KOREAN JOURNAL OF ENGLISH LANGUAGE AND LINGUISTICS

The Vocabulary List for ELT Textbooks in Language Acquisition and Teaching Methodology*

Hyeon-Okh Kim · Hye-Kyung Lee (Ajou University)

Hyeon-Okh Kim (1st author)
Professor, Dept. of English
Ajou University.
Email: hokim67@ajou.ac.kr

Hye-Kyung Lee (corresponding author)
Professor, Dept. of English
Ajou University.
Email: hklee@ajou.ac.kr

* A preliminary version of this paper was presented at the 2022 KATE SIG Conference, Seoul.

ABSTRACT

Kim, Hyeon Okh and Hye-Kyung Lee. 2023. The vocabulary list for ELT textbooks in language acquisition and teaching methodology. *Korean Journal of English Language and Linguistics* 23, 873-894.

The study aims to investigate vocabulary demands in academic textbooks for ELT majors, identifying the most frequent word list for preservice teachers and variations in lexical needs across sub-areas. For the purposes, the study compiled a corpus comprising approximately 1.6 million tokens from twelve university textbooks in the integral dimensions of ELT: language acquisition and teaching methodology. By analyzing the lexical coverage of the ELT textbook corpus against the twenty-five 1,000 word-family list from the British National Corpus and Corpus of Contemporary American English (BNC/COCA), the study assessed the lexical load of ELT textbooks and developed an essential vocabulary list. Findings revealed that achieving a 95% lexical coverage in ELT textbooks necessitates mastery of the top 4,000 word families, including proper nouns, interjections, transparent compounds, abbreviations, and glossary terms. To attain 98% coverage, however, ELT students require an 11,000-word family vocabulary. Further analyses show that textbooks in language acquisition demand a higher lexical requirement compared to those in teaching methodology. By applying a set of criteria for widespread use and pedagogical relevance, the study identified 513 word families beyond the initial 2,000 levels on the BNC/COCA, constituting 9.36% of the ELT textbooks. The study suggests practical pedagogical implications.

KEYWORDS

ELT vocabulary, ELT word list, technical vocabulary, discipline-specific word list, vocabulary load, ELT textbooks, corpus analysis

1. Introduction

Vocabulary is an essential part of language learning. As vocabulary knowledge constitutes the ability to communicate as the basic building blocks of language system, a robust vocabulary is a stepping stone to effective communication, better literacy skills, higher-level fluency, and greater academic performance (Laufter and Nation, 1995, Nation 2001). Previous research attests to the importance of vocabulary size as a strong predictor in language proficiency across the board: The richer vocabulary, the better fluency development in all aspects of language (Nation 1990). While the vocabulary size is a critical indicator of reading performance, the predictive power of academic vocabulary is much stronger than that of the vocabulary size itself (Moon 2017). Indeed, academic vocabulary emerges as an even more influential factor in determining academic performance, surpassing the significance of sheer vocabulary size.

Since academic vocabulary is characterized by "distinctive lexical, morphological, syntactic, and stylistic features," learning and using academic language could be a task of the greatest challenges (Scott, Nagy and Flinspach 2008). As the meaning of technical academic vocabulary is part of the subject knowledge (Armbruster 1992), L2 readers should have in-depth vocabulary knowledge specifically processed in their field of study. L2 students of ELT majors, in particular, are expected to acquire a higher degree of vocabulary knowledge not only because they are prospective English teachers, but also because their curriculum, materials, and teaching practices require them to master highly specialized academic English. ELT students need to develop an extensive repertoire of technical vocabulary to meet their academic challenges and requirements in their teacher-training.

Previous research has indeed produced word lists relevant to English education. However, the lists were often compiled from relatively small-sized corpora from textbooks used within specific curriculum courses (Lee and Kim 2013), which necessitates further investigation that employs larger corpora from a wider range of textbooks commonly used in practice. Other studies (Chung 2014, Ha 2018) sourced their corpora from research articles that might not be readily available to students majoring in ELT, making the findings less directly applicable to the broader population of preservice teachers. Moreover, no previous research has undertaken the task of evaluating the lexical threshold required for effective comprehension of ELT textbooks. Additionally, given the interdisciplinary nature of ELT, there could be differing lexical requirements depending upon specific sub-areas. Recognizing the essential nature of language acquisition and teaching methodologies as mandatory core dimensions in teacher professional development within ELT, this study seeks to assess the specific lexical demands and challenges inherent to ELT academic materials related to these sub-areas. Thorough lexical analyses of academic textbooks within these two sub-areas may reveal distinct lexical burdens in ELT academic materials for preservice teachers. Ultimately, the research aims to assess the lexical load of ELT textbooks and create a valuable inventory of technical vocabulary that caters to the precise needs of preservice ELT teachers, enhancing the quality of their training and professional competence. The study is guided by the following questions:

1) What size of vocabulary is required for preservice teachers to be able to adequately read their ELT textbooks in core courses?
2) What kinds of words constitute a discipline-specific academic word list that suffices to read ELT textbooks with 95% lexical coverage?
3) What differences may exist in the lexical threshold and the frequency of domain-specific academic words between ELT sub-corpora: language acquisition as opposed to teaching methodology?

2. Literature Review

2.1 Academic Vocabulary

Academic vocabulary refers to a range of words that are commonly used in educational contexts. As such, academic language employs "a different constellation of linguistic resources from what is typical or expected in everyday conversation" (Schleppegrell 2004). As a particular register used in school settings, academic vocabulary is classified into two categories: general core vocabulary and discipline-specific academic vocabulary (Beaumann and Graves 2010). The former encompasses commonly used, high-frequency versatile words that appear across various fields of study, while the latter includes domain-specific words specific to a particular discipline which may involve technical jargon.

This line of research has generated a significant body of research concerning the acquisition of fundamental vocabulary that is applicable across various academic domains, alongside vocabulary that is specific to a particular discipline, and how its familiarity relates to text comprehension within those specific fields. The former type of research involved inquiries into a compilation of commonly used words that are prevalent across diverse subjects as in West's (1953) General Service List (GSL, henceforth), whereas the latter focused on identifying subject-specific vocabulary lists, taking into account the specific registers whose meanings may vary across fields (Nation 2001). Though there is no clear-cut boundary between the general-service and specialist vocabulary, highly specialized academic vocabulary is rarely overlapped with general-service words (Xue and Nation 1984). What is more, words used in everyday language may have different meanings and uses in specialized academic texts (Hyland and Tse 2007), thus distinctively classified as technical or sub-technical vocabulary (Chung and Nation 2003). Since learning specialist or technical words requires students to know the content associated with the word, learning technical vocabulary is part of content learning (Armbruster 1992).

While the size of academic vocabulary needed for successful reading may vary, previous research introduced two lexical benchmarks based on the portion of running words the reader knows in the text: 95% lexical coverage for a minimum "acceptable" comprehension of texts and 98% for an "optimal" level of comprehension (Laufer and Ravenhorst-Kalovski 2010). For the minimally acceptable vocabulary threshold for achieving "adequate" comprehension of authentic texts, Laufer (1989) claimed, around 95% lexical coverage is necessary, assuming readers can infer the meaning of unknown words from contexts. For "successful" comprehension of a text, according to Nation (2006), readers should attain about 98% lexical coverage. Following Nation's (2006) estimation of vocabulary load based on word frequency within every 1000 word-families, achieving 95% vocabulary coverage amounts to the knowledge around 4,000 word-families in addition to proper nouns and exclamations, whereas reaching 98% vocabulary coverage of authentic texts implies a reader is acquainted with 8,000 to 9,000 word-families, thereby requiring a broader vocabulary knowledge. Given the substantial vocabulary size for effective comprehension of academic texts, the widely recognized the 2000 words in the GSL, accounting for about 85% of written academic English (Nation and Waring 1997), deems insufficient in ensuring effective comprehension of academic textbooks, especially in higher educational contexts. Even when incorporating Coxhead's (2000) Academic World List (AWL, henceforth) – additional 570 word families beyond the GSL, the combined coverage still only spans up to 86.1% of tokens in the corpus (Coxhead 2000, Li and Qian 2010).

Moreover, the representativeness of the general-purpose vocabulary lists varies across diverse fields and genres: For instance, West's 2000 GSL exhibits lexical coverage ranging from 78% to 92% (Nation and Waring 1997). Similarly, Coxhead's 570 academic word families, initially estimated to cover 10% across academic texts, demonstrated notably lower coverage rates: 2.4% of English presentations (Hincks 2003) and 6.72% of anatomy

text (Cobb and Horst 2002). Chung and Nation (2003) highlighted the considerable disparity in the coverage of academic vocabulary within specialized texts. Through a comparison of the distribution of technical vocabulary between two areas, Chung and Nation effectively demonstrated significant variations in the proportion of specialized vocabulary between applied linguistics and anatomy texts. Within the text of applied linguistics, around 68.5% of the total 93,445 tokens were derived from West's GSL, and an additional 6.9% were attributed to Coxhead's AWL. However, a considerable portion (21%) of the corpus remained unaccounted for by existing academic word lists. In contrast, in the field of anatomy, only about 53.3% of the total 350,000 words were traceable to GSL and 3.7% to AWL, leaving a significant 43%, comprising 31% technical vocabulary and 12% low-frequency words. These results emphasize that the differences in vocabulary distribution depend on the specific field of study and the importance of developing subject-specific vocabulary lists.

