LEARN Journal: Language Education and Acquisition Research Network
ISSN: 2630-0672 (Print) | ISSN: 2672-9431 (Online)
Volume: 17, No: 2, July – December 2024
Language Institute, Thammasat University
https://so04.tci-thaijo.org/index.php/LEARN/index

A Corpus-Based Vocabulary Analysis of First-Year Undergraduate Economics Textbooks in an International Program

Pongsathon Wasikarat[a,*], Kittitouch Soontornwipast[b]

[a] pongsathon.was@dome.tu.ac.th, Language Institute, Thammasat University, Thailand
[b] kittitouch.s@litu.tu.ac.th, Language Institute, Thammasat University, Thailand
[*] Corresponding author, pongsathon.was@dome.tu.ac.th

APA Citation:
Wasikarat, P. & Soontornwipast, K. (2024). A corpus-based vocabulary analysis of first-year undergraduate economics textbooks in an international program. *LEARN Journal: Language Education and Acquisition Research Network, 17*(2), 470-486.

Received 29/12/2023

Received in revised form 15/05/2024

Accepted 26/05/2024

ABSTRACT

The purpose of this current study was to (1) investigate the text coverage that the BNC/COCA Word Family Lists (Nation, 2017) and the Academic Word List (Coxhead, 2000) provided in the first-year undergraduate economics textbooks, and (2) estimate the vocabulary size required to read the textbooks. A corpus of 1,343,493 words from the economics textbooks was compiled into the ECON corpus, and it was then analyzed using the AntWordProfiler software program (Anthony, 2020). The study suggests that for 95% coverage, a vocabulary size of approximately 3,500-4,400 word families is necessary for a reasonable reading comprehension level, while for 98% coverage of an optimal level, around 8,500-9,900 word families are required to read economics textbooks in an international program. However, it does not imply that readers have reached a certain word level; it will automatically result in good reading comprehension (Laufer & Ravenhorst-Kalovski, 2010).

Keywords: corpus-based vocabulary analysis, text coverage, vocabulary size, economics textbooks, word frequency

Introduction

English is a widely spoken language, becoming a lingua franca for international communication in business, diplomacy, and education. Although the Ministry of Education of Thailand promotes English language teaching through government policy, implementation has been unsatisfactory (Kaur & Young, 2016). The country's CEFR-based proficiency levels are Basic User (A1 and A2), and the country was ranked in the low proficiency band in 2017, and the rank was lower in 2018 (Mala, 2018; Waluyo, 2019). Thai university students require a minimum level of English proficiency for admission and graduation. Learning through English encourages more English-medium instruction (EMI) practices and influences national education policies, contributing to a surge in interest in English programs and international programs across more diverse, specialized disciplines for 21st century education. EMI is a growing trend in teaching academic subjects, particularly in non-English-speaking countries, through English. It plays a critical role in achieving a country's globalization and internationalization policies. EMI courses in Thailand enhance students' English academic skills, resulting in high proficiency in reading, writing, and interactional English (Tang, 2021). However, EMI implementation in non-English-speaking peripheral countries faces challenges with limited English vocabulary among students. Students with less vocabulary face more difficulty (Aizawa & Rose, 2020). Nation and Waring (2002) argue that English has a rich vocabulary that makes it impossible to learn all words in English. Vocabulary size, or the breadth of vocabulary knowledge, is crucial for second language learners.

First-year undergraduate students face challenges in specialized courses and reading English-medium textbooks, making it crucial for them. For materials developers, understanding text coverage is essential for English proficiency development and vocabulary development and can help increase text comprehension and contribute to reading comprehension. Although there are many studies on corpus-based vocabulary analysis, it has been found so far that there are a few studies conducted independently of an EMI context, e.g., textbooks of English for Engineering (Chanchanglek & Sriussadaporn, 2011), textbooks for English majors (Hajiyeva, 2015), and materials for Business English majors (Liangpanit, 2010). Corpus-based studies in a specific discipline can be essential in teaching and learning and for students' academic success. Teachers and course designers should

consider text coverage, vocabulary size, or headwords in the textbooks as alternative criteria to improve students' English proficiency.

Based on the objectives of the study, the following research questions are formulated to guide the investigation:

(1) What text coverage do the BNC/COCA word family lists and the Academic Word List provide in the first-year undergraduate economics textbooks of the BEIP?

