textbooks and the business textbooks. The words included in the AWL are underlined and in bold.

After hours of not eating, there came a point where I wondered: what in the world was I thinking, deciding not to eat for 30 hours and getting people to pay me for it as well? You might ask why any sensible person would pay a teenage girl to skip three meals over the course of two days. The answer is "the 30-hour famine."

The 30-hour famine is a program of World **Vision**, an international Christian relief and development organization **established** in the 1950s and dedicated to helping children, because "when children have food, shelter and schooling and are protected, valued and loved, a **community** thrives."

[This excerpt is from "The 30-Hour Famine: The absence of food for thought" by Madeleine Irving, in M. Costantino & T. Dabbs (2001). *Reading for real-high intermediate*. (pp. 75-79). Vancouver, CA: Lynx.]

Although audit and tax services have **traditionally** been the mainstay of public accountants, another **area**, generally called management advisory services, has experienced rapid growth in recent years. While performing audits, accountants often discover defects in the client's accounting system or its operation. It is natural for the accountant to advise the client on how to correct defects and improve **procedures** so as to produce more efficient operations and related cost savings. Clients expect these recommendations and often engage the accountant to **undertake** additional **investigations** for improving operations. Public accountants offer a wide **range** of advisory services, some with little relationship to accounting. These services include advice on such things as mergers with other companies, installation or **modification** of accounting systems, **design** or **modification** of pension plans, and advice regarding budgeting, forecasting, and general **financial** planning.

[This excerpt is from *Accounting Principles* by Weygandt et al. (2007, p. 49).]

Among the total of 105 words in the GE excerpt, three words (underlined and in bold) belong to the AWL, whereas 10 words out of the 136-word business excerpt are in the AWL. The AWL lexical coverage of the GE passage and the business passage is 2.86% and 7.35% in tokens respectively. The results of the two excerpts are consistent with the results of the two textbook corpora (see Table 3 and Table 4). A business core textbook generally encompasses a greater amount of academic vocabulary than a GE textbook.

Table 5 is a list of the most and the least frequently-occurring academic headwords in the corpus of GE textbooks and Table 6 enumerates the most frequent academic headwords in the corpus of business textbooks.

Academic headword	Range (across the number of textbooks)	Word family occurrences
create	20	243
final	20	122
area	19	174
job	18	206
major	18	98
similar	18	83
adjacent	1	1
allocate	1	1
discrete	1	1
intrinsic	1	1
regime	1	1
hierarchy	1	1

Table 5: The most and the least frequent academic words in twenty GE textbooks

Academic headword	Range (across the number of textbooks)	Word family occurrences
income	8	1,487
purchase	7	1,239
economy	8	1,168
require	8	1,020
strategy	6	991
margin	8	967
finance	8	983
corporate	7	863
revenue	7	780
assume	8	754
invest	7	751
vary	8	713

Table 6: The most frequent academic headwords in eight business textbooks

In the GE textbook corpus, two academic words *create* and *final* as well as their family members, occurred in all the twenty GE textbooks (see Table 5). The word *create*, by itself appeared in all twenty textbooks. When its family members (created, creates and creating, etc.) were included, they showed up as many as 243 times. The word *area* as well as the three words, *job*, *major* and *similar* appeared in 19 and 18 textbooks respectively. Words such as *adjacent*, *discrete* and *regime*, though among the interdisciplinary academic words, only showed up once in one GE textbook. Words like *create*, *job* and *similar* do not show an academic flavor in a strong sense. These words appear to be common language components that slip invisibly in and out of everyday conversation and content domain talk. They are commonly used by different groups of people. It is highly possible that EFL students may have encountered and learned these words elsewhere before. On the other hand, least frequent words such as *intrinsic*, *allocate* and *hierarchy* reveal a sense of business content domain.

In the present business textbook corpus, words like *principal*, *credit*, *corporate*, *currency*, *exploit*, *offset*, *levy*, *innovative*, *valid*, *strategy* and *deviate* are common words in economics, accounting and marketing. Table 6 illustrates the most frequent academic headwords across eight business specialist textbooks, many of which naturally show a business flavor in connection with profit, money and numbers.

It may be taken for granted that sub-technical/academic words should be much more frequent in the specialist corpus than in the GE corpus. The results seem to imply that even though the general academic vocabulary goals set in GE courses are missed by a wide margin, business students can still have multiple chances to intensively encounter and learn a sufficient amount of high-frequency academic/sub-technical vocabulary across disciplinary domains within their own specialist fields. As far as this privilege is concerned (benefiting students with 421–537 academic words from a business textbook versus 49–415 academic words from a GE textbook), business English appears to favor such all-embracing academic vocabulary learning, but this interpretation needs further comparison with English-medium specialist textbooks of other academic subjects. One point that needs highlighting here is that English-medium business courses cannot substitute for GE courses simply on account of more academic words occurring. On the contrary, initial exposure to academic words in GE courses solidifies the learning of these vocabulary items simultaneously or later on when they are met in the specialist materials.

