International Journal of
Strategic Research in Education, Technology and Humanities
p-ISSN: 2465-731X | e-ISSN: 2467-818X
IJSRETH
April, 2024
Vol. 12, No. 1

Vocabulary Thresholds Study for Business English Major Based on the Range Corpus

[1]Wang Run & [2]Li Weina
[1&2]Department of English, Beijing International Studies University, Beijing, China

Article DOI:
10.48028/iiprds/ijsreth.v12.i1.02

Abstract

The texts have been selected from 45 articles of 15 units of Selected Readings of Business English and from prominent economic English newspapers in China and the United States—China Economic Net and The Economist. Employing the vocabulary corpus software, Range 32, we conducted a statistical examination of vocabulary coverage and the frequency of corresponding terms in the British National Corpus (BNC). This rigorous analysis aims to ascertain whether the content of the textbook aligns with the requisites of specific professional tests. Additionally, it serves to assist students and learners of Business English in China by aiding them in identifying suitable English-language reading materials. This research provides a valuable reference, guiding individuals toward reading newspapers that align with their vocabulary proficiency in English. The study findings indicate that to attain a fundamental comprehension of the reading materials, achieving a 95% vocabulary coverage requires a command of 4,000 to 5,000 vocabulary words for Selected Readings of Business English, 4,500 to 5,000 vocabulary words for China Economic Net, and 5,000 to 6,000 vocabulary words for The Economist. For a more fluent reading experience, corresponding to a 98% vocabulary coverage, a mastery of 7,500 to 8,000 vocabulary words is necessary for Selected Readings of Business English, 8,500 to 9,000 vocabulary words for the China Economic Net, and 10,000 to 11,000 vocabulary words for The Economist. And this study suggests that the vocabulary thresholds for Business English majors should be 4,000 to 6,000 (95% vocabulary coverage) and 8,000 to 11,000 (98% vocabulary coverage).

Keywords:
Vocabulary Threshold, Business English, Selected Readings of Business English, China Economic Net; The Economist

Corresponding Author:
Wang Run

First Published:
http://i-jeh.com/index.php/ijeh/index

Background to the Study

As the "Belt and Road Initiative" deepens and global trade strengthens, proficiency in business English skills has become an indispensable necessity. Mastery of business English reading skills, in particular, is now a vital criterion for business English majors. Beyond conventional textbook instruction, engaging with business newspapers serves as a pivotal avenue for business English majors to acquire language proficiency and gain insights into financial and current affairs. The core of effective text comprehension lies in vocabulary mastery, with its direct impact on the fluidity of reading and depth of understanding. Presently, the vocabulary requirements for business English majors remain unclear, leading to a lack of precision in determining whether relevant materials align with the teaching standards and vocabulary proficiency levels expected of business English majors. Therefore, it is imperative to elucidate the vocabulary thresholds embedded in business English textbooks and associated learning materials.

Vocabulary, as one of the elements of language, has been the object of research by many linguists in the study of business English teaching. Vocabulary has long been regarded as an important determinant of reading success (Biemiller, 2003). Wilkins (1972) argued that without grammar, people can only convey very little information; without vocabulary, people cannot express anything. Therefore, adequate vocabulary is an important foundation for reading comprehension and the most direct and effective way for English learners to acquire language knowledge. However, if the reading material is too difficult, it will make the readers lose interest in reading and frustrate their motivation. Of course, if the reading is too easy, there is little room for learning improvement (Zhao, 2013). Therefore, for the preparation of business English reading materials, we should find reading contents more suitable for students' vocabulary, and write them in a targeted way according to the different vocabularies of the materials; business English learners with different vocabularies should find economic reading materials more suitable for their own vocabularies to read, instead of choosing the reading materials blindly, so that the effect of the improvement of the reading level will be more significant. In recent years, people have paid more and more attention to English vocabulary.

However, so far, there are still fewer studies on business English professional vocabulary (Mao, 2016; Li, 2023; Wang & Lian, 2013), especially through vocabulary size (Li, Yu& Lin, 2015; Lin, Li & Yu, 2018) and there are almost no studies on argumentation and discussion from newspapers in combination with textbooks. Therefore, starting from 45 articles in 15 units of Selected Readings of Business English and business newspapers in China and the United States -China Economic Net and The Economist. This study analyses the word family size of the two newspapers using the lexical corpus Range32 software, based on the 14 basic word lists that come with the BNC corpus and lexical coverage. The study shows that a vocabulary coverage of more than 95% enables readers to basically understand what they are reading, and a vocabulary coverage of 98% enables readers to read easily. On this basis, this study examines what vocabulary base business English textbooks should be written, and how much vocabulary business English

learners need to understand and read these two business newspapers easily. It is hoped that this study will help to deepen the understanding of the relationship between vocabulary coverage and word families and make an important contribution to the field of business English textbook writing and the study of vocabulary learning and reading comprehension for business English learners.

Literature Review

Vocabulary

Based on Paul Nation's Range software settings, all references to vocabulary throughout the text should refer to the number of word families. Word family is composed of a basic word or root, its inflectional affixes, and derived affixes. They have the same rhyme, including vowels and subsequent consonants. Once a base word or even a derived word is known, the recognition of other members of this family requires little or no extra effort. Therefore, the words of the same family can be understood together by a learner without having to learn each form separately (Laufer & Nation, 1993; Coxhead, 2000). What counting unit should be used to judge the number of words? The counting unit of a word is selected according to the purpose of statistics. Some researchers chose a vocabulary list represented by Academic Vocabulary List (Gardner & Davies, 2014), and some researchers chose to use General Service List (West, 1953), Academic Word List (Coxhead, 2000), Medical Word List (Hsu, 2013), and Academic Spoken Word List (Dang, Coxhead, & Webb, 2017), etc. which represent vocabulary lists (Liu & Lei, 2020). In this paper, the main research is to study how large a vocabulary is needed to read Business English materials, in order to help business English textbook compilers be more precise in materials and to help learners find some newspapers that suit their vocabulary.