(2) What vocabulary size is needed to read the first-year undergraduate economics textbooks of the BEIP?

(3) What are the most frequently occurring AWL academic words in the first-year undergraduate economics textbooks of the BEIP?

A review of the literature then shows that there is a need to carry out a study that analyzes first-year undergraduate economics textbooks in an international program.

Literature Review

A corpus is a principled collection of authentic, machine-readable texts suitable for qualitative and quantitative research, transforming the way researchers think about vocabulary research studies (Biber et al., 1998; McEnery et al., 2006). Corpora in language instruction offers realistic examples and frequency data, focusing on the most used words in a text type for students.

According to Nation (2017), vocabulary is able to be divided into 25,000 frequency-based word families. The BNC/COCA word family lists consist of 29 lists, including 25 frequency and range data-based lists, and 4 additional lists. These lists represent a learner's vocabulary at higher frequency levels (Nation, 2006), selected from 100 million words based on word frequency. The 25-level theoretical research focuses on learners mastering vocabulary in order of frequency.

Vocabulary is crucial for developing English language skills and effective communication (Nation, 1993). English has a countless number of words, making it impossible for learners to master them like native speakers. Poor or limited vocabulary knowledge negatively impacts academic performance (Altahab, 2019; Pookcharoen, 2007). Vocabulary size is crucial for second language (L2) acquisition, as L2 learners have limited time and opportunities to learn words like native speakers (Laufer, 2014). Vocabulary size and text coverage are reciprocally related (Fan, 2013). Text coverage refers to the percentage of running words or tokens in the text a reader knows (Nation, 2006). It is important as a means of estimating how comprehensible

reading materials are, and it can assist researchers and teachers better tailor vocabulary learning objectives.

The BNC/COCA Word Family Lists

Not all words are equally useful for language instruction. Second language learners should focus on high-frequency words (Nation & Waring, 1997), which are the foundation of language use. Nation (2001) divided vocabulary into high, mid, and low-frequency categories, expanding vocabulary knowledge.

In the BNC/COCA word family lists, words are arranged according to how they are frequently perceived for listening and reading comprehension. Here are some examples: 1) high-frequency words, i.e., develop, communicate, incredible; 2) mid-frequency words, i.e., beautiful, important, government; and 3) low-frequency words, i.e., play, house, eat. Beginning learners are suggested to frequently concentrate on the BNC/COCA word family lists for key reasons: 1) these are frequency-based word lists and this implies that learners should spend more time on them and less time learning vocabulary they might rarely use; 2) the lists help learners understand how the base word and all the related forms are used; and 3) by focusing on word families, this can help learners infer the meaning of unknown words within a word family that they already partially know.

The General Service List (GSL)

Frequency-based word lists can assist students learn essential words that they need to know. Michael West (1953) compiled the General Service List (GSL), a well-known list of 2,000 headwords, comprising the first 1000 words and the second 1000 words, that was thought to be the most commonly used English words for learners of English as a Foreign Language (EFL). Here are some in the first 1,000 words: *account, allow, battle, believe, chance, choose, decide, divide, enter, explain, fail,* and *find*, and some in the second 1,000 words: *admire, apologize, borrow, butter, ceremony, civilize, dinner, drink, electricity, explore, fault,* and *frequent.* The GSL includes the most common words in English or the vocabulary needed to understand simple texts and conversations. Students who are familiar with these words will find it easier to acquire more difficult words.

The Academic Word List (AWL)

Academic vocabulary, known as sub-technical or semi-technical words, can be defined as words that are regularly used in academic texts but are uncommon in other genres (Coxhead & Nation, 2001). General (core) academic vocabulary accounts for a higher proportion of running words than general English words in academic texts. Learning academic vocabulary requires a combination of knowledge related to the discipline and the academic language, with high-frequency words essential for common speech and running words in all writings. Academic word lists, such as the Academic Word List (AWL) developed by Coxhead (2000), are essential for teaching and learning academic vocabulary. The AWL is a developed comprehensive word list of 570 academic words, not specific to any particular subject, making it beneficial for growing vocabulary for academic purposes. Besides learning technical vocabulary, learning core academic vocabulary in the AWL can assist students in enhancing comprehension in academic texts. It covers 10% of words in academic texts and are crucial in teaching materials of English as a Second Language (ESL)/English as a Foreign Language (EFL) and for academic learners. Word families, created by combining base words with inflectional and derivational forms, help learners understand and deduce meanings of frequently inflected and derived forms.