2.2 Discipline-Specific Word List

In the pursuit of enhancing the teaching and learning of academic lexis, a range of specialized vocabulary lists have been created over time. Stemming from a longer history of research in English for Science and Technology, the earliest investigations emerged from the field of engineering science. Since then, diverse efforts have been dedicated to constructing vocabulary lists pertinent to specific fields, including agricultural studies (Martinez, Beck and Panza 2009), business (Konstantakis 2007, Hsu 2011), chemistry (Valipouri and Nassaji 2013), engineering (Mudraya 2006, Ward 2009, Hsu 2014), environment (Liu and Han 2015), finance (Li and Qian 2010), food (Esfandiari and Moein 2015), medical science (Hsu 2013, Lei and Liu 2016, Wang, Liang and Ge 2008), nursing (Yang 2015), social studies (Kwary and Artha 2017), and more (further details in Kim and Lee 2019).
The recent growth in English for Specific Purposes (ESP) within the Korean research circle has also led to the development of specialized vocabulary lists tailored to various professional fields such as British drama (Jeong 2018), convention English (Kwon 2013), airline cabin crew services (Cho 2015), advertising (Ha 2020), film (Lee 2020), and military English (Kwon et al. 2023). Most of these subject-specific word lists were constructed through the compilation of corpora and generated a word list against the word families in GSL and/or AWL.

In linguistics, earlier lists were formulated based on research articles. Vongpuvititch, Huang and Chang (2009), drawing from a 1,554,032 word corpus derived from 200 research articles in 5 journals, compiled a list of 603 word families, featuring 475 AWL and 128 non-AWL word-families, representing 11.12% and 2.8% of linguistic research articles, respectively. Khani and Tazik (2013) expanded the corpus size by collecting 240 research articles out of 12 journals in applied linguistics and curated 573 AWL and 200 non-AWL word types amounting to 12.48% of 1,553,450 running words. Likewise, Moini and Islamizadeh (2016) extracted 1,263 word families, reporting a 72.48% coverage by GSL and a 10.2% coverage by AWL with additional 224 word families representing 5. 07% of the entire linguistics corpus. In contrast, more recent studies have shifted focus towards vocabulary in major academic textbooks. Kim and Lee (2019) analyzed a corpus of 1,141,830 running words from five linguistics textbooks widely employed in the foundation courses for English majors, uncovering that English majors need the knowledge of the most frequent 7,000 word families for a 95% coverage of linguistics textbooks. They also curated the Linguistics Academic Vocabulary List (LAVL), consisting of the most frequent 607 word families equating to 11.05 % of the entire corpus, with an overlap of 208 words sourced from AWL. Lee and Kim (2020) further delved into semantics and pragmatics, reporting that the acceptable comprehension of semantics/pragmatics textbooks requires students to acquire a vocabulary of 5,000 word families plus additional categories, with mastery of 409 word families specific to semantics and pragmatics. Among the 409 word families, 206 words overlapped with the LAVL, leaving 203 word families unique to semantics and pragmatics.

More relevant to the present study is the research on the vocabulary lists for English Education. Lee and Kim (2013) constructed English Education corpus (EECO) totaling 300,000 running words sourced from university textbooks used within specific curriculum courses. They found the combined lexical coverage of GSL and AWL reached 91.83%, notably higher than the coverage reported by Coxhead (2000), which ranged from 79.8 to 88.8%. Excluding words from GSL and AWL, they further generated a list of Technical Vocabulary for English Education (TV4EE) of 181 word families, representing 3.24% of EECO. Later studies (Chung 2014, Ha 2018) compiled a larger corpus based on research articles. Chung (2014) found the most frequent 45 word families beyond GSL and AWL made up 2.44% of a corpus of 55 research articles. He also examined the differing ratios of AWL coverages across sub-corpora: AWL in Korean-published research articles in ELT amounted to 11.83%, slightly exceeding the portion of 11.49% in ELT articles in England and the USA. In contrast, teaching English as a native language in the USA had only 9.52% AWL coverage. These findings indicated distinct lexical choices within ELT subfields, contingent on the specific subject or theme. More recently, Ha (2018) complied a corpus of 9,186,721 running words based on 387 research articles published in domestic and international journals. He developed a word list consisting of 744 word families for English education majors, integrating 457 GSL words, 215 AWL words, 23 technical words, and 49 additional terms. In Ha's study, the combined coverage by GSL and AWL averaged at 80.92%, with a coverage rate of 79.98% for texts by Korean writers and 81.82% for those by foreign writers, both of which were notably lower compared to the accumulated coverage of 91.83% in Lee and Kim (2013), which could be attributed to the nature of research articles. Table 1 summarizes vocabulary lists in linguistics and English education, including the corpus size, data types, and coverage percentages by word lists.

Table 1. Vocabulary List in Linguistics and English Education

Area	Study	Data	Corpus size[a]	GSL (%)	AWL (%)	Discipline-Specific WL	
						Size	%
Linguistics	Vongpumivitich et al. (2009)	RAs	1.55	NA	11.17	603	14.07
	Khani & Tazik (2013)	RAs	1.55	76.04	11.96	773	12.48
	Moini & Islamizadeh (2016)	RAs	4.00	72.48	10.18	224	5.07
	Kim & Lee (2019)	TBs	1.14	80.59[b]	NA	607	11.05
	Lee & Kim (2020)	TBs	1.37	79.57[b]	NA	409	10.35
English Education	Lee & Kim (2013)	TBs	0.30	80.71	11.12	181	3.24
	Chung (2014)	RAs	0.25	75.88	11.83	45	2.44
	Ha (2018)	RAs	9.19	70.18	10.74	72	3.95

Note. RAs = Research articles; TBs = Textbooks; NA = Not available
[a]The size of each corpus is presented on a million-word basis
[b]The coverage up to 2nd 1000 word families on BNC/COCA

Although the knowledge of GSL and AWL words aids students in comprehending textbooks and research articles in linguistics and English education, their combined utilization falls short of achieving the desired 95% lexical coverage for proficient reading. While GSL and AWL have proven instrumental for general usage in diverse contexts, they cannot effectively capture the specialized, technical vocabulary vital in the ELT field, thereby underscoring the necessity for a discipline-specific word list tailored to ELT requirements.

3. Research Methodology

Drawing from previous studies (Hsu, 2014; Kim and Lee 2019; Lee and Kim 2020), the current study adopted

a layered method to generate a discipline-specific vocabulary list pertaining to the fundamental domains of ELT. This approach compares the frequency of occurrence of a word in the specialized text against that of a well-established general large corpus. The present study curated a high-frequency vocabulary list against Nation (2012)'s twenty-five BNC/COCA 1000-word-family lists, which was designed for learners of English as a foreign language. Since this approach is designed for intermediate to advanced learners, assuming that the target learners are already acquainted with the vocabulary from the basic word list (Surtees and Horst 2013), it suits well with the context of preservice ELT teachers in the study.

3.1 Construction of the Corpus

A corpus was created from major ELT textbooks that were commonly utilized in the pre-service teacher training programs. By examining multiple ELT syllabi available online, the study collected a series of textbooks recurrently employed in ELT curricula and grouped them into two primary areas: Acquisition (A) and Methodology (M). Then all selected textbooks, initially available in PDF format, were converted into 12 separate text files. Guided by previous research (Hsu 2014; Kim and Lee, 2019; Lee and Kim, 2020), the materials were thoroughly reviewed and refined by removing extraneous elements such as the front matter, headers, illustrations, bibliographic references, and indices. The finalized corpus was named the ELT Textbooks Corpus (ELTTC henceforth), containing a cumulative total of 1,579,852 running words with its two distinct sub-corpora: ELTTC-A, comprising 659,578 words from acquisition textbooks and ELTTC-M incorporating 920,274 words from methodology textbooks. Table 2 gives the composition of ELTTC including the sizes of tokens, the type/token ratio (TTR) and standardized type/token ratio (STTR) measures.

Table 2. Composition of the ELT Textbooks Corpus (ELTTC)

Areas (corpora)	Textbooks[a]	Types	Tokens	TTR	STTR
Acquisition (ELTTC-A)	Book 1	9,636	130,649	7.38	43.21
	Book 2	9,877	195,844	5.04	38.12
	Book 3	5,128	71,460	7.18	36.22
	Book 4	5,523	87,508	6.31	38.81
	Book 5	6,703	94,943	7.06	40.02
	Book 6	5,994	79,174	7.57	41.55
	Sub-total	42,861	659,578	6.76	39.66
Methodology (ELTTC-M)	Book 7	9,319	161,255	5.78	42.25
	Book 8	12,966	284,715	4.55	41.26
	Book 9	9,953	198,836	5.01	39.18
	Book 10	6,321	86,987	7.27	37.71
	Book 11	7,725	105,334	7.33	42.73
	Book 12	6,162	83,147	7.41	39.75
	Sub-total	52,446	920,274	6.23	40.48
Total		95,307	1,579,852	6.49	40.27

[a]The list of selected textbooks is provided in Appendix A.