Therefore, in terms of receptive reading, using word family as the unit of word measurement is a good choice. Choosing word families as the unit of counting is that once learners have mastered the stem word, with knowledge of basic word-building processes, they will infer the meanings of regularly inflected and derived forms of that word without much effort (Laufer & Nation, 1993). Hirsh and Nation (1992) proposed that if the learner has the necessary knowledge of affixes, then the inflection or derivative form of a word is also considered a known vocabulary, based on which the vocabulary can be measured by the amount of word families. Therefore, using word family as a counting unit is an effective way for vocabulary teaching and memory. In the 1940s and 1950s, vocabulary began to be regarded as an important part of reading skills, which could be seen in the reports of Coleman and English language teaching expert Michel West. In 1953, West revised and published "A General Service List of English Words", which listed 2,285-word families frequently used in reading and writing and provided extremely detailed word frequency information. It received widespread attention in the 1960s and 1970s.

Some researchers believed that there was a vocabulary threshold in the reading process (that is, the increase in reading ability below this minimum vocabulary was not obvious), and the minimum vocabulary range was about 2,000 ~ 5,000word families, which is

mainly reflected in the research of Laufer and Nation (1993). Laufer (1989: 321) asserted "5,000 words seem to be the lexical threshold beneath which other facilitating factors in reading comprehension may not be very effective". The research of Hirsh and Nation (1992) showed that the most frequent 2,000 words didn't provide adequate coverage for pleasurable reading and that a word family size of around 5,000 would be needed to do this. Han and Geng (2007) believed that only after proficiently mastering 3,000 words (4,800 lexical items), could learners initially develop strong English reading ability and independently solve the general language problems encountered in English reading; if the familiar words reached 5,000 words (8,000 lexical items), readers could read the original English more freely, and vocabulary knowledge would gradually become a secondary factor restricting reading comprehension. Other scholars believed that 10,000-word families were the most basic requirement if foreign language learners wanted to read medium-difficulty articles fluently (Koda, 1989; Laufer, 1991).

Besides, some scholars (Gao, 2015; Quinn, Wagner, Petscher, & Lopez, 2015) have proved through experimental research that there is a positive correlation between vocabulary and reading comprehension (that is, as the number of word families mastered increases, his ability to read and comprehend (word coverage) also increases). Gao (2015) conducted a vocabulary size and reading comprehension test with 15 self-examined students in the English major of the full-time self-examination college as the research object. The research results showed that there was a significant positive correlation between the two; vocabulary size had a positive impact on reading comprehension ability, that is, the larger the vocabulary, the better the reading comprehension performance, and vice versa. Quinn et al. (2015) explored the potential developmental coupling of vocabulary and reading comprehension through the use of a latent change scoring model, and the results showed that improvements in reading comprehension do depend in part on vocabulary knowledge.

Lexical Coverage

The study of lexical coverage plays a very important role in second language acquisition and Business English learning. Nation (2006) defined lexical coverage as the ratio of the vocabulary that readers knew to the total number of words in the text. It is the best gauge and measure of whether a text is likely to be adequately understood and refers to the percentage of words in the text from a specific vocabulary. Most researchers believed that in order to have an overall understanding of the article, it was necessary to master 95% of all vocabulary and to fully understand it, at least 98% of the vocabulary coverage was required. Laufer (1989) suggested that lexical coverage of 95% could ensure reasonable reading comprehension. The research of Laufer (1997) showed that second-language readers needed to master 95% of all vocabulary in the article to get a general understanding of the article, and 98% or more of the vocabulary in the article could be fully understood. The study of Hu and Nation (2000) concluded that around 98% coverage of vocabulary was needed for learners to gain unassisted comprehension of a fiction text. Schmitt et al. (2011) suggested that 98% of lexical coverage was a more reasonable coverage target for readers of academic texts. In the realm of vocabulary

research, scholars have widely debated the correlation between vocabulary coverage and reading comprehension ability, and there exists a distinct boundary regarding the nature of this correlation. He and Hao (2012) demonstrated a direct relationship between vocabulary coverage and the level of reading comprehension, emphasizing the pivotal role vocabulary plays in comprehending written texts. Adding to this discourse, Zhang and Liu (2019) conducted an exploration into the relationship between the breadth of vocabulary knowledge and the reading comprehension skills of college English learners. The results of their study unequivocally indicated a significant and positive correlation between vocabulary knowledge breadth and reading comprehension, reinforcing the notion that a robust vocabulary foundation is intricately linked to enhanced reading comprehension abilities in the context of language learning. Some foreign vocabulary researchers believed that there existed a relatively linear relationship between the percentage of known vocabulary and the degree of reading comprehension (Schmitt et al., 2011). Ludewig, Hübner, & Schroeder (2023) showed that text coverage, especially above 56% text coverage, will better aid text comprehension, while reading below 56% text coverage will degrade the reader's reading experience and is unsuitable for instructional use. American scholars Francis and Kucera (1982) once found that 4,000-word families corresponded to 86.8% of text lexical coverage. To read novels and have coverage of 95% ~ 98% of the high-frequency words, a learner would need around 5,000-word families (Hirsh & Nation, 1992).