Nation (2006) emphasized the significance of English vocabulary knowledge for typical language use, focusing on three aspects: the number of words in English, in the native speaker's knowledge, and in the learner's vocabulary needs. This helps investigate an appropriate size for a non-native learner's vocabulary. Schmitt (2008) recommends a large vocabulary size of 8,000-9,900 word families for reading and 5,000-7,000 for oral discourse. L2 learners must expand their vocabulary to read diverse materials without unfamiliar vocabulary. Laufer and Ravenhorst-Kalovski (2010) and Nation (2006) suggest that textbooks require a minimal vocabulary size of 4,000-5,000 word families for 95% text coverage and 7,000-8,000 word families for 98%.

Methodology

The quantitative research methodology was used to analyze text coverage, vocabulary size, and academic words in first-year undergraduate economics core courses at an autonomous university in Thailand. The methodology focused on corpus design, compilation, data collection, and data analysis.

Research Instruments

AntWordProfiler

The AntWordProfiler (Anthony, 2022) software program was selected for quantitative text analysis to investigate text coverage and word frequency in first-year undergraduate economics textbooks of the BEIP. It was used to compare vocabulary across multiple text files (4 subcorpora) and a single corpus (the ECON corpus in this current study), explore word coverage, create word lists based on frequency and distribution, and discover specialized vocabulary in a specific text. The program's functions include comparing vocabulary usage, analyzing word frequency, and discovering specialized vocabulary in a text. It helps create vocabulary learning sequences by word frequency, target specific needs, and modify learning burdens and opportunities. This program categorized texts by word frequency, using two word lists: the BNC/COCA 1st to 25th 1,000 word families and Coxhead's (2000) AWL containing 570 word families. Word list studies use counting units based on skill concentration. Lemmas are more practical for productive skills, while word families are suggested for measuring receptive skills in reading and listening (Brezina & Gablasova, 2015; Nation, 2016). Word families represent a collection of words with implied meanings when the learner is aware of the base form. This current study used word families to count the most frequent words in the ECON corpus, as Nation and Webb (2011) suggest using word families as the most sensible unit when for measuring receptive knowledge in reading.

The Selected Undergraduate Economics Textbooks

This study investigates the economics textbooks used by first-year undergraduates in the Bachelor of Economics International Program (BEIP) at an autonomous university in Thailand. A corpus of first-year undergraduate economics textbooks used in the Bachelor of Economics International Program (BEIP) was compiled using a corpus-based vocabulary analysis. The 1,343,493-word ECON corpus has been created, comprising four subcorpora that were collected from the four economics course textbooks. These textbooks are mandatory for students to read in core courses listed in the BEIP study plan. The 2018 BEIP curriculum requires students to complete 131 credits in economics core courses to graduate. The study found 12 accredited textbooks in the first year of study, including Principles of Microeconomics (EE211), Principles of Macroeconomics (EE212), Calculus

for Social Science 1 (MA216), and Statistics for Social Science 1 (ST216). The researcher requested faculty permission to have a list of compulsory textbooks in first-year undergraduate economics core courses before a text corpus compilation stage.

Data Collection

The text corpus compilation for the Bachelor of Economics International Program (BEIP) consists of first-year undergraduate economics textbooks listed in the curriculum. The researcher collected texts in electronically-readable formats, inverting hardcopy or printed versions using a scanner. The corpus aims to represent the economics textbooks required for the first year of study. The researchers compiled the corpus, considering various issues and ensuring all text files were ready for compiling into subcorpora and the ECON corpus.

Data Analysis

The researcher prepared texts which represent large, electronic data sets of corpus data of the first-year undergraduate textbooks of the BEIP. The corpus texts were used to compile a corpus of the economics core courses. However, the aforementioned corpus was compiled to represent the first-year undergraduate economics textbooks of the BEIP. As mentioned earlier, a text with Microsoft Word (.doc) format file or any other format is not able to be processed on the AntWordProfiler (Anthony, 2022) software program. The texts were processed to electronically-readable data for the analysis. They were skimmed, and scanned to remove irrelevant items, including all graphics, tables, anchored items, numbers, formulas, symbols, abbreviations, the table of contents, indices, appendices, and any glossaries. To ensure validity, a spelling-checking feature and the Find and Replace feature in a word processor software program were used to recheck the corpus data. In addition, an electronic version of the first-year undergraduate textbooks of the BEIP was converted from PDF (.pdf) format to Microsoft Word (.doc or .docx) format using AntFileConverter (Anthony, 2022), a software program employed in this study to convert either PDF format and Microsoft Word (.doc or .docx) format into plain text for being able to be used or ready. After the texts were completely been stripped of irrelevant items, the researcher finally converted texts into plain text (.txt) files for the next stage of data analysis.