3.2 The Program

To assess the lexical threshold and retrieve the inventory of the word families of the ELTTC, the study utilized the RANGE program (Nation 2012). The RANGE software was based on comprehensive word family lists sourced from the large contemporary corpora: British National corpus (BNC) and the Corpus of Contemporary American

English (COCA). With its built-in reference BNC/COCA twenty-five 1,000-word-family lists [1] categorized according to the difficulty levels, RANGE is designed to tally the frequency of the most common words required until a specified level of lexical coverage is achieved. Since the lexical coverage progresses within the BNC/COCA word lists, incremented by every 1000-word-family band, the vocabulary demand is computed by adding the coverage percentage of each band starting from the initial 1000-word-family band and advancing through higher 1,000-word-family bands until the cumulative coverage reaches the lexical threshold of 95%, which corresponds to the vocabulary load necessary for the minimum desired level for comprehending ELT textbooks.

When the ELTTC was run on the RANGE program, the output displays the frequency of vocabularies for each 1000 word-list band of 25,000 BNC/COCA plus four additional lists including proper nouns, interjections, compounds, and abbreviations, categorized as Basewrd31, Basewrd32, Basewrd33, and Basewrd34 respectively. Following the previous studies (Hsu 2014, Kim and Lee 2019), the words extracted under 'Not found in any list' by RANGE were further examined and sorted into their respective Baseword lists as initially defined. This process resulted in the creation of an extra category 'glossary' to accommodate a list of technical words that are not present in the BNC/COCA word levels but have a domain-specific meaning and use.

3.3 Extracting the Word List

The discipline-specialist vocabulary list for reasonable comprehension of ELT textbooks was extracted following the method of prior research (Nation 2006, Hsu 2014). Initially, the cumulative coverage percentages of the first 2,000 word-families were established as the baseline vocabulary load and those of the additional categories were calculated and appropriately integrated prior to those of the first 2000 words, as these marginal words pose minimal learning challenges for vocabulary acquisition (Nation 2006, Hsu 2014). Then, the cumulative percentage of running words covered up to the first 2000 words was then deduced from the target 95 lexical coverage. This process resulted in the creation of the ELT Vocabulary List (ELTVL henceforth) – a specialized compilation of vocabulary pertinent to understanding ELT textbooks with a lexical coverage of 95% beyond the first 2000 words. Given that formal English education during the secondary level covers the 2,000 word families, the focus was shifted towards words beyond the first 2,000 words to cater to the need of ELT majors at the tertiary level.

When selecting target vocabulary, further criteria have been adopted from the relevant literature (e.g., Coxhead 2000, Hsu 2014, Kim and Lee 2019, Lee and Kim 2020):

(1) Specialized occurrence: The word families are beyond the first 2,000 word families.
(2) Range: Members of a word family appear at least across 7 out of 12 textbooks.
(3) Frequency: Members of a word family occur at least 72 times across textbooks in the ELTTC.

Thus, the range of 7 (60 %) and frequency of 72 (6 occurrences by 12 textbooks) were chosen after repeated experiments to fulfill the targeted lexical coverage of 95 percent. The detailed discussion of this procedure will be presented in Section 4.2 below.

---

[1]Following Nation, a word family consists of a base or headword, "its inflected forms and its closely related derived forms (2001)." Nation's (2017) word families in the BNC/COCA list adhere to Bauer and Nation's (1993) classification, including verb inflections, plurals, prefixes and suffixes, up to level 6, but not level 7, i.e. classical roots and affixes. As a result, Nation's BNC/COCA list includes a slightly different set of word families from AWL.

4. Results and Discussion

4.1 The Lexical Load of the ELTTC

Regarding Research Question 1, the vocabulary load for preservice teachers for moderate comprehension of the ELT textbooks was estimated by utilizing BNC/COCA based word lists. The results are summarized in Table 3 including the frequency of the words for each level of the BNC/COCA word list in addition to the coverage and cumulative percentage of the ELTTC.

Table 3. BNC/COCA base word list in the ELTTC

Level	Tokens	%	Cumulative %	Word Families
Proper noun	29,798	1.88	1.88	-
Interjection	2,214	0.14	2.02	-
Compound	8,182	0.52	2.54	-
Abbreviation	6,602	0.42	2.95	-
Glossary	2,478	0.16	3.11	-
1st	1,116,162	70.46	73.57	1000
2nd	191,188	12.07	**85.64**	969
3rd	134,732	8.51	94.15	960
4th	*23,032*	*1.45*	*95.60*	*830*
5th	14,207	0.90	96.50	685
6th	6,581	0.42	96.91	597
7th	7,654	0.48	97.40	493
8th	4,848	0.31	97.70	427
9th	1,468	0.09	97.80	353
10th	2,011	0.13	97.92	270
11th	*1,196*	*0.08*	*98.00*	*225*
12th	840	0.05	98.05	149
13th	797	0.05	98.10	135
14th	985	0.06	98.16	113
15th	407	0.03	98.19	83
16th	339	0.02	98.21	81
17th	587	0.04	98.25	71
18th	214	0.01	98.26	52
19th	189	0.01	98.27	40
20th	145	0.01	98.28	52
21st	169	0.01	98.29	37
22nd	89	0.01	98.30	32
23rd	96	0.01	98.30	22
25th	20	0.00	98.31	15
25th	108	0.01	98.31	19
Not in the lists	26, 713	1.51	100.00	
TOTAL	1,584,051	100.00		

The first 1000 word families on the BNC/COCA list accounted for 70.46% of the total words in ELTTC and the second 1000 word families contributed 12.07%. When combined, the cumulative coverage percentage of the

first two 1000 word families on BNC/COCA amounted to 82.53%. This coverage percentage slightly exceeded 80.59% of the general linguistics textbooks by Kim and Lee (2019) and 79.57% in semantics/pragmatics text by Lee and Kim (2020). When incorporating the additional categories, cumulative coverage point reached 95.60% as indicated in Table 3 (highlighted in bold), corresponding to the fourth 1000 word families. In other words, in order to attain a lexical coverage of 95% for ELT textbooks, prospective ELT teachers need to acquire vocabulary up to the fourth level of the BNC/COCA lists encompassing 4,000 word families, in addition to proper nouns, interjections, abbreviations, compounds and the glossary. Notably, this lexical requirement is comparatively lower than 7,000 word families in general linguistics textbooks in Kim and Lee (2019) and 5,000 word families in semantics/pragmatics texts in lee and Kim (2020). This discrepancy could be interpreted as ELT textbooks containing a broader selection of general-service words. However, opting for the 98% coverage level of comprehension of the ELT textbooks, there is a striking disparity in the vocabulary threshold needed. Despite the gradual decrease in the proportion of word families across frequency bands, preservice teachers would necessitate a vocabulary size of 11,000 word families to achieve the 98% lexical coverage. Given that they are mandated to pass the National Teacher Certification Exam, it becomes crucial for them to reach the level of unassisted reading.

Meanwhile, there found a marked distinction in the distribution of the additional categories within the ELTTC in comparison to their frequency of occurrences in Kim and Lee's (2019) linguistics textbooks. Specifically, the supplementary categories in the ELTTC –transparent compounds (.52%), abbreviations (.42%), and glossary (.16%) - displayed notably higher proportions. In contrast, these same categories had lower proportions within Kim and Lee's (2019) linguistic corpus: transparent compounds (.19), abbreviation (.25) and glossary (.05). As a result, the cumulative occurrences of these supplementary categories increased in the current study, lowering the lexical demand for the ELT textbooks. This approach ensures that the estimation of lexical demand remains focused on the core content of the ELT textbooks.

4.2 The ELT Word List (ELTVL)

In addressing Research Question 2, which aims to determine the specific types of words that make up the ELT Vocabulary List (ELTVL) necessary for a lexical coverage of 95% of ELT textbooks, two-stage analyses were conducted. First, the threshold for the ELTVL was calculated by deducting the cumulative coverage percentage obtained from the first 2000 word levels (amounting to 85.64% on Table 3) from the desired total coverage of 95%. As a result, the remaining 9.36 lexical coverage (95% - 85.64% = 9.36%) will constitute the ELTVL.