As for the study of business English textbooks and academic journals, Hsu (2011) compiled his own corpus of business core course textbooks as well as a corpus of business research articles, both of which are more than 7 million words, and found that the vocabulary thresholds of business textbooks were 3,500 (95%) to 5,000 (98%) word families, while the vocabulary thresholds of business research articles were 5,000 (95%) to 8,000 (98%) word families. This paper then compares the scope of research based on Hsu's business English textbook. Nation & Waring (2002) proposed that when the vocabulary size reached 2,000word families, readers could understand 79.7% of the text vocabulary; when the vocabulary size level reached 3,000word families, readers could understand 84% of lexical coverage; and when it reached 15,851word families, their comprehensible text lexical coverage was 97.8%. Nation 2006) selected 14,000-word families based on the BNC (British National Corpus) and concluded that 8,000 ~ 9,000word families are needed to comprehend written text if one can easily read and comprehend 98% of the article coverage. Nation has also studied the vocabulary thresholds for reading newspapers, and the result is that when the vocabulary coverage is 95% (1 unfamiliar word in 20words), the corresponding vocabulary size is 4,000 ~ 5,000word families; when the vocabulary coverage is 98% (1 unfamiliar word in 50 words), the corresponding vocabulary size is 8,000 ~ 9,000word families, based on which this paper conducts a comparison of the vocabulary of newspapers.

Summary

Existing foreign research predominantly focuses on word lists, textbooks, and experimentally designed databases, exploring vocabulary and vocabulary coverage

(Hirsh & Nation, 1992; Laufer & Ravenhorst-Kalovski, 2010; Nation, 2006). However, the research scope often encompasses broad perspectives, with limited detailed studies and a notable absence of investigations into the vocabulary specific to business English majors and Business English newspapers. Hsu's work (2011) stands out in this regard, although it lacks an analysis of specific newspaper vocabulary thresholds. Conversely, domestic scholars such as Wang (2017) primarily concentrate on examining vocabulary requirements within textbook syllabi, aiming to provide guidance for syllabus revision in second language teaching. The majority of domestic scholars contribute to the understanding of syllabus adequacy by comparing vocabulary mastery rules across different syllabus stages. Furthermore, researchers delve into English test question analysis to offer insights and strategies for test-takers. However, a limited number acknowledge the existence of a vocabulary threshold for reading (Gui et al., 2020; He & Hao, 2012; Qin & Yang, 2009; Lin et al., 2018; Li et al., 2015). As the research landscape on the correlation between reading and vocabulary evolves globally, the growing body of literature provides valuable data and references, offering substantial support for the focus of this paper.

Methods

Research Questions

This study seeks to furnish Business English learners and textbook writers with a valuable reference, offering insights to enable textbook authors to tailor their materials more effectively to students' writing proficiency levels. Additionally, the study aims to assist business English majors in identifying business-oriented newspapers that align better with their vocabulary proficiency. The research is primarily dedicated to addressing three key questions:

1. How much vocabulary does Selected Readings of Business English cover under the standard?
2. How much vocabulary is needed to read China Economic Net?
3. How much vocabulary is needed to read The Economist?

Research Materials

In this study, we selected a domestic business English reading textbook Selected Readings of Business English, and two prominent English-language newspapers renowned for their readership in both China and the United States: China Economic Net and The Economist. To ensure data accuracy, our research constructed a dedicated corpus by downloading approximately 63,000 words from 45 reading articles spanning 15 units of the textbook Selected Readings of Business English. These articles were sourced from the official CXStar platform. Additionally, we acquired content from the two newspapers—China Economic Net (http://en.ce.cn/) and The Economist (https://www.economist.com/). Each newspaper contributed approximately 105,000 words to the research text. Given variations in word counts among articles in the two newspapers, 50 articles from each of the thematic categories (China, World, Business, Life, and Insight) were selected from China Economic Net. For The Economist, 138 articles were randomly chosen from two issues dated October 28, 2023, and December 23, 2023. The selected articles covered diverse themes, including the latest news in 2023.

Table 1: Words Collection of the Two Newspapers' Text

No	Type	Name	Selection	Total Words
1	Textbook	Selected Readings of Business English	45 articles	63082
2	Newspaper	China Economic Net	250 articles	104489
3	Newspaper	The Economist	138 articles	104994

Among them, due to some emerging proper nouns as well as some compound words Range software cannot identify, so this study carried out corpus processing, self-constructed proper noun word list basewrd17, and some various forms of by two relatively simple belonging to the first 16word lists can be recognized by the word list of the composite word to split, so that Range can be accurately identified, and will not be treated as an unrecognizable raw word processing. The final word count was recognized by Word, and a total of 63082 words were selected from Selected Readings of Business English, 104489 words from China Economic Net, and 104994 words from The Economist.

Research Instruments

The software of vocabulary analysis—Range32 is a corpus software used to analyze the range and depth of text vocabulary. It was designed by two linguists—Professor Nation and Coxhead from Victoria University and written by Heatley. It comes with 16 basic vocabularies provided by Nation. 1-14word lists are based on the frequency and distribution of the BNC corpus and all include 1,000word families (except for the second basic vocabulary which has 998word families). BNC is the most authoritative and largest British English corpus nowadays. It uses written and spoken languages from a wide range of sources as samples. The word capacity exceeds 100 million, of which written corpus accounts for 90%, which is very suitable for analyzing the newspaper texts we are studying. The 15th basic vocabulary is a proper noun vocabulary, which has 13,535word families, and the 16th basic vocabulary is a list of interjections and modal particle vocabularies, with only 4word families (AH, HA, AW, OH).