In the process of analyzing Coxhead's (2000) 570 most frequent interdisciplinary academic words (AWL 570), the researcher noticed that some of the words that appeared very frequently in

over a half of the eight business textbooks are not included in the AWL 570. Below are some of such words.

administer	appraisal	accomplish	adverse	appeal	boom
boost	barrier	cope	counter	cancel	certify
compliance	campaign	demographic	elaborate	engage	efficiency

Instead of being finance-oriented, these frequent words that are not in the AWL but reveal empirical and action research nature in the business register have raised the researcher's concerns about a single core of high-frequency words for academic study.

Some word families like *inherent*, *convene* and *arbitrary* in the AWL did not appear in the business textbook corpus at all. This may be partly ascribed to the fact that the present corpus that was narrowed to one academic field (business-related only) was more specific than Coxhead's (2000) that consisted of 28 subjects from four different academic areas. This confirms that not all of these interdisciplinary/general academic word families (i.e. AWL 570) are frequently used in the business domain.

Consequently, a new concern is generated from the above two findings and leads us to query the completeness of AWL in representing business academic vocabulary. This issue echoing Hyland and Tse's (2007) contention for developing a more restricted, discipline-based lexical repertoire is beyond the scope of the present research but worth investigation.

4.2 The vocabulary levels of GE textbooks and English-medium business textbooks

The BNC HFWL was used to identify the division among the diverse vocabulary levels contained within the GE textbooks and business textbooks. As mentioned previously, there are fourteen 1,000-word bands created from the British National Corpus in the RANGE program. The lexical coverage of each 1,000-word band in the target textbook was calculated. Vocabulary levels were thereby defined as the number of 1,000-word bands needed until the total coverage reached 95%.

Referring back to the 136-word excerpt from the business textbook *Accounting Principles* (in Section 4.1), the following output of the RANGE program using the fourteen base word lists (see Figure 5) illustrates how many word families in the input text are found in each list. We can see 64 word families are in the first base word list, 8 in the second list, 3 in the third and so on. The business excerpt has 87.5% lexical coverage of the first and the second base word lists (79.41% + 8.09% in tokens). A shortage of 7.5% lexical coverage (95% – 87.5% = 7.5%) means that it is likely for a business text to contain many words in the base word list 3 and above, if extending from the present excerpt of 136 words to a whole book. In other words, one may encounter one new word beyond the 2,000 words for every 13 words of text (roughly one unknown word in each line). This could affect the reading ability of a student with such limited vocabulary if 95% comprehension and fluency in reading is required. Conversely, if one's vocabulary capacity approaches the level of the 4th 1,000 words along the BNC scale (i.e. one's vocabulary size is 4,000 words), s/he may not face a daunting amount of dictionary work while reading *Accounting Principles*.

WORD LIST	TOKENS/%	TYPES/%	FAMILIES
one	108/79.41	75/78.13	64
two	11/ 8.09	9/ 9.38	8
three	7/ 5.15	4/ 4.17	3
four	6/ 4.41	5/ 5.21	4
five	2/ 1.47	1/ 1.04	1
six	1/ 0.74	1/ 1.04	1
seven	0/ 0.00	0/ 0.00	0
eight	0/ 0.00	0/ 0.00	0
nine	0/ 0.00	0/ 0.00	0
ten	0/ 0.00	0/ 0.00	0
11	1/ 0.74	1/ 1.04	1
12	0/ 0.00	0/ 0.00	0
13	0/ 0.00	0/ 0.00	0
14	0/ 0.00	0/ 0.00	0
not in the lists	0/ 0.00	0/ 0.00	????
Total	136	96	82

Number of BASEWRD1.txt types: 6348	Number of BASEWRD1.txt families: 1000
Number of BASEWRD2.txt types: 5593	Number of BASEWRD2.txt families: 1000
Number of BASEWRD3.txt types: 4517	Number of BASEWRD3.txt families: 1000
Number of BASEWRD4.txt types: 4288	Number of BASEWRD4.txt families: 1000
Number of BASEWRD5.txt types: 3991	Number of BASEWRD5.txt families: 1000
Number of BASEWRD6.txt types: 3493	Number of BASEWRD6.txt families: 1000
Number of BASEWRD7.txt types: 3273	Number of BASEWRD7.txt families: 1000
Number of BASEWRD8.txt types: 3192	Number of BASEWRD8.txt families: 1000
Number of BASEWRD9.txt types: 3050	Number of BASEWRD9.txt families: 1000
Number of BASEWRD10.txt types: 2840	Number of BASEWRD10.txt families: 1000
Number of BASEWRD11.txt types: 2794	Number of BASEWRD11.txt families: 1000
Number of BASEWRD12.txt types: 2568	Number of BASEWRD12.txt families: 1000
Number of BASEWRD13.txt types: 2426	Number of BASEWRD13.txt families: 1000
Number of BASEWRD14.txt types: 2225	Number of BASEWRD14.txt families: 1000

Table of Ranges: Types

Figure 5: The RANGE output of the 136-word excerpt from *Accounting Principles*

Table 7 displays the vocabulary levels of the four GE reading textbooks in the same series: *Interactions 1*, *Interactions 2*, *Mosaic 1* and *Mosaic 2* for learners of different proficiency levels.