Then, with the predefined inclusion criteria requiring a minimum of 72 instances in at least 7 out of 12 textbooks (as discussed in section 3.3), the compilation of ELTVL resulted in a total of 513 word families. The last entry in the ELTVL was *metalinguistic* at the 17th level of BNC/COCA, with a collective count of 99 appearances distributed across 9 textbooks. Among the 513 word families included in the finalized list, 235 were found to overlap with Coxhead's (2000) AWL, while the remaining 278 word families were distinctively associated with the ELT domain. These overlapping items accounted for a lexical coverage of 4.29%, thereby affirming the instrumental nature of the AWL. While the portion of AWL words within the ELTVL again reiterates its vital role in comprehending academic textbooks, it is valid to question the generality of AWL, given the significant amount of specialized vocabulary within ELT. Thus, the integration of the AWL and specialized non-AWL vocabulary unique to the ELT field can effectively guide pre-service teachers in their academic pursuits. Table 4 provides the detailed overview of word frequency and distribution of the selected words across the BNC/COCA lists with the proportions of AWL coverage. The details of the ELTVL can be found in Appendix B.

Table 4. ELTVL 513 Word Families across the BNC/COCA

BNC/COCA	AWL	Non-AWL	ELTVL in Total
3rd	213	150	363
4th	18	44	62
5th	4	22	26
6th	0	15	15
7th	0	16	16
8th	0	14	14
9th	0	2	2
10th	0	4	4
11th	0	2	2
12th	0	2	2
13th	0	2	2
14th	0	3	3
15th	0	1	1
17th	0	1	1
Total	235	278	513

As the frequency occurrences of the ELTVL based on the BNC/COCA show in Table 4, only the words in the 3rd through 5th 1000 word families are identified as part of AWL. Furthermore, as the BNC/COCA levels progress through higher levels, there is a noticeable decrease in the proportion of AWL words and a simultaneous increase in the presence of the ELT-specific vocabulary not covered by the AWL. Among the ELTVL belonging to the 3rd through 5th BNC/COCA levels, the top 30 words are displayed in the order of total frequency in Table 5, where AWL words are highlighted in bold.

Table 5. Top 30 Most Frequently Used Words in the ELTTC on BNC/COCA

Rank	Word	Range	Frequency	Level	Rank	Word	Range	Frequency	Level
1	**communicate**	12	3003	3	16	content	12	1422	3
2	**task**	12	2908	3	17	**error**	12	1400	3
3	**acquisition**	12	2601	3	18	*grammatical*	12	1370	7
4	**interact**	12	2389	3	19	**analyze**	12	1279	3
5	**method**	12	2151	3	20	**appropriate**	12	1265	3
6	*linguistic*	12	2098	4	21	**principle**	12	1223	3
7	**context**	12	2097	3	22	*vocabulary*	12	1163	5
8	grammar	12	2048	5	23	*cognitive*	12	1057	4
9	**structure**	12	2044	3	24	**target**	12	1055	3
10	focus	12	2004	3	25	**motive**	12	1004	3
11	**text**	12	1818	3	26	*competence*	12	999	4
12	**strategy**	12	1674	3	27	**data**	12	974	3
13	**input**	12	1644	3	28	**technique**	12	974	3
14	**theory**	12	1569	3	29	verb	12	968	5
15	**function**	12	1481	3	30	**factor**	12	964	3

Note. The **bold** words align with Coxhead's (2000) AWL; The *italicized* words correspond to Lee and Kim's (2013) TV4EE.
Note. Nation's (2012) BNC/COCA list treats *acquisition* and *acquire* as separate word families, whereas Coxhead (2000) classifies them to belong to the same category. The current study adopts Nation's classification.

While more than half of the high-frequency words listed in Table 5 above are identified as part of Coxhead's

(2000) AWL, a set of non-AWL word families, such as "cognitive," "grammatical," "vocabulary," "linguistic," and "competence," also fall within the highest frequency group of 45 words in Lee and Kim's (2013) Technical Vocabulary for English Education (TV4EE), yet with "linguistics" holding the highest total count.

During the process of vocabulary selection, it became apparent that certain world families demonstrated notable prominence within one sub-domain, but not in the other. In essence, the distribution of vocabulary seems to hinge upon the specific sub-domains. For example, the term "collocate" was found in 7 textbooks, but its total occurrences reached 140 with a notable concentration in 5 methodology textbooks (119 cases in ELTTC-M). Similarly, "declarative" appeared 94 times across 7 textbooks, predominantly within the 6 acquisition textbooks (85 times in ELTTC-A), with just 9 instances in a single methodology textbook. This pattern also extends to words that fail to meet the aforementioned inclusion criteria of at least 72 occurrences across 7 textbooks. The word "automaticity," for instance, appeared across 9 ELT textbooks, but their overall occurrences were limited to 63 instances with higher frequency in acquisition textbooks (40 cases in ELTTC-A). Likewise, the term "classification" occurred 55 times across 9 ELT textbooks, but the instances were largely concentrated in methodology textbooks (41 cases ELTTC-M). This observation calls for the need for further analyses to delve into the distinction between different ELT sub-domains and further elucidate the differences.

4.3 Cross-Comparison of ELT Sub-Corpora: Acquisition versus Methodology

Research Question 3 examines potential variations in the lexical threshold and the frequency of discipline-specific academic words between sub-corpora, which incorporates the following questions: How many words are required for students to comprehend the ELT textbooks in acquisition compared to those in methodology? What is the frequency and dispersion pattern of discipline-specific academic words in each corpus?

4.3.1. The Lexical Demands across ELT Sub-Domains

In order to examine the variations in lexical demands across different ELT domains, the study repeated the same procedure (as used in 4.1) with two sets of sub-corpora: ELTTC-A (acquisition) and ELTTC-M (Methodology). Results indicated that the textbooks in acquisition demand higher level of lexical load than those in methodology. To achieve the lexical coverage of 95%, acquisition textbooks require preservice teachers to have a minimum of 5,000 words, whereas methodology textbooks necessitate a minimum of 3,000 words. Table 6 displays the word frequencies and coverage percentage of each corpus.

In acquisition textbooks, the initial 1000 word families on the BNC/COCA list constituted 67.42% of the total words in the ELTTC-A, while the subsequent 1000 word families contributed 12.66%. When combined, the cumulative coverage percentage of the first two sets of 1000 word families from BNC/COCA reached 80.08%. Taking into account the occurrences of supplementary categories including proper nouns, interjections, compounds, abbreviations, and glossary, the cumulative coverage percentage up to the two thousand BNC/COCA levels rose to 83.80 (highlighted in bold in Table 5). This indicates an additional lexical load of 11.2% (95%-83.80%=11.2%) for achieving acceptable comprehension of acquisition textbooks. Alternatively, to attain a lexical coverage of 95% for acquisition textbooks, prospective ELT teachers need to master the vocabulary up to the fifth level around 95.59% encompassing 5,000 word families. Conversely, methodology textbooks demonstrated a comparatively lower level of lexical demand. The combined coverage by the first and second 1000 word families reaching a slightly higher lexical coverage of 84.30%. The cumulative percentage of these initial two sets of 1000 word families, in conjunction with additional categories, increased to 86.97% of the ELTTC-M, indicating a lesser lexical load of 8.03% (95%-86.97% = 8.03%), compared to acquisition textbooks. As a result, in order to achieve

a comprehension level of 95% for methodology textbooks, preservice ELT teachers are required to familiarize themselves with a minimum of 3,000 word families as highlighted in Table 6, corresponding to 95.11%.

Table 6. Comparison of Lexical Thresholds between ELTTC-A and ELTTC-M

Level	ELTTC-A				ELTTC-M			
	Tokens	%	Cum. %	WF	Tokens	%	Cum. %	WF
Proper N.	16,208	2.44	2.44	-	13,590	1.48	1.48	-
Interj.	1,103	0.17	2.61	-	1,111	0.12	1.60	-
Comp.	2,292	0.35	2.95	-	5,890	0.64	2.24	-
Abbrev.	3,737	0.56	3.52	-	2,865	0.31	2.55	-
Glossary	1,351	0.20	3.72	-	1,127	0.12	2.67	-
1st	447,041	67.42	71.14	986	669,121	72.65	75.32	999
2nd	83,924	12.66	**83.80**	867	107,264	11.65	**86.97**	937
3rd	59,778	9.02	92.81	866	**74,954**	**8.14**	*95.11*	**933**
4th	11,141	1.68	94.49	596	11,891	1.29	96.40	757
5th	**7,283**	**1.10**	*95.59*	**445**	6,924	0.75	97.15	591
6th	2,830	0.43	96.02	371	3,751	0.41	97.56	506
7th	3,632	0.55	96.57	311	**4,022**	**0.44**	*97.99*	**391**
8th	2,688	0.41	96.97	250	2,160	0.23	98.23	335
9th	725	0.11	97.08	194	743	0.08	98.31	258
10th	1,084	0.16	97.24	141	927	0.10	98.41	198
11th	558	0.08	97.33	139	638	0.07	98.48	158
12th	413	0.06	97.39	77	427	0.05	98.53	111
13th	432	0.07	97.46	77	365	0.04	98.56	87
14th	565	0.09	97.54	66	420	0.05	98.61	76
15th	212	0.03	97.57	42	195	0.02	98.63	54
16th	227	0.03	97.61	56	112	0.01	98.64	45
17th	344	0.05	97.66	38	243	0.03	98.67	50
18th	87	0.01	97.67	27	127	0.01	98.68	33
19th	105	0.02	97.69	25	84	0.01	98.69	27
20th	63	0.01	97.70	30	82	0.01	98.70	33
21th	62	0.01	97.71	20	107	0.01	98.71	22
22th	55	0.01	97.71	15	34	0.00	98.72	21
23th	40	0.01	97.72	15	56	0.01	98.72	9
24th	8	0.00	97.72	8	12	0.00	98.72	6
25th	**45**	**0.01**	*97.73*	**9**	63	0.01	98.73	14
Not listed	15,029	2.27	100.00		11,684	1.27	100.00	
TOTAL	663,062				920, 989			