Besides, when Range software and 16 basic vocabularies were used to analyze China Economic Net and The Economist, a new proper noun vocabulary based on the vocabulary in the text was compiled that was not in the vocabulary and inserted the name basewrd17, a total of 2,419word families. Therefore, the final word list used in this study is seventeen. The software is a very powerful program that can run 32 texts at the same time and compare the usage of different text vocabulary. Moreover, the speed is so fast that it can be completed in less than one second. Most researchers will use it to study the vocabulary of teaching materials to study the correlation between vocabulary and writing quality or English reading, thereby assisting English.

Research Procedures

The research procedure is mainly divided into three parts: text collection, text preprocessing, and final statistical analysis. This study first selects the materials about

6,3000 words from the textbook Selected Readings of Business English and randomly selects several issues of newspapers (including only the title and text) from the official websites of China Economic Net and The Economist, each with a total of about 105,000 words, and then copies and pasts them into the text file and summarizes them.

Text preprocessing. Firstly, this study uses Range software to conduct statistical analysis on the summarized original texts of newspapers and selects hyphenated vocabulary, compound words, acronyms, and proper nouns from the "not in the lists" of the statistical results, and then generates proper noun vocabulary for Selected Readings of Business English, China Economic Net and The Economist. Secondly, the "hyphens" in some compound words are replaced with "spaces" in batches (such as quake-hit, sun-synchronous, new-generation, remote-sensing, home-based, at-home, worst-ever, right-wing, counter-offensive, etc.). These words individually belong to the first 14 basic vocabularies, but because of the hyphens, RANGE will count them as a whole and exclude them from the basic vocabularies. In addition, some compound words and phrases (such as spokesperson, hydropower, waterproof, etc.) are divided into two independent words, because they also belong to the 14 basic vocabularies individually, and both are familiar to everyone in daily use.

After that, some of the URLs in the original text are deleted, and after the processing is completed, multiple analysis experiments are carried out. Then the researcher manually filters out proper nouns that overlap with other 16 base word lists, generating the final version of the proper noun vocabulary (including e.g., COVID-19, Shenzhou-17, Changshu, Chongqing, etc.)—basewrd17. Final statistical analysis. According to the adjusted and supplemented 17 vocabularies, the Range software is used to conduct a final statistical analysis of the textbook and two newspapers text. The following statistical results are the final data used in the study. Because the RANGE software counts the total number of tokens, which is different from the word count method of the software, the total number of tokens of the two newspapers finally shown in the table is also somewhat different.

Table 2: Words Collection of the Textbook and Newspaper Text

No	Materials	Total Words	Total Tokens
1	Selected Readings of Business English	63082	61483
2	China Economic Net	104489	103574
3	The Economist	104994	105246

Findings and Discussion

According to the 17 vocabularies mentioned above, the researcher uses Range software to perform statistical analysis on the 63,000-word text of the textbook and the 105,000-word text of the two newspapers, which can be seen from the data in Table 3, Table 4, and Table 5.

Vocabulary of Selected Readings of Business English

Table 3 shows the combined statistical results of Selected Readings of Business English. In the header of Table 3, the first column is the word list, which represents the serial numbers of the basic word list from 1 to 17 and the words that aren't in the list; the second column is token, which refers to the number of words in each list in the merged text, and the third column is token percent; the fourth column is type, which refers to tokens that are not calculated repeatedly. That is, if multiple identical tokens appear repeatedly, only one type can be counted. For example, boy is boy. There are three tokens, but only two types: boy and is. The fifth column is the percentage of types; the sixth column is the number of word families; the seventh column is the cumulative percentage of the number of tokens (excluding proper nouns and interjections); the data in this column are the cumulative addition of the percentage of tokens in the second column; the eighth column is the cumulative addition of the percentage of tokens (including proper nouns and interjections), which can be obtained by adding each data in the fifth column to the percentage of tokens in the combined text of the three word lists of the 15th, 16th, and 17th. The data in the last row is the sum of all the data in the column. Besides, based on this table, we can infer how many word families correspond to different vocabulary coverage, and from this we can know how much vocabulary is needed in order to understand Selected Readings of Business English.

Table 3: Statistical Results of Selected Readings of Business English

WORD LIST	TOKENS	/%	TYPES	/%	FAMILIES	Character percentage accumulation (excluding proper nouns and interjections)	Character percentage accumulation (including proper nouns and interjections)
1	46217	75.17	2560	31.08	938	75.17	78.89
2	6227	10.13	1584	19.23	786	85.30	89.02
3	1775	2.89	730	8.86	498	88.19	91.91
4	1697	2.76	643	7.81	436	90.95	94.67
5	787	1.28	387	4.70	295	92.23	95.95
6	481	0.78	267	3.24	214	93.01	96.73
7	415	0.67	206	2.50	180	93.68	97.4
8	392	0.64	173	2.10	148	94.32	98.04
9	233	0.38	117	1.42	101	94.70	98.42
10	163	0.27	117	1.42	106	94.97	98.69
11	133	0.22	83	1.01	80	95.19	98.91
12	87	0.14	65	0.79	62	95.33	99.05
13	108	0.18	70	0.85	65	95.51	99.23
14	72	0.12	51	0.62	47	95.63	
15	780	1.27	371	4.50	371		
16	3	0.00	2	0.02	1		
17	1509	2.45	519	6.30	519		
Not in the lists	404	0.66	293	3.56	—		
合计	61483	100.01	8238	99.99	4847		

*RANGE doesn't count out-of-vocabulary words.