The AntWordProfiler software program (Anthony, 2022) was used to analyze the vocabulary of the BNC/COCA word family lists (Nation, 2017) and the Academic Word List (Coxhead, 2000) occurring in the first-year undergraduate economics textbooks of the BEIP. In other words, it informs how much and what vocabulary occurs in a particular text or a group of texts. The textbooks were compiled into 4 subcorpora and the ECON corpus and were analyzed using the program. After the program analysis, it would illustrate (1) tokens, also known as running words that appear in the whole corpus and its subcorpora of first-year undergraduate economics textbooks in this current study; (2) word families which refer to groups of words that share a common root word with different derivations (e.g. prefixes and suffixes) and inflections (different forms); (3) text coverage or percentage of known words in the textbooks which may indicate text comprehension, leading to reading comprehension.

Results

The Text Coverage of the BNC/COCA Frequency-Based Word Family Lists Provided and the Vocabulary Size Needed to Read the Textbooks

The findings show the text coverage of the BNC/COCA word family lists and the estimated vocabulary size in the textbooks (subcorpora) of the four different economics core courses as follows: principles of microeconomics (MIC), principles of macroeconomics (MAC), calculus for social science 1 (CAL), statistics for social science 1 (STA), and the whole corpus.

Table 1

Word Levels, Tokens, No. of Word Families, Text Coverage, and Cumulative Text Coverage of the BNC/COCA Word Family Lists in Each Subcorpus and the ECON Corpus

BNC/ COCA Word List	MIC Subcorpus	MAC Subcorpus	CAL Subcorpus	STA Subcorpus	ECON Corpus
	Text Coverage/ Cumulative Text Coverage	Text Coverage/ Cumulative Text Coverage	Text Coverage/ Cumulative Text Coverage	Text Coverage/ Cumulative Text Coverage	Text Coverage/ Cumulative Text Coverage
Proper Nouns	1.20%	5.06%	26.77%	2.35%	8.85%

BNC/ COCA Word List	MIC Subcorpus	MAC Subcorpus	CAL Subcorpus	STA Subcorpus	ECON Corpus
	1.20%	5.06%	26.77%	2.35%	8.85%
1st 1,000	72.32%	68.36%	52.43%	65.98%	64.77%
	73.51%	73.42%	79.20%	68.33%	73.62%
2nd 1,000	13.53%	13.62%	6.90%	13.47%	11.88%
	87.05%	87.04%	86.09%	81.80%	85.50%
3rd 1,000	8.62%	8.35%	5.89%	12.25%	8.78%
	***95.66%**	***95.39%**	91.99%	94.04%	94.27%
4th 1,000	2.02%	1.27%	2.13%	1.89%	1.83%
	97.69%	96.66%	94.12%	***95.93%**	***96.10%**
5th 1,000	1.00%	1.15%	1.06%	0.95%	1.04%
	****98.69%**	97.81%	***95.18%**	96.88%	97.14%
6th 1,000	0.30%	0.23%	0.52%	0.47%	0.38%
	98.98%	****98.04%**	95.70%	97.35%	97.52%
7th 1,000	0.18%	0.19%	0.17%	0.25%	0.20%
	99.17%	98.23%	95.87%	97.61%	97.72%
8th 1,000	0.16%	0.25%	0.37%	0.18%	0.24%
	99.33%	98.48%	96.23%	97.79%	97.96%
9th 1,000	0.03%	0.05%	0.39%	0.14%	0.15%
	99.36%	98.53%	96.62%	97.92%	****98.11%**
10th 1,000	0.05%	0.05%	0.09%	0.06%	0.06%
	99.40%	98.58%	96.71%	97.98%	98.17%
11th 1,000	0.04%	0.11%	0.10%	0.07%	0.08%
	99.44%	98.69%	96.82%	****98.05%**	98.25%
12th 1,000	0.02%	0.04%	0.11%	0.05%	0.06%
	99.46%	98.73%	96.93%	98.10%	98.31%
13th 1,000	0.03%	0.04%	0.13%	0.14%	0.09%
	99.49%	98.77%	97.06%	98.24%	98.39%
14th 1,000	0.05%	0.07%	0.04%	0.10%	0.07%
	99.54%	98.84%	97.10%	98.33%	98.45%
15th 1,000	0.02%	0.01%	0.03%	0.02%	0.02%
	99.56%	98.85%	97.12%	98.35%	98.47%
16th 1,000	0.00%	0.02%	0.06%	0.01%	0.02%
	99.56%	98.87%	97.19%	98.37%	98.50%
17th 1,000	0.01%	0.02%	0.01%	0.03%	0.02%
	99.57%	98.89%	97.20%	98.40%	98.52%
18th 1,000	0.01%	0.01%	0.03%	0.00%	0.01%
	99.58%	98.90%	97.23%	98.40%	98.53%
19th 1,000	0.00%	0.01%	0.01%	0.00%	0.01%
	99.58%	98.91%	97.24%	98.40%	98.54%
20th 1,000	0.01%	0.01%	0.03%	0.00%	0.01%
	99.59%	98.92%	97.27%	98.40%	98.55%
21st 1,000	0.00%	0.00%	0.01%	0.00%	0.00%
	99.59%	98.92%	97.28%	98.41%	98.55%
22nd 1,000	0.00%	0.01%	0.03%	0.00%	0.01%
	99.60%	98.93%	97.31%	98.41%	98.56%