Running these four books on the BNC HFWL 1st–14th 1,000 through the RANGE program, it can be seen that the vocabulary levels of *Interactions 1*, *Interactions 2*, *Mosaic 1* and *Mosaic 2* were 4,000, 5,000, 6,000 and 13,000 words with the accumulated lexical coverage at 95%.

Table 7 shows that the vocabulary distribution of *Mosaic Reading 2* among the different bands of the ranked BNC high-frequency word lists is more diverse than that of the other three books in the same series. There is a progression of a 1,000-word increase across the three books, *Interactions 1*, *Interactions 2* and *Mosaic 1*. However, there is a striking difference in vocabulary levels between these two books, *Mosaic 1* and *Mosaic Reading 2*, one for high-intermediate learners and the other for advanced learners. The 13,000-word level reveals that *Mosaic Reading 2* had a much larger vocabulary and using this textbook would result in learners working on a wider variety of vocabulary, some of which they would encounter only once or no more than a few times throughout the book. In contrast, *Interactions 1* used a smaller vocabulary, converging at the 4,000 word level along the scale of the BNC HWFL. Because the BNC 1st–14th 1,000 English words are ranked in accordance with their frequency of occurrence, with the 1st 1,000 words being the most frequent and correspondingly the 14th 1,000 words the least frequent, a book with a higher vocabulary level can be interpreted as having more English words appearing in the latter 1,000-word bands.

BNC HFWL	<i>Interactions 1</i> tokens/ coverage	<i>Interactions 2</i> tokens/ coverage	<i>Mosaic 1</i> tokens/ coverage	<i>Mosaic 2</i> tokens/ coverage
1 st 1,000	11,568/ 81.98%	7,891/ 82.42%	14,341/ 81.40%	21,070/ 76.58%
2 nd 1,000	1,331/ 9.43%	784/ 8.19%	1,266/ 7.19%	2,376/ 8.64%
3 rd 1,000	378/ 2.68%	237/ 2.48%	479/ 2.72%	803/ 2.92%
4th 1,000	**230/ 1.63%	182/ 1.9%	312/ 1.77%	662/ 2.41 %
5th 1,000	136/ 0.96%	** 95/ 0.99%	179/ 1.02%	390/ 1.42%
6th 1,000	78/ 0.55%	84/ 0.88%	**165/ 0.94%	217/ 0.79%
7 th 1,000	38/ 0.27%	38/ 0.40%	87/ 0.49%	145/ 0.53%
8 th 1,000	55/ 0.39%	11/ 0.11%	59/ 0.33%	180/ 0.65%
9 th 1,000	23/ 0.16%	52/ 0.54%	45/ 0.26%	94/ 0.34%
10 th 1,000	16/ 0.11%	41/ 0.43%	35/ 0.20%	64/ 0.23%
11 th 1,000	10/ 0.07%	14/ 0.15%	29/ 0.16%	54/ 0.20%
12 th 1,000	14/ 0.10%	3/ 0.03%	15/ 0.09%	41/ 0.15%
13th 1,000	23/ 0.16%	6/ 0.06%	27/ 0.15%	**51/ 0.19%
14 th 1,000	3/ 0.02%	5/ 0.05%	18/ 0.10%	36/ 0.13 %

“***” means that by this level, the accumulated lexical coverage of each 1,000 word list reached 95%.

Table 7: Vocabulary levels of *Interactions 1*, *Interactions 2*, *Mosaic 1* and *Mosaic 2*

Table 8 and Table 9 are two complete lists of vocabulary levels across the twenty GE textbooks and the eight business textbooks and their vocabulary distribution among the BNC HFWL 1st–14th 1,000, with the accumulated text coverage reaching 95%.