The answer to the first question can be acquired from Table 3, which shows that if the Business English majors want to roughly understand *Selected Readings of Business English*, they only need 5,000word families (92.23%) plus the coverage of proper nouns and interjections (3.72%) to reach 95% vocabulary coverage, reaching 95.95%. 4,000word families can only reach 94.67%, which is still some distance from 95%. Therefore, if they want to understand *Selected Readings of Business English*'s content basically, 4,000 ~ 5,000word families are needed. This result is nearly 500 ~ 1000word families higher than Hsu's (2011) study of 3,500 (95%).

If Business English learners want to understand the content of *Selected Readings of Business English* easily, that is, to reach a vocabulary recognition rate of 98%, 8,000word families (94.32%) plus proper nouns (3.72%) are needed, and their coverage rate in *Selected Readings of Business English* text reaching 98.04 %. The coverage of 7,000word families plus proper nouns and interjections in the text is 97.4%, which is a little bit far from 98%. Therefore, learners need about 7,500 ~ 8,000word families to read *Selected Readings of Business English* to read fluently. This result is significantly higher than Hsu's (2011) study by 5,000 (98%) and nearly 3,000word families. The researcher speculates that this phenomenon is due to the fact that Hsu's (2011) study has a wider scope and a larger corpus, which includes many basic business English textbooks, whereas the objects used in this paper are business English reading books, and most of the content is extracted from *The Economist*, which has a larger vocabulary requirement, and the overall base of the corpus is smaller and more difficult, which is an advanced expansion of the material in the business English textbooks, and so the result is higher than Hsu's.

Vocabulary of China Economic Net

Table 4 shows the combined statistical results of *China Economic Net* and the content of the header in Table 4 is the same as that in Table 3. Based on this table, we can infer how many word families correspond to different vocabulary coverage, and from this, we can know how much vocabulary is needed in order to understand *China Economic Net*.

Table 4: Statistical Results of *China Economic Net*

WORD LIST	TOKENS	/%	TYPES	/%	FAMILIES	Character percentage accumulation (excluding proper nouns and interjections)	Character percentage accumulation (including proper nouns and interjections)
1	69827	67.42	2613	27.02	933	67.42	72.41
2	13695	13.22	1714	17.72	779	80.64	85.63
3	3715	3.59	852	8.81	511	84.23	89.22
4	4258	4.11	726	7.51	449	88.34	93.33
5	1940	1.87	485	5.01	337	90.21	95.20
6	1018	0.98	335	3.46	256	91.19	96.18
7	786	0.76	240	2.48	188	91.95	96.94
8	627	0.61	212	2.19	166	92.56	97.55
9	600	0.58	167	1.73	134	93.14	98.13
10	243	0.23	116	1.20	103	93.37	98.36
11	307	0.30	97	1.00	89	93.67	98.66
12	156	0.15	63	0.65	59	93.82	98.81
13	280	0.27	80	0.83	68	94.09	99.08
14	146	0.14	51	0.53	48	94.23	99.22
15	1777	1.72	402	4.16	402		
16	5	0.00	1	0.01	1		
17	3387	3.27	1177	12.17	1174		
Not in the lists	807	0.78	340	3.52	—		

*RANGE doesn't count out-of-vocabulary words.

The answer to the second question can be acquired from Table 4, which shows that if the readers want to roughly understand *China Economic Net*, they only need 5,000-word families (90.21%) plus the coverage of proper nouns and interjections (4.99%) to reach 95% vocabulary coverage, reaching 95.20%. While 4,000-word families can only reach 93.33%, which is still a long way from 95%. Therefore, if they want to understand *China Economic Net*'s content basically, 4,500 ~ 5,000-word families are needed (Hirsh & Nation, 1992; Laufer, 1991). This study is basically consistent with the findings of these researchers.

And if English learners want to understand the content of this newspaper easily, that is, to reach a vocabulary recognition rate of 98%, 9,000-word families (93.14%) plus proper nouns (4.99%) are needed, and their coverage rate in *China Economic Net* text reaching

98.13 %. The coverage of 8,000-word families plus proper nouns and interjections in the text is 97.55%, which is close to 98%. Therefore, business English learners need about 8,000 ~ 9,000-word families to read *China Economic Net* to read fluently (Li & Yu, 2018; Lin et al., 2018; Nation, 2006). This result is almost identical to the research results of Nation's 8,000 ~ 9,000-word families (2006).

Vocabulary of The Economist

The statistical results of *The Economist* have shown in Table 5 and the content of the header in Table 5 is the same as that in Table 3.