BNC/ COCA Word List	MIC Subcorpus	MAC Subcorpus	CAL Subcorpus	STA Subcorpus	ECON Corpus
23rd 1,000	0.00% 99.60%	0.02% 98.95%	0.00% 97.31%	0.01% 98.42%	0.01% 98.57%
24th 1,000	0.00% 99.60%	0.00% 98.95%	0.00% 97.31%	0.00% 98.42%	0.00% 98.57%
25th 1,000	0.00% 99.60%	0.00% 98.95%	0.05% 97.36%	0.00% 98.42%	0.01% 98.58%
Off-list	0.40% 100.00%	1.05% 100.00%	2.64% 100.00%	1.58% 100.00%	1.42% 100.00%
Total	**100.00%**	**100.00%**	**100.00%**	**100.00%**	**100.00%**

Table 1 shows text coverage of BNC/COCA word family lists in BEIP's first-year economics core courses and cumulative coverage of these lists in the textbooks (remarks: *meaning that the 95% of text coverage was reached, **meaning that the 98% of text coverage was reached).

In the MIC subcorpus, the results have shown that the 3rd 1,000 BNC/COCA word family list reached 95.66% of text coverage, while the 5th 1,000 list reached 98.69% of text coverage. In the MAC subcorpus, the 3rd 1,000 BNC/COCA word family list reached 95.39%, and the 6th 1,000 list reached 98.04%. In the CAL subcorpus, the 5th 1,000 BNC/COCA word family list reached 95.18%; however, the 98% of text coverage was not reached. In the STA subcorpus, the 4th 1,000 BNC/COCA word family list reached 96.10%, and the 11th 1,000 list reached 98.05% of text coverage. As a result, in the ECON corpus, the 4th 1,000 BNC/COCA word family list reached 96.10%, resulting in having a vocabulary size of around 3,500–4,000 word families, and the 9th 1,000 BNC/COCA word family list reached 98.11%, resulting in having a vocabulary size of around 8,500–9,000 word families.

Text Coverage of the Academic Word List

This part reports the text coverage of the Academic Word List appearing in the aforementioned textbooks, and the corpus was analyzed to report: (1) text coverage of the General Service List (GSL) and the Academic Word List (AWL) in each subcorpus and the ECON corpus; and (2) cumulative text coverage of the GSL and AWL in each subcorpus and the ECON corpus.