There are four apparent exceptions in Table 8 with regard to the book level claimed by the publishers and the vocabulary level measured by the BNC HFWL 1st–14th 1,000 words. Two low-intermediate GE textbooks, *Reading for Success 2* and *What a World 2*, had a vocabulary level surpassing 7,000 words and a wider dispersion along the scale of BNC high-frequency words. Counter to expectations, these two books at the low-intermediate level actually had more middle- and low-frequency English words. By contrast, two advanced books, *NorthStar-advanced* and *Reading for Real-advanced*, had a denser distribution of words, scattering in the range of the 4th and 5th 1,000-word bands. The above two situations give some evidence that it is possible to select an advanced-level GE textbook with a lower vocabulary level. The discrepancies in vocabulary levels among GE textbooks imply that textbook authors may not apply the same standard in their selection of words while writing their teaching materials for the publishers. It can also be attributed to the fact that articles in textbooks for different reading purposes may involve different levels of difficulty. Some low-frequency words appear more frequently in a certain genre or subject matter and can function as technical words in certain fields. By the same token, some vocabulary presumed difficult by some authors may be considered easy by others.

GE Textbook	Book Level	Vocabulary Level
Mosaic Reading 2	advanced	13,000
Reading Matters 3	high-intermediate	9,000-9,500
Reading for Success 2	low-intermediate	7,000-7,500
What a World 2	low-intermediate	7,000-7,500
Reading for the Real World 3	high-intermediate	6,000-6,500
Mosaic Reading 1	high-intermediate	6,000
Reading for Success 3	intermediate	6,000
Reading for the Real World 2	intermediate	5,500-6,000
Reading Matters 4	advanced	5,500-6,000
Hot Topics 3	high-intermediate	5,500-6,000
What a Word 3	intermediate	5,000-5,500
NorthStar-focus on reading-advanced	advanced	4,500-5,000
NorthStar-high intermediate	high-intermediate	4,500-5,000
Reading for Real-advanced	advanced	4,000-4,500
Reading for Real-high-intermediate	high-intermediate	4,000-4,500
Select Readings-high-intermediate	high-intermediate	4,000-4,500
Interactions 2 for reading	intermediate	5,000
Interactions 1 for reading	low-intermediate	4,000
Hot Topics 2	intermediate	3,500-4,000
Select Readings-intermediate	intermediate	2,500-3,000

The vocabulary level of each GE textbook was obtained by counting the number of words from the top of BNC HFWL until the accumulated lexical coverage reached 95%.

Table 8: Vocabulary levels of the GE Textbooks

There are four apparent exceptions in Table 8 with regard to the book level claimed by the publishers and the vocabulary level measured by the BNC HFWL 1st–14th 1,000 words. Two low-intermediate GE textbooks, *Reading for Success 2* and *What a World 2*, had a vocabulary level surpassing 7,000 words and a wider dispersion along the scale of BNC high-frequency words. Counter to expectations, these two books at the low-intermediate level actually had more middle- and low-frequency English words. By contrast, two advanced books, *NorthStar-advanced* and *Reading for Real-advanced*, had a denser distribution of words, scattering in the range of the 4th and 5th 1,000-word bands. The above two situations give some evidence that it is possible to select an advanced-level GE textbook with a lower vocabulary level. The discrepancies in vocabulary levels among GE textbooks imply that textbook authors may not apply the same standard in their selection of words while writing their teaching materials for the publishers. It can also be attributed to the fact that articles in textbooks for different reading purposes may involve different levels of difficulty. Some low-frequency words appear more frequently in a certain genre or subject matter and can function as technical words in certain fields. By the same token, some vocabulary presumed difficult by some authors may be considered easy by others.

<i>Business Textbook</i>	<i>Vocabulary Level</i>	<i>Business Textbook</i>	<i>Vocabulary Level</i>
Money, Banking and Financial Markets	4,500–5,000	Management	4,000–4,500
Microeconomics: A Modern Approach	4,000	Business Essentials	5,000
Principles of Marketing	5,000	Foundations of Production & Operations Management	4,000
Accounting Principles	4,000–4,500	Introductory Statistics	4,000–4,500

The vocabulary level of each business textbook was obtained by counting the number of words from the top of BNC HFWL until the accumulated lexical coverage reached 95%.

Table 9: Vocabulary levels of the business textbooks

Comparing Table 8 with Table 9, we can see that the vocabulary levels of the business-specific course books were more focused than those of GE textbooks, which used a much larger vocabulary. There was no apparent difference in vocabulary level among the business core textbooks. Overall, the eight business textbooks involving specialist knowledge of different content areas included consistent vocabulary levels of around 4,000–5,000 words. That is to say, if a student has a vocabulary capacity of 5,000 words, he/she would be able to gain an adequate understanding of the eight English-medium business textbooks with equal ease in terms of the frequency of consulting a dictionary. Regardless of departments, say, economics, accounting or management, studying any of the business core textbooks did not entail a higher vocabulary capacity than the others. Nevertheless, it is highly likely that business students' English vocabulary size would level off at 5,000 words because they do not need to read English articles outside of their technical textbooks, which is often the case in EFL settings.