Table 5: Statistical Results of *The Economist*

WORD LIST	TOKENS	/%	TYPES	/%	FAMILIES	Character percentage accumulation (excluding proper nouns and interjections)	Character percentage accumulation (including proper nouns and interjections)
1	76204	72.41	3082	22.70	976	72.41	76.36
2	10010	9.51	2297	16.92	936	81.92	85.87
3	3497	3.32	1353	9.96	771	85.24	89.19
4	3498	3.32	1109	8.17	664	88.56	92.51
WORD LIST	TOKENS	/%	TYPES	/%	FAMILIES	Character percentage accumulation (excluding proper nouns and interjections)	Character percentage accumulation (including proper nouns and interjections)
5	1739	1.65	792	5.83	526	90.21	94.16
6	1258	1.20	581	4.28	422	91.41	95.36
7	912	0.87	405	2.98	321	92.28	96.23
8	851	0.81	380	2.80	305	93.09	97.04
9	547	0.52	299	2.20	255	93.61	97.56
10	359	0.34	226	1.66	207	93.95	97.90
11	333	0.32	221	1.63	190	94.27	98.22
12	260	0.25	180	1.33	158	94.52	98.47
13	280	0.27	161	1.19	146	94.79	98.74
14	190	0.18	104	0.77	99	94.97	98.92
15	2356	2.24	730	5.38	729		
16	8	0.01	4	0.03	2		
17	1790	1.70	874	6.44	874		
Not in the list	1154	1.10	780	5.74	—		
合计	105246	100.00	13578	100.01	7581		

*RANGE doesn't count out-of-vocabulary words.

Researchers such as Nation (2006) believed that, because proper nouns are easy to recognize and do not need to be learned in advance, that is, the burden of learning is

small, proper nouns are usually classified as vocabulary that readers already know. This study also classifies proper nouns (the basewrd15 and basewrd17) into vocabulary recognized by readers in China because proper nouns are usually composed of human names, place names, time, acronyms, etc., which are convenient for identification and recognition. In addition, there are 4-word families (AH, AW, HA, OH) of interjections and modal particles appearing in the results of this research, so this research also classified the basewrd16 (interjections and modal particles) as this type of vocabulary. From Table 5 and Table 6, it can be seen that the sum of the percentages of the 15th, 16th, and 17th word lists in the *China Economic Net* research text is 4.99%, and 3.95% in *The Economist*, which is also a very high proportion.

The answer to the third question can be acquired from Table 5, which shows the combined statistical results of *The Economist*. As can be seen from it, to basically understand *The Economist*, that is, to achieve 95% lexical coverage requires 6,000 word families (91.41%) plus proper nouns and interjections (3.95%), reaching 95.36% lexical coverage; and 5,000 word families reaching 94.16% lexical coverage, so if English learners want to roughly understand *The Economist*, 5,000 ~ 6,000 word families are needed, which is consistent with the research results of Hu and Nation (2000). If they want to understand the content of the newspaper almost completely, 10,000-word families (93.95%) plus proper nouns (3.95%) are needed, and its coverage rate is 97.90%, which basically meets the requirement of 98%; when 11,000-word families (94.27%) are added with specific nouns (3.95%), the vocabulary coverage rate reaches 98.22%. It can be concluded that second language learners need approximately 10,000 ~ 11,000-word families to read *The Economist*, which is consistent with the research of Hsu (2011) and Li (2013). Compared with the research result of Nation (2006), this result is much higher than its standard, about 2,000-word families.

Vocabulary Comparison of the Textbook and Two Newspapers

From the results of the above study, it can be found that the overall vocabulary requirement of the business English reading textbook *Selected Readings of Business English* is relatively lower than that of the two newspapers. This is due to the fact that the textbook is needed to enable business English students to learn the knowledge points, so it is not possible to select reading articles that are too difficult and require too much vocabulary, which will increase the difficulty of students' reading comprehension, thus losing their interest in learning and probably failing to learn the knowledge points they should have learned.

Among the two newspapers, *China Economic Net* also requires a lower vocabulary than that of *The Economist*, which is about 2500 ~ 3000 words. This is due to the fact that *China Economic Net*, as an English-language newspaper in China, sources articles from domestic newspapers, and its contents are mostly about the economic situation in China, so that the audience groups and the writers may still be Chinese, whereas *The Economist*'s economic news is spread all over the world, and its main audiences and writers are mostly English-speaking people. *The Economist*'s economic news covers the whole world,

and the main audience groups and writers are mostly English language speakers. As a native language, the articles will be written in a more authentic way, and the vocabulary will be more difficult because of the wider scope of coverage and more events.

Given that the teaching syllabi for Business English programs and related examinations, such as the Business English Certificate (BEC), do not explicitly outline specific vocabulary requirements, many learners in the field of Business English find themselves uncertain about their proficiency levels. This uncertainty extends to determining the suitable material for study and discerning the appropriate level of examination for their skill set. The official BEC website categorizes proficiency into Preliminary, Vantage, and Higher levels, drawing a parallel with benchmarks such as CET4 or CET6 and TEM4 or TEM8. Broadly speaking, the BEC Preliminary aligns roughly between the proficiency spectrum of CET4 to CET6. Correspondingly, BEC Vantage is positioned within the range defined by CET6 and TEM4. Meanwhile, BEC Higher is envisaged to cover the proficiency domain situated between TEM4 and TEM8. It is crucial to note that due to BEC's specialized focus on evaluating specific business English competencies, an overall increase in difficulty is warranted when compared to general language proficiency assessments.

In the newly revised College English Syllabus in 1999, the vocabulary required for Level 4 is 4,200 words, and the vocabulary required for Level 6 is 5,500 words. Huang and Xu (2003) believed that English majors should have a vocabulary of 8000 in TEM-4 and 12,000 ~ 13,000 in TEM-8. Accordingly, BEC Preliminary should have a vocabulary of about 5,000 words, BEC Vantage should have a vocabulary of about 6,000 ~ 8,000 words and BEC Higher should have a vocabulary of about 10,000 words, with a certain proportion of business vocabulary. Besides, the CATTI test syllabus shows that at least 5000 words are required for Level 3 and at least 8000 words for Level 2, which is similar to the requirement of BEC Preliminary and BEC Vantage.