Table 2

Word Levels, Tokens, No. of Word Families, Text Coverage, and Cumulative Text Coverage of the GSL and AWL in Each Subcorpus and the ECON Corpus

Word List	MIC Subcorpus	MAC Subcorpus	CAL Subcorpus	STA Subcorpus	ECON Corpus
	Text Coverage/ Cumulative Text Coverage	Text Coverage/ Cumulative Text Coverage	Text Coverage/ Cumulative Text Coverage	Text Coverage/ Cumulative Text Coverage	Text Coverage/ Cumulative Text Coverage
General Service List (GSL)					
1st 1,000 GSL	80.62%	77.78%	66.81%	74.20%	74.85%
	80.62%	77.78%	66.81%	74.20%	74.85%
2nd 1,000 GSL	5.79%	4.62%	3.10%	5.93%	4.86%
	86.41%	82.40%	69.91%	80.13%	79.71%
Academic Word List	**8.73%**	**9.53%**	**6.44%**	**13.80%**	**9.62%**
	95.13%	91.93%	76.35%	93.93%	89.33%
Off-List	7.42%	8.07%	23.65%	6.07%	10.67%
	100.00%	100.00%	100.00%	100.00%	100.00%
Total	**100.00%**	**100.00%**	**100.00%**	**100.00%**	**100.00%**

Table 2 displays the text coverage and the cumulative text coverage of the General Service List (GSL), Academic Word List (AWL), and ECON corpus in the first-year undergraduate economics textbooks of the Bachelor of Economics International Program (BEIP). The study used the General Service List (GSL) as the base list for general English words and the Academic Word List (AWL) as the base list for academic English words.

According to the results of GSL's text coverage, the subcorpora in the study were as follows: (1) the subcorpus of the course principles of microeconomics (hereafter referred to as the MIC subcorpus) comprising 80.62% and 5.79% of the 1st 1,000 GSL and 2nd 1,000 GSL of text coverage and resulting in the cumulative text coverage of the GSL at 86.41%, (2) the subcorpus of the course principles of macroeconomics (hereafter referred to as the MAC subcorpus) comprising 77.78% and 4.62% of the 1st 1,000 GSL and 2nd 1,000 GSL of text coverage with 82.40% of the cumulative text coverage, (3) the subcorpus of the course calculus for social science 1 (hereafter referred to as the CAL subcorpus) comprising 66.81% and 3.100% of the 1st 1,000 GSL and 2nd 1,000 GSL of text coverage and resulting in the cumulative text coverage of the GSL at 69.91%, and (4) the subcorpus of the

course statistics for social science 1 (hereafter referred to as the STA subcorpus) comprising 74.20% and 5.93% of the 1st 1,000 GSL and 2nd 1,000 GSL of text coverage and the cumulative text coverage at 80.13%. According to the text coverage of GSL in the whole corpus (ECON corpus) in this study, the 1st 1,000 GSL and the 2nd 1,000 GSL constituted 74.85% and 4.86%, respectively, of text coverage. Consequently, the GSL constituted 79.71% of text coverage in the ECON corpus. The text coverage analysis of the AWL in various text subcorpora revealed that the AWL provided 8.73% coverage in the MIC subcorpus, 9.53% in the MAC subcorpus, 6.44% in the CAL subcorpus, and 13.80% in the STA subcorpus. In the ECON corpus, the AWL provided 9.62% of text coverage.

The Most Frequently Occurring AWL Academic Words

The third research question in this study will identify the AWL academic words used most frequently in BEIP's first-year undergraduate textbooks. To begin addressing this question, a list of AWL academic words that meet the standardized criteria for frequency uniformity according to other studies, which include word frequency, needs to be compiled. Based on Coxhead's (2000) criteria for word selection of AWL words, the occurrences were then determined. After the mathematical calculation, the frequency of occurrence of AWL academic words should be at least 38 times in the ECON corpus and at least 4 times in each subcorpus. It was discovered that there were 187 AWL academic words, ranked in frequency order, and they appeared at least 38 times overall in the whole corpus and at least 4 times across the subcorpora. The most frequent AWL academic words were data (2,978 times), function (2,895 times), equation (2,808 times), chapter (2,398 times), income (2,053 times), series (1,575 times), distribution (1,549 times), area (1,306 times), economy (1,232 times), estimate (1,222 times), labor (1,187 times), section (1,063 times), percent (981 times), compute (965 times), functions (905 times), and so on.