Based on Nation's (2001) classification of 4 kinds of vocabulary, technical words and low-frequency words in the present business English corpus were all that were left after high-frequency 2,000 words and academic vocabulary were counted. In view of convergent vocabulary levels of 4,000–5,000 words at the 95% lexical coverage in business specialist texts, business technical vocabulary can be regarded as being scattered over the BNC HFWL 3rd–5th 1,000. This may in part explain why some business terms and jargon are common in ordinary English and the more generally accessible nature of business English, as opposed to medical and legal English as well as the English of science and technology.

4.3 The number of new words an EFL student may learn from a GE textbook beyond the top 2,000-word level and the commonly-used academic vocabulary

As is evident from the results of Research Question 2, students may learn more new words from a higher-level GE textbook than a business core textbook in English. Upon admission to college, business freshmen may take a GE course and business foundation courses at the same time. Accordingly, GE textbooks as well as English-medium business textbooks complement each other in helping to meet students' need of interdisciplinary academic words. When students are equipped with the 2,000 most frequent English words and commonly-used academic words, we become more concerned about how much further their vocabulary can be developed within one year of the GE course. The results of Research Question 3 would not only show how many new words our students may learn from GE textbooks of assorted levels but also reveal the ultimate vocabulary levels for graduating college students. This is due to the fact that in the current EFL context, class time is probably the only learning opportunity for non-English majors to learn English when they are not interested in English and do not have a learning motive. After the first two years of GE courses (reading mainly in the first year and listening with full focus in the second year), there will be no formal/required English courses in the third and fourth years.

To answer Research Question 3, a mixed word-family list of BNC HFWL 2,000 and AWL 570 was hence adopted as the computing basis for analysis, since it represents the average level of a college student's vocabulary capacity while taking English-medium academic courses. Running

the AWL 570 on the BNC HFWL 2,000 via the RANGE software, it was detected that there is an overlap of 277 word families between the AWL 570 and the BNC HFWL 2,000³. Excluding the 277 repetitive words, a non-English major may have a vocabulary size of 2,293 words on average ($2,000+570-277=2,293$) at the stage of taking foundation courses, namely in the first year of college.

Table 10 gives some indication of how many new words a college student may learn from a GE textbook. The columns *in the mixed base word list of BNC HFWL 2,000 and AWL 570* and *not in the mixed base list* present how much of the vocabulary in a GE textbook is familiar to a learner and how many words the learner may not know.

GE Textbook	The number of word families in the mixed base word list of BNC HFWL2000 and AWL570	The number of word families NOT in the mixed base list	Total number of word families
Reading Matters 3	2,089	1,327	3,416
Reading Matters 4	2,053	1,234	3,287
Mosaic Reading 2	1,939	1,170	3,109
Mosaic Reading 1	1,645	657	2,302
NorthStar-advanced	1,630	633	2,263
Reading for Success 3	1,638	625	2,263
Hot Topics 3	1,696	566	2,262
NorthStar-high intermediate	1,562	559	2,121
Select Readings-high intermediate	1,572	537	2,109
Reading for Real-advanced	1,301	492	1,793
Reading for the Real World 2	1,619	437	2,056
Reading for the Real World 3	1,556	437	1,993
What a World 3	1,322	405	1,727
Reading for Success 2	1,285	375	1,660
Reading for Real hi-intermediate	1,177	374	1,551
Interactions 1	1,277	298	1,575
Hot Topics 2	1,391	274	1,665
What a World 2	950	218	1,168
Interactions 2	1,136	217	1,353
Select Readings-intermediate	1,306	190	1,496

Twenty textbooks are tabulated in descending order according to the number of word families *not in the mixed base list*.

Table 10: The number of word families covered by a mixed base word list of BNC HFWL 2,000 and AWL 570 across GE textbooks

For example, 2,089 of the word families in *Reading Matters 3* were in the mixed base word list of BNC HFWL 2,000 and AWL 570, while 1,327 word families were not. These 1,327 word families could possibly be college students' new words. By and large, if a college student reads all the texts in a GE course book within an academic year, he/she would come across 190–1,327 new words beyond the 2,000-word level and general academic vocabulary, appearing in different types.

This broad range of 190 to 1,327 new words implies that the vocabulary learning goals are quite wide-ranging. Adding 190 new words to the 2,293 words he/she already possesses, the student would have a vocabulary of 2,483 words in total if he/she studies a lower-level GE textbook. The 2,483 words may be viewed as a low-achieving student's graduation level in English, due to the fact that there are no more required English courses afterwards. On the other hand, adding 1,327 words to the 2,293 words he/she already possesses, the student may reach a vocabulary of 3,620 words if a higher-level GE textbook is chosen. If the 3,620 words constitute a high-achieving college student's English level before graduation, it may serve as a warning that there is insufficient English instruction under the current curriculum design. In consequence, it would not be surprising to predict and see a low passing rate in any accredited English proficiency test (e.g. TOEFL, IELTS and TOEIC, etc.) among Taiwanese students.