Tables 3, 4, and 5 collectively demonstrate that the coverage of 2,000 high-frequency word families in both the textbook and the two newspapers exceeds 80%. This finding aligns with the research outcomes of Nation & Warning (1997), Ward (1999), and Lin et al. (2018). Consequently, scholars recommend that proficiency in business English necessitates a foundational mastery of 2,000 to 3,000 high-frequency word families (Nation, 2001; Gui, 2006).

The significance of high-frequency words is underscored, and the researcher advocates that, for enhanced proficiency in Business English, learners should prioritize acquiring an additional 2,000 to 5,000 intermediate-frequency business English words (Lin et al., 2018). This assertion reinforces the notion that a comprehensive command of both high-frequency and intermediate- frequency vocabulary is paramount for Business English learners aspiring to enhance their language proficiency in the domain.

Conclusion

This study focuses on the textbook Selected Readings of Business English and two newspapers, namely China's China Economic Net and America's The Economist, using them as samples. The analysis is based on 95% and 98% lexical coverage rates, incorporating 16 foundational word lists developed by Nation, along with the addition of the 17th proper noun list. The Range32 text analysis software is employed to investigate the requisite vocabulary for compiling suitable textbooks by business English teachers or for facilitating ease in reading for business English learners. The findings reveal that business English learners possessing a vocabulary of 2,000-word families can comprehend 80% of the content within the textbook and the two newspapers. To achieve a basic understanding of the Selected Readings of Business English, a mastery of 4,000-to-5,000-word families is necessary, while a more comprehensive understanding is facilitated with 7,500-to-8,000-word families. Similarly, for China Economic Net, a rudimentary understanding requires 4,500-to-5,000-word families, while fluent reading necessitates 8,000-to-9,000-word families. In the case of The Economist, basic comprehension entails 5,000-to-6,000-word families, whereas fluent reading demands approximately 10,000-to-11,000-word families. Therefore, for Business English majors, if they want to have a basic understanding of Business English textbooks or prepare for the related qualification exam (95%) at preliminary or vantage level, they need to master at least 4,000-to-6,000-word families; while if they want to read related materials fluently or prepare for the advanced qualification exam, they need to master at least 8,000 to 11,000 vocabularies.

To a certain extent, the data provide support for Business English teachers and learners. However, there are some limitations worth noting. Firstly, although this paper is researched based on 105,000 words per newspaper and 63,083 words of the textbook, it is only a small part of its total word number. Therefore, even if it is more innovative and accurate than other articles that only use a few thousand words as the research material, the statistical materials still need to be further expanded to make the research results more accurate. What's more, although English learners knowing a basic word can recognize other words in its word family, the meanings of words in the same word family will also be very different. If the readers know a base word, they may not necessarily recognize all its derivative words, which are easy to garble. This brings certain difficulties to English learners in the acquisition process. For example, if English learners know that "value" is "belief" or "worth", they can guess that "valuable" is "very important", but "invaluable" is not "valueless", but "extremely valuable or important". Therefore, some words that are similar or in the same word family relationship but have different meanings require readers to pay special attention. Thirdly, for business English learners, there are many words in the common basic vocabulary that have different meanings and a set of special names in Business English field. For example, "action" is not "the process of doing things in a broad sense", but "legal process to make someone pay for a mistake"; "party" is not "a social occasion", but "a person involved in a legal argument"; "offer" is not "giving suggestion", but "making an invitation". Although these words can be recognized in the BNC vocabulary list, they are the most basic meanings, which are very

easy to be confused by Business English students, so some basic Business English vocabulary is the first thing that Business English learners must master.

This research also raises some questions worthy of discussion for other future studies: Firstly, considering the limitation in the size of the data sample, it prompts the need for exploring larger datasets to enhance the comprehensiveness of the study. Examining a broader range of words in diverse contexts could provide a more nuanced understanding of language acquisition and usage. We can continue to explore the vocabulary and vocabulary coverage of other business English textbooks and business newspapers, so that we can help more business English learners with different vocabulary levels and provide references for textbook compilers. Secondly, the challenge posed by words in the same family having different meanings calls for a deeper investigation into the complexities of word relationships. Future studies could explore the cognitive processes involved in distinguishing between related words and the strategies employed by learners to overcome such challenges. Understanding the nuances within word families is crucial for effective language acquisition. Besides, with the continuous progress and development of the times, the vocabulary is constantly updated. Some low-frequency English vocabulary from a dozen years ago has become a household name nowadays, for example, words such as “CPU”, “dataset”, “GDP”, “tech”, “ShenZhou-17”, “e-commerce”, etc. The above-mentioned words that belong to unknown words in the BNC's 14-word lists become popular. Therefore, many new vocabulary corpora, especially business English lexical corpora, need to be researched and updated urgently. Fourthly, the unique challenges faced by business English learners in navigating words with multiple meanings and specialized terminology highlight the importance of tailoring language learning materials to meet their specific needs. Future research could delve into developing targeted instructional approaches that address the distinct vocabulary demands of business English, ensuring learners can confidently navigate the diverse meanings associated with common words in a professional context. Despite the somewhat restricted sample size, the outcomes of this study serve as a valuable complement to prior research investigating the correlation between word family, lexical coverage, and reading. These findings carry significant implications for enhancing Business English learners' comprehension of reading materials and can contribute to the informed development of business English textbooks. Furthermore, the insights gleaned from this study can potentially inform future updates to vocabulary corpora, ensuring their relevance and applicability in evolving language learning contexts.