Note. Cum. % = Cumulative % coverage in tokens; WF = Word Families

   In summary, achieving a 95% lexical coverage of acquisition textbooks entails familiarity with the most frequent 5,000 word families alongside proper nouns, interjections, abbreviations, transparent compounds, and glossary, whereas methodology textbooks require a smaller vocabulary size of 3,000 word families. The varying lexical requirements between these two sub-domains highlight a notable distinction – a greater proportion of the initial 2,000 words in methodology books surpasses the coverage found in acquisition textbooks. The gap in vocabulary demands becomes more pronounced when the preservice teachers aim for a 98% lexical coverage, as acquisition textbooks require a substantial amount of vocabulary beyond the 25th BNC/COCA lists. While not a fixed parameter, the clear gap in lexical thresholds between these two sub-corpora underscores that ELT majors need a more extensive vocabulary repertoire when engaging with acquisition textbooks.

4.3.2 Cross-Comparison of the Most Frequent 30 Words in the ELTVL

　　Further comparison is made on the frequency and dispersion of discipline-specific academic words within the two sub-corpora. Notably, an analysis of the top 30 words in each list also marks a discrepancy between the two lists. Table 7 presents the most repeatedly occurring words in the following comparative lists with the information about the range of occurrences and the frequency on the corresponding BNC/COCA level of each word.

Table 7. Top 30 Words in the ELTVL

ELTVL Rank	ELTVL-A Word	Frequency	Level	ELTVL Rank	ELTVL-M Word	Frequency	Level
3	acquisition	2239	3	2	task	2009	3
4	interact	1374	3	1	communicate	1981	3
6	linguistic	1362	4	5	method	1724	3
13	input	1251	3	11	text	1620	3
14	theory	1084	3	16	content	1218	3
9	structure	1060	3	10	focus	1185	3
7	context	1032	3	8	grammar	1166	5
1	communicate	1022	3	7	context	1065	3
2	task	899	3	4	interact	1015	3
15	function	899	3	9	structure	984	3
8	grammar	882	5	12	strategy	983	3
17	error	876	3	28	technique	876	3
10	focus	819	3	20	appropriate	856	3
19	analyze	778	3	22	vocabulary	797	5
27	data	774	3	21	principle	743	3
18	grammatical	766	7	6	linguistic	736	4
36	hypothesis	738	3	18	grammatical	604	7
23	cognitive	732	4	15	function	581	3
12	strategy	691	3	24	target	564	3
31	acquire	681	3	41	oral	557	3
26	competence	662	4	52	assess	554	3
29	verb	613	5	53	evaluate	551	3
30	factor	599	3	40	effective	550	3
59	universe	546	3	34	phrase	530	3
47	utter	543	4	45	proficient	528	7
25	motive	525	3	17	error	524	3
38	concept	494	3	71	syllabus	523	7
24	target	491	3	70	curriculum	517	3
21	principle	480	3	32	response	509	3
37	complex	468	3	49	procedure	504	3

Note. Underlining is used to indicate the absence of overlap between the two lists.

　　Among the top 30 words in the ELTVL, the following 14 words are overlapped: *communicate, task, interact, linguistic, context, grammar, structure, focus, strategy, function, error, grammatical, principle,* and *target.* Noticeable is that both lists contain specific terms relevant to their respective sub-domains as underlined in Table 7 above. This divergence is predictable given the inherent nature of the two corpora being examined. Excluding the overlapping words, the acquisition-specific word list is characterized by lexemes such as *acquisition, input, theory, analyze, data, hypothesis* and *cognitive,* which are undoubtedly pivotal for formulating theories about the learning process itself. Conversely, the methodology-specific word list features terms like *method, text, content, technique, appropriate, vocabulary,* and *oral,* which are closely tied to the practical strategies and methodologies

employed in the instructional process. It is clear that the non-overlapping words in each sub-corpus are highly relevant to their respective domains, underscoring the specificity of vocabulary within each sub-corpus. These findings highlight the importance of developing specialized vocabulary lists tailored to specific domains. Such domain-specific knowledge would better meet the needs of ELT majors on the distinct academic requirements.

4.4 Features of Additional Categories

The distinct nature of ELT vocabulary was further elucidated with the additional categories. Of particular significance are the glossary words, due to their inclusion of a substantial number of technical terms directly pertinent to ELT domains. The glossary category comprises words that are not found in any of the BNC/COCA word lists but possess a strong technical sense in the context of ELT textbooks. Examples of such terms include *"audiolingual," "interlanguage,"* and *"sociocultural."* Another group of words with a strong technical use in high frequency took the form of transparent compounds, such as *"classroom," "feedback,"* and *"textbook"* as well as abbreviations like *"L1," "L2," "SLA," "ESL."* Each of these terms indeed represents a distinct component in the ELT domain. Based on the same criteria for the inclusion in the ELTVL (requiring a minimum of 72 occurrences within a range of 7 out of 12 textbooks as outlined in Sections 3.3 and 4.2), a sum of 26 ELT-specific words were identified from the glossary, compounds, and abbreviations categories. Table 8 provides a comprehensive overview of the frequency and range of this ELT-specific list retrieved in the additional categories in the ELTTC, juxtaposed with their corresponding occurrences in each sub-corpus.

Although these categories were not included in the compilation of the ELTVL due to their minimal learning load (as stipulated by Hsu 2014), it is essential for the students of ELT major to familiarize themselves with these terms, since these high-frequency terms carry domain-specific meanings. Following Chung and Nation (2004), these words are functionally classified as technical vocabulary since their meanings are closely related to the subject area with a significantly higher frequency of occurrences within the specific field, contributing to subject-specific knowledge. As expected, several words in Table 8 coincide with the ELT vocabulary highlighted in prior research. Among these words, *classroom, feedback,* and *textbook* have consistently been classified as a part of the high-frequency group in previous studies. About half of the high-frequency words in Table 8 were also recognized as ELT-specific vocabulary by Lee and Kim (2013); *classroom, L1/L2, feedback, SLA, ESL, textbook, interlanguage, EFL, sociocultural, classmate, online,* and *audiolingual* (boldfaced in Table 8). Similarly, Ha's (2018) technical and other word lists overlap with the underlined terms like *feedback, nonnative, sociocultural, classroom, textbook, online,* and *long-term*. The commonalities across these words provide insight into the nature of ELT vocabulary, as many of these jargons point to the interactional aspects of language learning and instruction in a classroom setting.

A noteworthy observation regarding ELT textbooks is their abundant utilization of abbreviations and compound nouns to represent various ELT methods and principles, which could also fall under the category of glossary. While the strategic use of such condensed language helps convey complex concepts efficiently, the concentrated presence of such language in one sub-corpus highlights the distinctiveness of the two sub-domains, particularly concerning words that cannot meet the selection criteria of the minimum of 72 occurrences across 7 textbooks. For instance, the abbreviation "ESP" (English for Specific Purposes) was present in 5 methodology books, totaling 121 instances, while "ELT" was found in 6 methodology books with 87 occurrences. Remarkably, both terms were entirely absent from acquisition textbooks. Another recurrent abbreviation was "TBLT" (Task-Based Language Teaching), along with such variations as "TBL" and "TBI," which amounted to 79 instances across 5 methodology books. In case of compounds, the term "jigsaw" appeared across 9 textbooks, yet the total occurrences were

confined to 45 with 37 instances exclusively present in all 6 methodology books. Similarly, the word "website" occurred a total of 113 times across 5 methodology textbooks, but only once in a single acquisition textbook.