Furthermore, we may speculate on the possibility if the 1,327 new words for a high-achieving student beyond the top 2,000 frequent words (75% coverage) and the AWL 570 (around 10% coverage) are sufficient to achieve an extra lexical coverage of 10%, since the 3rd 1,000 frequent words only make up a 3–5% proportion of a running text, as stated in Section 2.2. No matter what level of GE book is adopted, studying just one GE textbook apparently does not suffice. However, it is often the scenario in GE classrooms to use one textbook through two semesters.

A commonly used GE textbook contains 16 to 32 lessons⁴ based on separate texts often with little thematic coherence. The lack of coherence makes it unlikely that learners will meet the higher-level vocabulary just studied in the following lessons except in the case of high frequency words. From a pedagogical perspective, the retention of a new word is much dependent on repeated exposure to that word. The acquisition of a word is grounded in the hypothesis that the effect from short-term learning cannot be retained over a longer period of time and that the long-term effect is often sustained by long-term training. We are inclined to believe that recycling higher-level vocabulary in stages is better in terms of learning effect than providing a single massive dose. As there is only one year of the GE reading course (in the second year, GE aims at listening), the sporadic occurrences of higher-level vocabulary make the learning burden become even heavier for non-English majors.

In light of the potential role of GE courses in the present EFL context in expanding vocabulary, the researcher wishes to propose extending General English courses from the first two years up till the fourth year. The vocabulary gap can be narrowed, but only after a number of years of English-medium education.

5 Conclusion

5.1 Findings and pedagogical implications

This research continues from Hsu's (2009) study of college English textbooks for general purposes by adding a business textbook corpus for comparison. Drawing upon the notion of lexical coverage and the ranking of frequent word lists, the principal concern was threefold:

1. Coxhead's (2000) 570 academic word families accounted for 1.3–6.54% of the total words in a GE textbook and 10.52–14.52% of the tokens in a business textbook. Approximately 49 to 415 interdisciplinary academic words could be learned from a GE reading textbook as opposed to 421–537 academic words learned from a business textbook.
2. The English-medium business core textbooks used a smaller vocabulary than the higher-level GE textbooks. The English vocabulary levels of the business specialist books converged at the BNC HFWL 4,000–5,000. The GE textbooks offered texts of varying vocabulary levels, ranging from the BNC HFWL top 3,000 most frequently-occurring words to the BNC HFWL 13,000. After the 5,000-word level, there was not much resemblance between the two corpora of the textbooks. At the dividing point of 5,000 words, a GE textbook above this level has greater vocabulary breadth than any English-medium business textbook.
3. Apart from the 2,000-word level and the AWL 570, a GE textbook in the corpus can supply

students with 190 to 1,327 new word families.

The results of this research point to one problem concerning English instruction hours in the present tertiary context. One year of GE reading course with 2–3 credit hours per semester is meager. When complaining about Taiwanese college students' low proficiency of English and the little effort made to study English, we may have neglected to examine the present curriculum design for English programs in relation to the number of study years. If a threshold vocabulary of 5,000 words is the goal, then extending GE courses to four full years is justifiable on the grounds that the vocabulary size of a non-English majoring student may stop at 3,620 words with only one year of General English study.

Thanks to non-English subject teachers' preference for English-medium specialist textbooks, increasing students' lexical knowledge to the 4,000–5,000 vocabulary level becomes a more promising prospect. Although in content area classrooms, students may pay little attention to the rule-governed aspect of language form and syntax, the repetition of vocabulary in a longer text and sustained exposure to several English-medium specialist textbooks will help in strengthening the retention of the vocabulary. For this reason, content-based instruction is highly advocated.

Parallel to English-medium academic disciplines, GE courses make up for their insufficiency by concentrating on language use. Accordingly, the researcher would like to conclude that the fulfillment of vocabulary goals requires no radical new teaching approach but rather the integration of existing GE courses and content courses such that they complement each other. It is urged that GE courses be exploited flexibly by supplementing teaching materials to consolidate less frequent vocabulary learning or giving students extracurricular reading assignments.

5.2 *Limitations and recommendations*

Although the current survey contributes to the literature of the lexical coverage and vocabulary levels in an EFL setting, it has worked within a narrow focus on the field of business. The findings may serve as a comparison basis for research into other academic areas. A side issue arising from this study is the query about whether Coxhead's (2000) single core of academic vocabulary is applicable to all academic areas. The results show that her academic word list (AWL) does not fully represent the frequently-used business academic vocabulary in business textbooks. Hence, establishing a more specific, discipline-based academic word list may be interesting to pursue. Moreover, it is not sufficient merely to teach vocabulary in isolation. Rather, lexis should be taught in context. To deepen lexical knowledge and to raise students' sense of discovery learning, corpus-based concordancing materials are also worthwhile exploring.