References

Biemiller, A. (2003). Vocabulary needed if more children are to read well, *Reading Psychology*, *24*, 323– 335.

Coxhead, A. (2000). A New Academic Word List. *TESOL Quarterly*, *34*(2), 213–238.

Dang, T. N. Y., Coxhead, A., & Webb, S. (2007). The academic spoken word list. *Language Learning*, *4*, 959–997.

Gao, Y. (2015). *A study on the correlation between vocabulary and reading comprehension – An empirical study with self-taught English majors*. Zhejiang: Zhejiang University.

Gardner, D., & Davies, M. (2014). A new academic vocabulary list, *Applied Linguistics, 35*(3), 305-327.

Gui, M., Lu, N. N., & Shang, Y. J. (2020). The relationship between the proportion of college students' text literacy and English reading comprehension, *Contemporary Foreign Language Studies* (4), 73–81.

Han, G., & Geng, F. Z. (2007). Vocabulary threshold of English reading, *Journal of Qilu Normal University* (4), 102–104.

He, L. X., & Hao, J. (2012). How much vocabulary do English major learners need to read the New York Times, *Northern Literature* (5), 206.

Hirsh, D., & Nation, P. (1992). What vocabulary size is needed to read unsimplified texts for pleasure?. *Reading in a Foreign Language*, *8*, 689–696.

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

Hsu, W. (2013). Bridging the vocabulary gap for EFL medical undergraduates: The establishment of a medical word list, *Language Teaching Research, 17*(4), 454-484.

Hu, M., & Nation, P. (2000). Unknown vocabulary density and reading comprehension, *Reading in a Foreign Language*, *13*(1), 403–430.

Koda, K. (1989). The effects of transferred vocabulary knowledge on the development of L2 Reading proficiency, *Foreign Language Annals, 22*, 529–540.

Laufer, B. (1989). What percentage of text-lexis is essential for comprehension? Lauren, Ch., & Nordman, M. *Special Language: From Humans Thinking to thinking machines*. England: Multilingual Matters Press. 316–323

Laufer, B. (1991). The development of L2 lexis in the expression of advanced language learner, *Modern Language Journal*, *75*, 440–448.

Laufer, B. (1997). What' s in a word that makes it hard or easy: Some interlexical factors that affect the learning of words. In Schmitt, N., & McCarthy, M(Eds.), *Vocabulary: Description, Acquisition, and Pedagogy*, Cambridge: Cambridge University Press, 140–155.

Laufer, B., & Nation, P. (1993). Word families. *International Journal of Lexicography*, *6*, 253–279.

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.

Li, C. F., Yu, J., & Lin, J. (2015). Business English vocabulary: Description, identification and teaching – A Data Analysis Based on fortune's business English Word List, *Foreign Language World* (03), 83– 88, 96.

Li, Q. D. (2023). A study on the word class distribution of business English listening and speaking textbooks based on Range, *Journal of Heilongjiang Vocational Institute of Ecological Engineering* (02), 155–160.

Lin, J., Li, C. F. & Yu, J. (2018). Setting vocabulary thresholds for business English majors – Based on Range's analysis of Financial Times and Fortune business English news from 2006-2015. *Journal of Xiamen University (Arts & Social Sciences)* (02), 165–172.

Liu, D. L., & Lei, L. (2020). A review of research on academic word lists, *Foreign Language Education*
(2), 34–38, 50.

Ludewig, U., Hübner, N. & Schroeder, S. (2023). Vocabulary, text coverage, word frequency and the lexical threshold in elementary school reading comprehension, *Read Writ, 36*, 2409-2431. Doi: 10.1007/s11145-022-10385-0

Mao, J. (2016). The application of the economist in university business English teaching, *Education Modernisation* (14), 123–124.

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

Nation, P., & Waring, R. (2002). Vocabulary size, text coverage and word lists. In Schmitt, N. & McCarthy, M. *Vocabulary: Description, Acquisition and Pedagogy*. Shanghai: Shanghai Foreign Language Education Press. 6-19.

Qin, X. F. & Yang, D. Y. (2009). How much vocabulary is needed to read Chinese English newspapers, *Journal of Hebei University of Technology*, (4), 132–138.

Quinn, J. M., Wagner, R. K., Petscher, Y. & Lopez, D. (2015). Developmental relations between vocabulary knowledge and reading comprehension: A Latent change score modeling study, *Child Dev, 86*, 159-175. Doi: 10.1111/cdev.12292

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

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

Wang, F., & Lian, T. X. (2013). A corpus-based comparative study of vocabulary in business English and general English. *Journal of Dalian University of Technology (Social Science Edition)* (03), 130–133.

Wang, Y. L. (2017). *A corpus-based study of vocabulary coverage and gradient in New High School English.* Hohhot: Inner Mongolia Normal University.

Ward, J. (1999). How large a vocabulary do EAP engineering students need, *Reading in a Foreign Language*, *12*, 309–323.

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

Wilkins, D. A. (1972). *Linguistics in language teaching*, Cambridge: MIT Press.

Zhang, N. N., & Liu, Y. (2019). An empirical study on the relationship between English learners' vocabulary knowledge breadth and reading comprehension, *Journal of Tianjin Vocational University* (1), 76–82, 91.

Zhao, Z. G. (2013). *Readability research and RALCS system development based on vocabulary coverage*, Xinxiang: Henan Normal University.