Table 8. ELT-Specific Words in Additional Categories of the ELTTC

Category	Word	Range	Frequency	ELTTC-A			ELTTC-M		
				Rank	Range	Frequency	Rank	Range	Frequency
CMP	**classroom**	12	3042	3	6	791	1	6	2251
ABBR	**L2**	12	1743	1	6	1380	12	6	119
CMP	**feedback**	12	1065	5	6	518	3	6	547
ABBR	**SLA**	12	867	2	6	827	21	6	40
ABBR	**ESL**	12	836	10	6	120	2	6	716
ABBR	**L1**	12	803	4	6	684	5	6	363
CMP	**textbook**	11	498	16	5	37	4	6	461
GLS	**interlanguage**	10	416	6	6	390	24	4	26
ABBR	i.e.	11	368	9	6	175	9	6	193
ABBR	**EFL**	8	336	25	2	10	6	6	326
GLS	**sociocultural**	12	334	8	6	241	17	6	93
ABBR	NS	7	329	7	5	299	23	2	32
ABBR	CLT	9	246	17	3	32	7	6	214
GLS	group-work	9	223	19	4	20	8	5	203
GLS	role-play	10	183	19	4	20	10	6	171
GLS	pair-work	10	163	18	4	25	11	6	138
CMP	**classmate**	9	134	19	4	20	14	5	114
CMP	online	8	130	24	4	13	13	4	117
ABBR	TESOL	8	129	22	2	19	15	6	110
GLS	**audiolingual**	12	124	14	6	48	18	6	76
CMP	non-native	7	113	11	4	94	25	3	21
CMP	long-term	10	110	15	5	41	20	5	69
CMP	homework	10	105	26	4	7	16	6	98
CMP	self-esteem	9	97	12	4	60	22	5	37
CMP	**checklist**	7	87	23	2	15	19	5	72
ABBR	CA	10	75	13	5	57	26	5	18

Note. CMP = transparent compounds; ABBR = abbreviations; GLS = glossary
Note. **Bold** words overlap with Lee and Kim's (2013) list; <u>Underlined</u> words correspond to Ha's (2018) list.

A parallel pattern also emerges with acquisition textbooks, where there is a clear prominence of vocabulary used in theorizing perspectives and models in language learning. For example, the term "connectionism" appeared a total of 98 times, with a noteworthy concentration of 93 instances across 6 acquisition textbooks. Comparably, among abbreviations, "UG" occurred 260 times exclusively within 6 acquisition books, while "ZPD" appeared 68 instances across acquisition books. An intriguing pattern was further identified with the use of the abbreviations that designate interaction partners or interlocutors in conversations, as well as the roles of student and teacher in instructional practices. The acronym "NS" (Native Speaker) appeared in various forms like "NES (Native English Speakers)" and "NSE (Native Speakers of English)," while its counterpart "NNS" (Non-Native Speaker) was seen in forms like "NNS(s)" and "NNSE(s)." The abbreviation "NS" surfaced in 329 instances, mainly in 5 acquisition textbooks with a single methodology textbook. In contrast, its counterpart "NNS" occurred 338 times across 10 textbooks with its full compound noun "non-native" appearing 113 times, rather evenly distributed between the two domains. In acquisition textbooks, "NS" and "NNS" were typically used in an authentic NS - NNS dyad often engaged in negotiated interaction, but the use of "NS" did not automatically imply the role of a teacher providing

input or feedback facilitating English Language Learners' (ELL) in their communication skills. Meanwhile, the acronym "NNS" with the meaning of the teacher as opposed to ELL had another pair of variation, NEST (The native English-speaking teacher) versus non-NEST, though this unique type was not counted in the total occurrences as the identical word families. This particular pattern suggests the two domains within ELT differ not only in the distribution of these technical vocabulary but also vary in their types.

To sum up, the analysis of ELT sub-domains reveals a unique and specialized vocabulary that is integral to the discipline. Glossary words, transparent compounds, and abbreviations are all considered functionally technical, as they serve specific purposes in conveying complex ELT concepts efficiently, with their meanings closely tied to the ELT domain. Notably, the distribution of these technical words varies across sub-domains, further highlighting the distinct features of academic language within ELT. Therefore, familiarity with subject-specific technical vocabulary enhances a deeper understanding of ELT textbooks and the field as a whole.

5. Conclusion

The study aims to analyze the vocabulary load in academic textbooks for students majoring in ELT and to construct a comprehensive list of the most frequently used vocabulary for preservice ELT teachers. For the purposes, the study compiled a corpus (ELTTC) consisting of approximately 1.6 million tokens out of 12 university academic textbooks in ELT. First, it assessed the lexical coverage for adequate understanding of these major ELT textbooks by running ELTTC on RANGE using the built-in twenty-five BNC/COCA 1000-word-family list, Next, the study established a specialist academic vocabulary list in ELT (ELTVL) based on the criteria of at least 72 occurrences across a range of 7 out of 12 ELT textbooks. Additionally, the study compared the specific vocabulary demands between two pivotal sub-areas in ELT- language acquisition and teaching methodology and analyzed the frequency and distribution of discipline-specific academic vocabulary within each corpus, highlighting the patterns of specialized words within each domain.

The findings revealed that reaching a 95% lexical threshold in ELT textbooks requires preservice ELT teachers to be familiar with approximately 4,000 most frequent word families, in addition to proper nouns, interjections, transparent compounds, abbreviations, and glossary terms. Interestingly, this lexical requirement appears to be lower compared to the vocabulary demands of general linguistics textbooks as estimated by Kim and Lee (2019) and semantics/pragmatics texts as noted by Lee and Kim (2020). One possible interpretation of this difference is that ELT textbooks tend to incorporate a broader range of general service words, likely for the purpose of exemplifying discussion questions, classroom practices, activities and exercises. In contrast, linguistics textbooks might focus more on specialized terminology and concepts, which could explain the higher lexical threshold needed for comprehension in that area.

The study also identified a total of 513 word families beyond the initial 2,000 word families, accounting for 9.36% of the entire corpus. Among these 513 word families, 235 were identified to be part of Coxhead's (2000) AWL representing 4.29% of ELT textbooks, while the remaining 278 word families were specifically associated with the field of ELT. Given the considerable amount of AWL in the ELTVL, the findings underscore the crucial role of AWL in comprehending academic textbooks. However, the study also raises questions about the usefulness and universality of the AWL, which is evidenced by the greater portion of additional 278 ELT-specific word families in the current study. These specialized terms are pivotal for ELT majors in achieving a sufficient level of comprehension when studying their academic textbooks.

Further investigation indicated that acquisition textbooks imposed a more substantial lexical demand on ELT

majors. Achieving a 95% lexical coverage in acquisition textbooks necessitates knowledge of the most frequent 5,000 word families, whereas methodology textbooks had a lower vocabulary requirement, with a lexical size of 3,000 word families. Moreover, words extracted from glossaries, transparent compounds, and abbreviations were also considered functionally technical because they serve specific functions in efficiently conveying complex ELT concepts, with their meanings closely linked to each ELT domain. Importantly, the distribution of these technical words varies across domains of acquisition and methodology, underscoring the unique characteristics of academic language within the specific domains of ELT, the knowledge of which enriches the comprehension of ELT textbooks and the entire discipline as well.

The findings suggest some pedagogical implications. First, the assessment of the vocabulary requirements and the identification of word lists in a specific field hold valuable insights for both students and faculty members. This knowledge informs students about their academic preparation and relevant goal to be reached. Familiarity with the high-frequency academic vocabulary list can significantly benefit students in enhancing their academic literacy by focusing their efforts on vocabulary of relevance and immediate needs, leading to higher levels of achievement in their coursework, and ultimately benefit them beyond their academic studies. Instructors can also leverage these insights from vocabulary analysis. They can design or adapt course materials that cater to students' linguistic needs and readiness. Access to a well-curated essential vocabulary list ensures that course content is both suitable and appropriately challenging, allowing them to tailor their teaching strategies and materials to better support their students' comprehension. Additionally, they can utilize the vocabulary index to precisely gauge students' language proficiency and academic readiness, thereby improving their teaching practices and further developing own expertise in teaching profession. Second, the study offers valuable information for materials development and curriculum design, encompassing both the minimum lexical requirements and a comprehensive range of key vocabulary necessary for students' successful academic performance. This resource can empower the materials writers to craft more effective learning materials and equip educators to develop curricula that are finely attuned to the specific linguistic needs of their students. Third, the analysis of the ELT-specific vocabulary not only reveals the common vocabulary but also sheds light on the distinctive terminologies within sub-areas. This insight carries significant implications, emphasizing the necessity of developing customized vocabulary resources for the acquisition and methodology domains within ELT. These resources would be instrumental in meeting the specific academic needs of ELT majors, equipping them with the linguistic tools necessary to engage with the ELT textbooks more effectively. Consequently, the study highlights the practical application of corpus analysis in ELT methodologies, evaluation, and materials development by illustrating how corpus analysis aids in the identification of the specific vocabulary and lexical loads.

The vocabulary list serves as a powerful resource, equipping students with the linguistic capabilities essential for effective communication, greater text comprehension, and meaningful engagement across personal, academic, and professional contexts. For preservice teachers, it becomes imperative to build a strong foundation in academic vocabulary, as the demands for vocabulary become even more critical as they progress in their studies, they encounter an increasingly diverse range of texts across various content areas. Furthermore, the demands placed on preservice teachers for vocabulary growth are continuous and dynamic, given the evolving nature of language and the changing academic landscape. Since such demands require ongoing growth in their academic repertoire, the comprehensive vocabulary toolkit is essential, allowing them not only to excel in their academic pursuits but also to be effective educators who can facilitate meaningful language learning experiences for their students. Ultimately, the ELT vocabulary list serves as a bridge between language learning and professional competence, enabling preservice teachers to navigate their educational journey and excel in their professional development.