Last but not least, the data results add to our better understanding of how many new words students may learn with limited class hours and the extent to which their vocabulary can be expanded. The aim of this research has been to generate that awareness, to prepare ourselves to be in a position to make these transitions in enhancing students' English abilities.

Notes

¹ The twelve departments in connection with business and management in the researched school are the departments of accounting, finance, economics, money and banking, risk management and insurance, marketing, business administration, international trade, statistics, industrial engineering and management, information management, and tourism.

² The figures shown in the earlier issue of the *e-FLT Journal* (June, 2009) are reproduced here for illustration and comparison with the business English corpus.

³ Nation (2006) admitted that the disadvantage of the BNC HFWL is that the AWL is not separated from the frequency levels. The AWL vocabulary is spread from the 1st 1,000 words to the 10th 1,000 words.

⁴ There are 18 weeks per semester in universities and colleges in Taiwan. Excluding the midterm and final exam weeks, the actual number of class weeks is 16.

Acknowledgements

1. The researcher wishes to thank Professor Paul Nation, whose website containing high-frequency word lists and the RANGE program was indispensable to this study.
2. The researcher wishes to express her profound gratitude to the following publishers for their permission to reproduce copyright material for the current research.
Corpus use of content: From *Select Readings Intermediate Student's Book* and *Select Readings Upper-Intermediate Student's Book* copyright © 2004 Reproduced by permission of Oxford University Press; from *What a World 2*, *What a World 3* copyright © 2005, *NorthStar: Focus on Reading and Writing-high intermediate* and *NorthStar: Focus on Reading and Writing-advanced* copyright © 2004 Used by permission of Pearson Education, Inc.; from *Reading for Real High-Intermediate* and *Reading for Real Advanced* copyright © 2001 Reproduced by permission of Lynx Publishing; from *Reading for Success 2* and *Reading for Success 3* copyright © 2004 Reproduced by permission of Taiwan ELT; from *Interactions 1 & 2*, *Mosaic 1 & 2: Reading* copyright © 2007 Reproduced by permission of McGraw-Hill; from *Reading for the Real World 2*, *Reading for the Real World 3* copyright © 2004 Reproduced by permission of Compass.
3. The researcher also wishes to thank the National Science Council (NSC) in Taiwan for its research grant for the time-consuming corpus building and manual proofreading.

References

- Bauer, L., & Nation, I.S.P. (1993). Word families. *International Journal of Lexicography*, 6, 253–279.
- Brinton, D.M., Snow, M.A., & Wesche, M.B. (1989). *Content-based second language instruction*. New York: Newbury House.
- Cowan, J. R. (1974). Lexical and syntactic research for the design of EFL reading materials. *TESOL Quarterly*, 8(4), 389–400.
- Coxhead, A. (2000). A new academic word list. *TESOL Quarterly*, 34(2), 213–238.
- Farrell, P. (1990). Vocabulary in ESP: A lexical analysis of the English of electronics and a study of semi-technical vocabulary (*CLCS occasional paper No. 25*). London: Trinity College.
- Goulden, R., Nation, I.S.P., & Read, J. (1990). How large can a receptive vocabulary be? *Applied Linguistics*, 11(4), 341–363.
- Hwang, K. (1989). *Reading newspapers for the improvement of vocabulary and reading skills*. Unpublished M.A. thesis. Victoria University of Wellington, New Zealand.
- Hirsh, D. (1993). *The vocabulary demands and vocabulary learning opportunities in short novels*. Unpublished M.A. thesis. Victoria University of Wellington, New Zealand.
- Hsu, W. (2006). Pedagogical concerns: Some common features of CBI, TBL and BCS. In W. M. Chan, K. N. Chin & T. Suthiwan (Eds.), *Foreign language teaching in Asia and beyond: Current perspectives and future directions* (pp. 217–241). Singapore: Centre for Language Studies, National University of Singapore.
- Hsu, W. (2009). College English textbooks for general purposes: A corpus-based analysis of lexical coverage. *Electronic Journal of Foreign Language Teaching*, 6, 42–62. Retrieved June 11th, 2009, from <http://e-flt.nus.edu.sg/v6n12009/hsu.pdf>
- Hyland, K., & Tse, P. (2007). Is there an “Academic Vocabulary”? *TESOL Quarterly*, 41(2), 235–253.
- Krueger, M., & Ryan, F. (1993). Resituating foreign language in the curriculum. In M. Krueger & F. Ryan (Eds.), *Language and content: Discipline- and content-based approaches to language study* (pp. 316–323). Clevedon: Multilingual Matters.
- Laufer, B. (1989). What percentage of text lexis is essential for comprehension? In C. Lauren & M. Nordman (Eds.), *Special language: From human thinking to thinking machines* (pp. 316–323). Clevedon: Multilingual Matters.
- Leech, G., Rayson, P., & Wilson, A. (2001). *Word frequencies in written and spoken English*. Harlow: Pearson Educational.
- Nagy, W., Anderson, R., Schommer, M., Scott, J.A., & Stallman, A. (1989). Morphological families in the internal lexicon. *Reading Research Quarterly*, 24, 262–281.
- Nation, I.S.P. (2001). *Learning vocabulary in another language*. Cambridge: Cambridge University Press.
- Nation, I.S.P. (2006). How large a vocabulary is needed for reading and listening? *The Canadian Modern Language Review*, 63(1), 59–82.
- Nation, I.S.P. & Heatley, A. (2002). RANGE [Computer software]. Retrieved March 5, 2009, from <http://www.vuw.ac.nz/lals/staff/paul-nation/nation.aspx>
- Nation, I.S.P., & Waring, R. (1997). Vocabulary size, text coverage and word lists. In N. Schmitt & M. McCarthy (Eds.), *Vocabulary: Description, acquisition and pedagogy* (pp. 6–19). Cambridge: Cambridge University Press.