Discussion

Vocabulary is crucial for language abilities, as it enhances basic skills like reading, writing, listening, and speaking. According to the previous studies done by Laufer and Ravenhorst-Kalovski (2010) and Nation (2006), the minimal vocabulary necessary to read and comprehend the textbooks is that having a vocabulary size of 4,000–5,000 word families is for 95% text coverage and that of 7,000–8,000 word families is for 98% coverage. The

results of this current study suggest that a vocabulary size of 3,500–4,000 word families in an economics textbook is needed for reasonable reading comprehension as the BNC/COCA word family lists account for 95% of the running words in the ECON corpus, and a vocabulary size of approximately 9,000 word families is needed for optimal reading comprehension as the lists account for 98% of the running words in the ECON corpus. This concurs with Schmitt's (2008) suggestion that it requires a large vocabulary of 8,000–9,000 word families for reading. Nevertheless, it does not claim that reasonable comprehension can occur even if readers have reached a particular word band, which will automatically lead to good reading comprehension (Laufer & Ravenhorst-Kalovski, 2010).

The current study found that text coverage increased after reaching 95%, showing small increases in cumulative text coverage. Even a small increase in text coverage leads to the highest improvement in reading scores (Laufer & Ravenhorst-Kalovski, 2010). Low-frequency words were essential for comprehension and decoding, whereas high-frequency words expanded vocabulary coverage for beginning-level students (Nation, 2001). Word selection based on frequency should focus on balanced and effective teaching and learning.

The AWL accounts for 10% across academic texts, while this current study gave a 9.62% text coverage of the AWL, indicating lower text coverage than Coxhead's (2000) AWL. However, the vocabulary in economics core courses like Principles of Macroeconomics, Principles of Microeconomics, and Statistics for Social Science 1 was relatively relevant to Coxhead's (2000) AWL. Calculus for Social Science 1 had lower AWL coverage, possibly due to the assumption of some deviations from academic texts.

Conclusions and Implications

Corpora are essential for authentic English language usage and quality course materials development. As vocabulary mastery is crucial for ESL/EFL students to develop English language skills, non-native speakers have a smaller vocabulary than native speakers and often lack the necessary vocabulary knowledge, making it difficult for them to learn new words and use them purposefully. This current study analyzed a 1,343,493-word corpus of the first-year undergraduate economics textbooks of an international program using the AntWordProfiler (Anthony, 2022) software program. The results showed that AWL accounted for 9.62% of text coverage, while the first and second 1000-word bands of the GSL provided 79.71%. These words can serve as an alternative guide for future implications when developing curricula, courses, and instructional materials.

Laufer and Nation (2012) indicated that curriculum design often focused on grammar, topics, or tasks and overlooked vocabulary instruction. The results of the current study can be useful for curriculum developers, materials designers, subject content lecturers, and English language lecturers. They can use the results to redesign or modify instructional materials so that vocabulary will be presented in a meaningful way. Careful word selection is essential for effective learning in instructional materials, considering pedagogical purpose, context, and students' needs. Curriculum developers and materials designers need to support individuals with different needs. For example, introducing words with their definitions both in a general context and in an economics context, using vocabulary learning strategies with guidance in learning new vocabulary in an economics textbook, and providing multiple opportunities for vocabulary instruction to enhance comprehension are all pinpoints that first-year economics students with limited English proficiency or those who struggle with learning may need extra attention and support from economics textbooks.

Limitations and Recommendations for Further Research

This current study analyzed first-year economics undergraduate textbooks for core courses in an international program in Thailand. The 1,343,493-word corpus would allow researchers, teachers, and materials developers to improve vocabulary input in class or prepare their first-year students. The study focused on word families as a counting unit, but did not analyze multi-word units. Understanding word family patterns is crucial for comprehending the complexity of English. Further research could explore multi-word units or other expressions found in textbooks, such as lexical bundles and idiomatic expressions. Further studies might explore economics textbooks from other years of study, regardless of the first year. Researchers could also explore textbooks in other economics-related disciplines or in other different pedagogical and non-pedagogical genres, as follows: journal articles, research abstracts, research proposals, etc.

About the Authors

Pongsathon Wasikarat: A Ph.D. candidate in the English Language Teaching International Program, Language Institute, Thammasat University. His research interests include English Language Teaching, Corpus Linguistics, Language Assessment, Self-Directed Learning, and Educational Measurement and Evaluation.

Kittitouch Soontornwipast: An assistant professor at the Language Institute, Thammasat University. His research interests include Teacher Training, Instructional Media, Action Research for English Teachers, Teaching Methodology, Qualitative Research, Mixed-Methods Research, and Materials Development.