References

Armbruster, B. B. 1992. Vocabulary in content area lessons. *The Reading Teacher* 45(7), 550–551.

Bauer, L. and I. S. P. Nation. (1993). Word families. *International Journal of Lexicography* 6(4), 253–279.

Beaumann, J. F. and M. F. Graves. 2010. What is academic vocabulary? *Journal of Adolescent & Adult Literacy* 54(1), 4-12.

Cho, H. 2015. Creating a specialized word list for ESP college students. *The Jungang Journal of English Language and Literature* 57(1), 541-564.

Chung, S.-W. 2014. A corpus-based analysis of the frequency and text coverage of Coxhead (2000)'s academic words in research articles on English education. *Korean Journal of Linguistics* 39(4), 997-1019.

Chung, T. M. and I. S. P. Nation. 2003. Technical vocabulary in specialised texts. *Reading in a Foreign Language* 15(2), 103-116.

Chung, T. M. and I. S. P. Nation. 2004. Identifying technical vocabulary. *System* 32(2), 251-263.

Cobb, T. and M. Horst. 2004. Is there room for an academic word list in French. In P. Boggards and B. Laufer, eds., *Vocabulary in a Second Language: Selection, Acquisition, and Testing*, 15-38. Amsterdam: John Benjamins.

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

Esfandiari, R. and G. Moein. 2015. A corpus-driven food science and technology academic word list. *Issues in Language Teaching* 4(2), 131-157.

Ha, M. 2018. Development of an English education word List: A corpus-based study using Korean foreign articles corpora. *The Jungang Journal of English Language and Literature* 60(4), 377-398

Ha, M. 2020. Exploring academic vocabulary in advertising English corpus. *Studies in Linguistics 57*, 223-243.

Hincks, R. 2003. Pronouncing the Academic Word List: Features of L2 student oral presentations. *In Proceedings of the 15th International Congress of Phonetics Science*, Volume 28, 1545-1548, Barcelona, Universitat Autònoma de Barcelona.

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

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.

Hsu, W. 2014. Measuring the vocabulary load of engineering textbooks for EFL undergraduates. *English for Specific Purposes 33*, 54-65.

Hyland, K. and P. Tse. 2007. Is there an "academic" vocabulary? *TESOL Quarterly* 41(2), 235-253.

Jeong, H. 2018. Development of an English word list for British drama. *The Jungang Journal of English Language and Literature* 60(1), 281-306.

Khani, R. and K. Tazik. 2013. Towards the development of an academic word list for applied linguistics research articles. *RELC Journal* 44(2), 209-232.

Kim, H.-O. and H.-K. Lee. 2019. The academic vocabulary list in linguistics for EFL university students. *Korean Journal of English Language and Linguistics* 19(1), 27-53.

Konstantakis, N. 2007. Creating a business word list for teaching business English. *Estudios de Lingüística Inglesa Aplicada* 7, 79-102.

Kwary, D. A., and A. F. Artha. 2017. The academic article word list for social sciences. *MEXTESOL Journal* 41(4), 1-11.

Kwon, E.-Y., Woo, K. Jin, S. and T.-Y. Jeong. 2023. Development of a basic word list for military English using

big data. *Korean Journal of English Language and Linguistics* 23, 109-130.

Kwon, S. 2013. Profiling the vocabulary lists in a convention English corpus. *English Language Teaching* 25(4), 145-165.

Laufer, B. 1989. What percentage of text lexis is essential for comprehension? In C. Lauren and M. Nordman, eds., *Special Language: From Human Thinking to Thinking Machines*, 316-323. Clevedon, England: Multilingual Matters.

Laufer, B. and I.S.P. Nation. 1995. Lexical richness in L2 written production: Can it be measured? *Applied Linguistics* 16(3), 307-322.

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

Lee, H.-K. and H.-O. Kim. 2020. Constructing a vocabulary list of English textbooks in semantics and pragmatics. *Korean Journal of English Language and Linguistics* 20, 384-410.

Lee, J.-Y. and J.-R. Kim. 2013. Extracting technical vocabulary list for English Education: A corpus-based research. *Foreign Languages Education* 20(1), 99-122.

Lee, S. 2020. A corpus-based study of vocabulary for film English. *Journal of Language Sciences* 27(3), 139-168.

Lei, L. and D. Liu. 2016. A new medical academic word list: A corpus-based study with enhanced methodology. *Journal of English for Specific Purposes* 22, 42-53.

Li, Y. and D. D. Qian. 2010. Profiling the academic word list (AWL) in a financial corpus. *System* 38(3), 402-411.

Liu, J. and L. Han. 2015. A corpus based environmental academic word - list building and its validity test. *English for Specific Purposes* 39, 1-11.

Malvern, D. and B. Richards. 2002. Investigating accommodation in language proficiency interviews using a new measure of lexical diversity. *Language Testing* 19(1), 85-104.

Martinez, I. A., S. C. Beck and C. B. Panza. 2009. Academic vocabulary in agriculture research articles: A corpus-based study. *English for Specific Purposes* 28(3), 183-198.

Moini, R. and Z. Islamizadeh. 2016. Do we need discipline-specific academic word lists? Linguistics Academic Word List (LAWL). *Journal of Teaching Language Skills* 35(3), 65-90.

Moon, Y. (2017). Are different types of vocabulary needed for comprehending general and academic texts? *English Teaching* 72(3), 47-68.

Mudraya, O. 2006. Engineering English: A lexical frequency instructional model. *English for Specific Purposes* 25(2), 235-256.

Nation, I. S. P. 1990. *Teaching and Learning Vocabulary*. New Work: Newbury House.

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. 2012. Range with 25,000 words from the BNC/COCA Word Family Lists (version 1.0.0). Available online at https://www.wgtn.ac.nz/lals/resources/paul-nations-resources/vocabulary-analysis-programs.

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

Nation, I. S. P. and R. Waring. 1997. Vocabulary size, text coverage and word lists. In N. Schmitt and M. McCarthy, eds., *Vocabulary: Description, Acquisition and Pedagogy*, 6-19. Cambridge: Cambridge University Press.

Schleppegrell, M. J. 2004. *The Language of Schooling: A Functional Linguistics Perspective.* New York: Routledge.

Schmitt, N., X. Jiang and W. P. Grabe. 2011. The percentage of words known in a text and reading comprehension.

The Modern Language Journal 95(1), 26-43.

Scott, J. A., Nagy, W. E., & Flinspach, S. L. 2008. More than merely words: Redefining vocabulary learning in a culturally and linguistically diverse society. In A. E. Farstrup and S. J. Samuels, eds., *What Research Has to Say about Vocabulary Instruction*, 182-210. Newark, DE: International Reading Association.

Valipouri, L. and H. Nassaji. 2013. A corpus-based study of academic vocabulary in chemistry research articles. *Journal of English for Academic Purposes* 12, 248-263.

Vongpumivitich, V., J. Huang and Y. Chung. 2008. Frequency analysis of the words in the Academic Word List (AWL) and non-AWL content words in applied linguistics papers. *English for Specific Purposes* 28(1), 33-41.

Wang, J., S. Liang and G. Ge. 2008. Establishment of a medical academic word list. *English for Specific Purposes* 27(4), 442–458.

Ward, J. 2009. A basic engineering English word list for less proficient foundation engineering undergraduates. *English for Specific Purposes* 28(3), 170–182.

West, M. 1953. *A General Service List of English Words*. London: Longman.

Wright, T. S. and G. N. Cervetti. 2017. A systematic review of the research on vocabulary instruction that impacts text comprehension. *Reading Research Quarterly* 52, 203-226.

Xue, G. and I. S. P. Nation. 1984. A university word list. *Language Learning and Communication* 3(2), 215-229.

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

Examples in: English
Applicable Languages: English
Applicable Level: Tertiary

Appendix A

Selected Academic Textbooks in ELT

Brown, H. D. 2007. *Principles of Language Learning and Teaching*, 4th ed. White Plain, NY: Addison Wesley Longman.

Brown, H. D. 2007. *Teaching by Principles: An Interactive Approach to Language Pedagogy*, 3rd ed. Pearson Education.

Celce-Murcia, M., D. M. Brinton, and M. A. Snow. Eds. 2014. *Teaching English as a second or foreign Language*, 4th ed. Boston, MA: Heinle & Heinle.

Gass, S. M., and Selinker, L. 2008. *Second Language Acquisition: An Introductory Course*, 3rd ed. New York, NY: Routledge.

Harmer, J .2007. *The Practice of English Language Teaching*, 4th ed. Harlow, England: Pearson Longman.