- Read, J. (2000). *Assessing vocabulary*. Cambridge: Cambridge University Press.
- Schmitt, N., & McCarthy, M. (1997). *Vocabulary, description, acquisition and pedagogy*. Cambridge: Cambridge University Press.
- West, M. (1953). *A general service list of English words*. London: Longman, Green, & Co.

Appendix A

Eight Business Core Textbooks

- Cecchetti, S.G. (2007). *Money, banking and financial markets*. Taipei, Taiwan: McGraw-Hill International.
- Ebert, R.J., & Griffin, R.W. (2007). *Business essentials*. Englewood Cliffs, NJ: Prentice Hall.
- Griffin, R.W., & Moorehead, G. (2007). *Management*. Boston, MA: Houghton Mifflin.
- Kotler, P., & Armstrong, G. (2008). *Principles of marketing*. Englewood Cliffs, NJ: Prentice Hall.
- Schotter, A. (2008). *Microeconomics: A modern approach*. Singapore: Cengage Learning.
- Starr, M. (2007). *Foundations of production and operations management*. Singapore: Cengage Learning.
- Weiss, N.A. (2007). *Introductory statistics*. Reading, MA: Addison-Wesley.
- Weygandt, J.J., Kieso, D.E., & Kimmel, P.D. (2007). *Accounting principles*. Singapore: John Wiley & Sons (Asia) Pte.

Appendix B

Twenty General English Textbooks

- Bernard, J., & Lee, L. (2004). *Select readings-upper intermediate*. Oxford: Oxford University Press.
- Broukal, M. (2005). *What a world 2: Amazing stories from around the globe*. New York: Pearson Education.
- Broukal, M. (2005). *What a world 3: Amazing stories from around the globe*. New York: Pearson Education.
- Constantino, M., & Dabbs, T. (2001). *Reading for real-high intermediate*. Vancouver, CA: Lynx.
- Constantino, M., & Dabbs, T. (2001). *Reading for real-advanced*. Vancouver, CA: Lynx.
- English, A.K., & English, L.M. (2004). *NorthStar: Focus on reading and writing-high intermediate*. New York: Addison Wesley Longman.
- Graber, B., & Babcock, P. (2004). *Reading for the real world 3*. Sachse, TX: Compass.
- Hartmann, P., & Kirn, E. (2007). *Interactions 2: Reading*. New York: McGraw-Hill.
- Kirn, E., & Hartmann, P. (2007). *Interactions 1: Reading*. New York: McGraw-Hill.
- Lee, L., & Gundersen, E. (2004). *Select readings-intermediate*. Oxford: Oxford University Press.
- Miller, J.L., & Cohen, R.F. (2004). *NorthStar: Focus on reading and writing-advanced*. New York: Addison Wesley Longman.
- Pavlik, C. (2006). *Hot topics 2*. Boston, MA: Thomson Heinle.
- Pavlik, C. (2006). *Hot topics 3*. Boston, MA: Thomson Heinle.
- Rashap, A., Thomas, V., & Edwards, W. (2004). *Reading for success 2*. Baltimore, MD: American ELT.
- Rashap, A., Tong, V., & Edwards, W. (2004). *Reading for success 3*. Baltimore, MD: American ELT.
- Wegmann, B., & Knezevic M. (2007). *Mosaic 1: Reading*. New York: McGraw-Hill.
- Wegmann, B., & Knezevic M. (2007). *Mosaic 2: Reading*. New York: McGraw-Hill.
- Wholey, M.L., & Henein, N. (2007). *Reading Matters 3: An integrated approach to reading*. New York: Houghton Mifflin.
- Wholey, M.L., & Henein, N. (2002). *Reading Matters 4: An integrated approach to reading*. New York: Houghton Mifflin.
- Zwier, L.J., & Stafford-Yilmaz, L. (2004). *Reading for the real world 2*. Sachse, TX: Compass.