References

Aizawa, I., & Rose, H. (2020). High school to university transitional challenges in English-medium instruction in Japan. *System, 95,* 1-11.

Anthony, L. (2022). *AntFileConverter (Version 2.0.2)* [Computer Software]. Waseda University. https://www.laurenceanthony.net/software

Anthony, L. (2022). *AntWordProfiler (Version 2.0.1)* [Computer Software]. Waseda University. https://www.laurenceanthony.net/software

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

Brezina, V., & Gablasova, D. (2015). Is there a core general vocabulary? Introducing the new general service list. *Applied Linguistics, 36*(1), 1–22.

Chanchanglek, S., & Sriussadaporn, N. (2011). A corpus analysis of English vocabulary input in course materials used for Engineering students. *Humanities Journal, 18*(1), 141–149.

Cobb, T. (2010). Learning about language and learners from computer programs. *Reading in a Foreign Language, 22*(1), 181–200.

Cobb, T. (2020). *Familizer + Lemmatizer v.2.2* [computer program]. https://lextutor.ca/familizer.

Coxhead, A. (2000). A new academic word list. *TESOL Quarterly, 34*(2), 213–238.

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

Fan, F. (2013). Text length, vocabulary size and text coverage constancy. *Journal of Quantitative Linguistics, 20*(4), 288–300.

Hajiyeva, K. (2015). A corpus-based lexical analysis of subject-specific university textbooks for English majors. *Ampersand, 2*, 136–144.

Kaur, A., & Young, D. (2016). English education policy in Thailand: Why the poor results? In R. Kirkpatrick (Ed.), *English language education policy in Asia* (pp. 345–361). Springer International Publishing.

Laufer, B. (2014). Vocabulary in a second language: Selection, acquisition, and testing: A commentary on four studies for JALT vocabulary SIG. *Vocabulary learning and instruction, 3*(2), 38–46.

Laufer, B. & Nation, I. S. P. (2012). Vocabulary. In S. M. Gass & A. Mackey (Eds.), *The Routledge handbook of second language acquisition* (pp. 163–176). Routledge.

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.

Liangpanit, C. (2010). *The development of a corpus-based vocabulary package for Business English majors* [Doctoral dissertation, Suranaree University of Technology, Nakhon Ratchasima, Thailand]. http://sutir.sut.ac.th:8080/jspui/bitstream/123456789/3469/1/Full text.pdf

Mala, D. (2018, November 5). *Thai English proficiency drops*. Bangkok Post. https://www.bangkokpost.com/thailand/general/1570042/thai-english-proficiency-drops

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

Nation, I. S. P. (1993). Vocabulary size, growth, and use. *The bilingual lexicon, 6*, 115–134.

Nation, I. S. P. (2001). *Learning vocabulary in another language.* 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. (2016). *Making and using word lists for language learning and testing*. John Benjamins.

Nation, I. S. P. (2017). The BNC/COCA Level 6 word family lists (Version 1.0.0) [Data file]. http://www.victoria.ac.nz/lals/staff/paul-nation.aspx

Nation, I. S. P., & Webb, S. A. (2011). *Researching and analyzing vocabulary*. Boston, MA: Heinle, Cengage Learning.

Nation, I. S. P., & Waring, R. (1997). Vocabulary size, text coverage and word lists. *Vocabulary: Description, acquisition and pedagogy*, *14*(1), 6–19.

Pookcharoen, S. (2007). Exploring how teaching morphemic analysis foster Thai EFL students' vocabulary knowledge. *Journal of English Studies, 3*, 85–105.

Schmitt, N. (2008). Review article: Instructed second language vocabulary learning. *Language Teaching Research, 12*(3), 329–363.

Schmitt, N., & Schmitt, D. (2014). A reassessment of frequency and vocabulary size in L2 vocabulary teaching. *Language Teaching, 47*(4), 484–503.

Tang, K. N. (2021). Implementing English-medium instruction (EMI) in Thailand: University students' perspectives. *KKU Research Journal (Graduate Studies) Humanities and Social Sciences, 9*(2), 95–108.

Waluyo, B. (2019). Thai first-year university students' English proficiency on CEFR levels: A case study of Walailak University, Thailand. *The New English Teacher, 13*(2), 51–71.

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