Johnson, M. 2004. *A Philosophy of Second Language Acquisition*. New Haven, CT: Yale University Pres.:

Larsen-Freeman, D. and M. Anderson. 2011. *Techniques and Principles in Language Teaching*, 2nd ed. Oxford: Oxford University Press.

Lightbown, P. M. and N. Spada. 2006. *How Languages are Learned*, 3rd ed. Oxford: Oxford University Press.

McDonough, J., Shaw, C. and H. Masuhara. 2013. *Materials and Methods in ELT: A Teacher's Guide*, 3rd ed. Chichester, West Sussex: Wiley-Blackwell.

Mitchell, R. and F. Myles. 2004. Second Language Learning Theories, 2nd ed. London: Hodder Arnold

Richards, J. C. and T. S. Rodgers. 2001. *Approaches and Methods in Language Teaching*, 2nd ed. Cambridge: Cambridge University Press.

Saville-Troike, M. 2006. *Introducing Second Language Acquisition.* Cambridge: Cambridge University Press.

Appendix B
The ELT Vocabulary List (ELTVL)

Note: The ELTVL includes the 513 word families beyond the BNC/COCA 2000 words including 235 AWL words which are highlighted in bold. A number is given after each word to indicate the total frequency occurrences across 12 ELT textbooks.

abstract_203	**contrast_561**	**focus_2004**	mere_143	**relevant_387**
academy_610	**controversy_96**	formal_575	metalinguistic_94	reproduce_75
accompany_114	convention_157	**format_151**	metaphor_164	request_200
accomplish_215	convey_182	formation_206	**method_2151**	**resolve_82**
accurate_532	**cooperate_264**	former_108	**minimal_132**	**resource_479**
achieve_585	**core_118**	**formula_242**	**mode_179**	**respond_579**
acknowledge_120	corpus_170	fossil_138	**modify_279**	response_876
acquire_885	correlate_105	foster_84	module_**101**	**restrict_136**
acquisition_2601	**correspond_182**	**foundation_144**	**monitor_258**	**retain_91**
activate_142	counsel_126	**framework_410**	monolingual_94	**reveal_160**
adequate_153	**criteria_273**	frequency_224	moreover_98	review_619
adjective_133	critic_347	frequent_393	morpheme_264	**revise_147**
adjust_111	criticism_100	**function_1481**	morphology_267	reward_135
administer_80	**crucial_169**	**fundamental_175**	**motive_1004**	rhetoric_93
adopt_221	cue_256	**furthermore_136**	multilingual_161	rhythm_102
adverb_123	curriculum_555	gap_189	multiple_310	rod_150
advocate_141	**data_974**	**gender_157**	narrate_162	rote_70
affection_359	**debate_173**	**generate_253**	**negative_445**	routine_137
aim_256	**decade_158**	genre_250	negotiate_446	salient_103
aloud_97	declarative_94	genuine_112	**network_104**	sample_240
alternative_277	deduct_89	gesture_168	**nevertheless_192**	scaffold_101
ambiguity_72	**define_748**	global_223	**norm_195**	scan_80
analytic_116	delete_92	gradual_176	**notion_382**	schema_123
analyze_1279	**demonstrate_303**	grammar_2048	noun_428	**scope_81**
anticipate_107	**derive_156**	grammatical_1370	novel_76	script_105
anxiety_196	description_462	graphic_97	numerous_133	segment_99
appropriate_1265	**despite_149**	**guideline_156**	**objective_470**	semantic_329
approximate_78	**device_194**	hemisphere_80	oblige_101	sensitive_169
aptitude_252	devise_81	**hence_123**	**obtain_140**	**sequence_722**
arise_124	**devote_88**	**hierarchy_155**	ongoing_103	session_109
articulate_96	dialogue_514	**highlight_125**	optimal_91	significant_319
aspect_755	diary_97	**hypothesis_835**	oral_756	silent_194
assert_72	dictate_100	ideal_**160**	oriented_388	**simulate_107**
assess_671	dictionary_261	idiom_87	**outcome_299**	simultaneous_112
assign_364	differ_360	imitate_216	outline_174	singular_85
assumption_277	**differential_73**	immerse_199	**output_366**	socialise_75
attain_129	differentiate_101	immigrant_124	**overall_226**	sociolinguistic_170
attribute_167	**dimension_215**	**impact_147**	overt_127	software_113
audience_170	disadvantage_86	**implement_197**	overview_133	solution_142
audio_144	disagree_94	implicate_279	pace_95	solve_298
authentic_346	discourse_704	**implicit_243**	**paradigm_119**	**source_523**
author_235	**discriminate_88**	**imply_155**	paragraph_171	**specify_159**
authority_76	**display_143**	importance_410	**parallel_99**	spontaneous_112
autonomous_102	**distinct_449**	**impose_73**	**parameter_249**	**status_153**
autonomy_138	distinguish_258	incomprehensible_252	paraphrase_73	stimulate_144
auxiliary_118	**distribute_111**	**incorporate_253**	participant_436	stimulus_237
behaviour_873	**diverse_108**	independent_220	**participate_432**	**strategy_1674**
belief_260	**document_127**	indirect_174	passage_274	**structure_2044**
bilingual_493	**domain_230**	inductive_105	**passive_146**	subconscious_81

boundary_74
brainstorm_75
candidate_105
capacity_248
capture_92
category_695
characteristic_504
chart_267
chunk_124
cite_183
clarify_185
classify_93
clause_360
cluster_94
code_173
cognition_85
cognitive_1057
coherent_72
collaborate_323
colleague_289
collocate_140
column_97
com_91
communicate_3003
compensate_72
compete_79
competence_999
competent_220
complement_123
complex_831
component_385
compose_247
comprehend_864
comprehensive_105
concept_827
conclude_192
conclusion_282
conduct_300
confer_112
confidence_145
confirm_101
conflict_79
consequence_87
considerable_223
consist_238
consistent_195
consonant_109
constitute_163
constrain_273
construct_824
consult_95
content_1422
context_2097
continuum_145

dominant_130
draft_146
drill_325
dynamic_96
effective_775
efficient_170
ego_73
elaborate_119
electronic_81
element_570
elementary_167
elicit_245
embed_102
emerge_291
empathy_90
emphasis_274
emphasize_310
empirical_159
enable_219
encounter_243
enhance_230
ensure_142
entry_113
equivalent_108
error_1400
essay_138
essential_355
estimate_92
et_331
etc_564
evaluate_655
evident_82
exception_99
excerpt_72
exclusive_116
exemplify_83
expand_123
experiment_291
expert_137
explicit_556
explore_321
extensive_269
extent_352
external_210
extract_114
extrovert_88
facilitate_318
factor_964
faculty_79
failure_115
false_81
filter_88
flexible_94
fluent_436

infer_175
inflect_80
informal_206
inhibit_105
initial_387
initiate_204
innate_223
innovate_90
input_1644
inquire_81
insight_183
institution_287
intake_142
integrate_631
intellectual_92
intelligence_346
interact_2389
interfere_200
interlocutor_145
intermediate_289
internal_384
international_207
interpret_682
intervene_99
intonation_244
intrinsic_190
intuition_88
isolate_172
journal_399
justify_91
label_115
latter_183
lecture_210
lexical_558
lexicon_112
likeness_118
linguist_236
linguistic_2098
link_393
literacy_368
literal_84
literary_114
literature_296
logic_195
longitudinal_96
majority_106
manifest_109
manipulate_139
mature_85
mechanic_87
mechanism_186
media_182
mediate_143
medium_119

pedagogy_405
peer_304
pencil_74
perceive_328
perception_211
periphery_73
permission_113
permit_76
personality_178
perspective_591
phase_199
phenomenon_204
philosophy_122
phoneme_116
phonetic_85
phonology_346
phrase_865
plural_191
pose_98
potential_398
practitioner_77
pragmatic_346
pre_189
precede_180
precise_119
predict_619
preposition_115
primary_515
principal_77
principle_1223
prior_200
priority_100
procedure_696
proceed_94
proficient_724
promote_325
prompt_132
pronoun_177
proponent_78
psycholinguistic_106
psychology_536
publish_329
pursue_77
quantity_124
questionnaire_107
recast_287
receptive_108
recognition_156
reflect_572
regulate_167
reinforce_202
reject_132
relative_572
relevance_78

subordinate_76
subsequent_163
substitute_95
succeed_85
sufficient_173
sum_96
summary_336
superior_75
survey_187
sustain_74
syllable_199
syllabus_545
symbol_173
synonym_75
syntactic_234
syntax_235
target_1055
task_2908
technical_111
technique_974
text_1818
theme_250
theoretical_321
theory_1569
trait_82
transcript_137
transform_119
transition_109
translate_494
trend_109
trigger_73
tutor_161
ultimate_201
underlie_310
undertake_72
unique_136
universe_587
usage_120
utilise_108
utter_714
valid_202
variety_730
verb_968
verbal_384
versus_287
via_83
virtual_137
visual_233
vocabulary_1163
volume_141
vowel_172
zone_86