

The Academic Vocabulary List in Linguistics for EFL University Students*

Hyeon–Okh Kim (Ajou University)
Hye–Kyung Lee (Ajou University)

Kim, Hyeon–Okh and Hye–Kyung Lee. 2019. The academic vocabulary list and amount in linguistics for EFL university students. *Korean Journal of English Language and Linguistics* 19–1, 27–52. The study investigates the vocabulary loads of English–medium linguistics textbooks for EFL students at tertiary level and identifies a Linguistic Academic Vocabulary List (LAVL) in a way of promoting students' academic literacy by enhancing lexical knowledge in English. For the purposes, the study developed a Linguistics Textbooks Corpus (LTC) containing 1,141,830 running words compiled from five major linguistics textbooks used in the foundation courses for English majors. The results indicated the knowledge of the most frequent 7,000 word families in addition to proper nouns, interjections, abbreviations, glossary, and transparent compounds would be in need for 95% lexical coverage of a linguistics textbook written in English. Beyond the first 2,000 word lists covering up to 84.95%, the study further identified a list of 607 word families that correspond to 11.05% of the entire corpus of the linguistics textbooks. The study hopes to provide a way of assisting EFL university students in enriching their lexical competence to attain a reasonable level of understanding in the main textbooks, giving practical information and useful directions to ESP instructors and materials writers so that the EFL students could better benefit from focusing on the words of imminent need and relevance.

Keywords: vocabulary load, linguistics word list, academic vocabulary, EAP, EFL

1. Introduction

Academic vocabulary has long been a topic of interest in linguistics and language education. While a number of factors may contribute to high levels of academic literacy, adequate comprehension of a text equipped with the knowledge of a rich repertoire of academic lexis is key to successful academic achievement (Coxhead 2000). Since lexical

* This work was supported by the Ajou University research fund. An earlier version of this paper was presented at the third Asia Pacific Corpus Linguistics Conference, held in Takamatsu, 2018.

knowledge plays a far more critical role for second language reading (Schmitt, Jiang and Grabe 2011), developing ways to help students enhance their proficiency in English vocabulary has emerged as one of the main concerns in English education.

Academic vocabulary can be broadly defined as the language used in school settings to understand the content written in special registers that are not used in everyday conversation. In the framework of English for Academic Purposes (EAP), academic vocabulary refers to “high-frequency context independent words occurring across disciplines” (Martin 1976: 92). Earlier research, therefore, centered upon the common core vocabulary most frequently and widely used across disciplines, including the classic West’s (1953) General Service List (GSL), Xue and Nation’s (1984) University Word List (UWL), and later Coxhead’s (2000) Academic Word List (AWL). While the lists of the most frequent words in English still provide a considerable portion of words in specialized texts (Chung and Nation 2003), highly discipline-specific academic vocabulary is rarely overlapped with general-service words; the occurrence frequency of the particular registers in one field is not maintained in other fields (Nation 2001). Since each discipline is characterized with its academic specialist varieties (Gee 2004), domain-specific academic lexis needs to be sought and taught in a systematic way. Recent studies, therefore, have shifted its focus toward the discipline-specific registers.

As textbooks, materials, and references increase in difficulty and complexity through the stages of learning, academic discourse at the college level no doubt requires students to possess an adequate amount of English lexis. As previous studies suggest different indexes ranging from 4,000 to 9,000 word families corresponding to the text coverages of the minimum 95% to the maximum 99% for the vocabulary size attained by second language learners at higher levels, the target lexical threshold is an important pedagogical issue to EAP instructors and researchers at the university level. Though numerous studies on L1 reading have reported vocabulary size as one of the strongest predictors of text comprehension (Wright and Cervetti 2017), research on second language reading has not given its proper attention to *what words* and *how much* the students in the college level should know in order for them to understand the text in a given discipline. Students of linguistics majors, in particular, are expected to acquire a degree of breadth and depth in vocabulary knowledge at their disposal, since their curriculum often implements English mediated instruction with compulsory readings written in highly specialized academic English. The present study, therefore, investigates the vocabulary demands of Korean EFL students of linguistics and provides useful insights into the way of assisting them in meeting their academic challenges and

requirements in the curriculum. Specifically, this research was guided by the following two research questions:

- 1) What size of vocabulary is required for EFL students to be able to adequately read their primary textbooks in linguistics courses?
- 2) What kinds of words constitute a Linguistics Academic Vocabulary List (LAVL) that suffices to read a linguistic textbook with 95% lexical coverage?

2. Theoretical Background

Previous research findings generally agree that there is the minimum vocabulary load for an *adequate* level of comprehension of authentic texts. According to Laufer (1989), 95% lexical coverage is required to grasp the *acceptable* comprehension of academic texts when aided by guessing unknown words in the context. Nation (2006) further claimed that readers should attain the lexical coverage of 98% for *successful* text comprehension. Later, Laufer and Ravenhorst-Kalovski (2010) proposed two lexical thresholds: 95% coverage of running words for a minimal *acceptable* degree of comprehension with 4,000 to 5,000 word families and 98% coverage for an *optimal* level of comprehension with 8,000 word families.

The vocabulary load can also be estimated in terms of the frequency at every 1,000 word-family, which was based on the general vocabulary list such as Michael West's (1953) General Service List (GSL). According to Nation and Waring (1997), the basic 2,000 word families in West's GSL cover up to 85% of text in written academic English with a variation of lexical coverage from 78% to 92% across different genres. Following Nation (2006), 95% vocabulary coverage corresponds to the size of 4,000 word families, while 98% vocabulary coverage amounts to 8,000 to 9,000 word families which cover a lower range of vocabulary.

Other studies attempted to identify academic vocabulary in university settings. By combining the word lists based on textbooks, lectures and examinations in previous studies, Xue and Nation (1984) presented a University Word List (UWL), which contains a list of 808 high frequent non-GSL academic words, which account for approximately 8% of the words in typical academic texts (Nation 1990). While the UWL serves as a good index for academic vocabulary, the selection principles in the study were questioned because of the limited range in topics, genres and text types. By

building a more balanced data across 28 subject areas under 4 disciplines including humanities, science, commerce, and law, Coxhead (2000) compiled the Academic Word List (AWL), which consists of 570 word families beyond the 2,000 levels of West's GSL with the coverage of 10% of general academic texts. Since the AWL was created by considering frequency, range, and specialized occurrence, a number of subsequent studies have used Coxhead's AWL as another indicator of lexical coverage of an English text.

Although it is quite useful to identify the core academic vocabulary across disciplines, questions have been made regarding the viability of a vocabulary list common across different areas, since many high-frequency words vary their meanings from discipline to discipline (Hyland and Tse 2007). In this line of research, the list and amount of discipline-specific academic vocabulary have been pursued to meet the needs for students in a given area under the framework of EAP. Research articles, therefore, have been a main source of analyses in the majority of studies on academic vocabulary, since they are structured following the Introduction-Method-Results-Discussion (IMRD, Swales 1990) format with specialist registers varying depending upon areas. Studies have proposed word lists for specific academic disciplines: agricultural studies (Martinez, Beck and Panza 2009), business (Hsu 2011, Konstantakis 2007), chemistry (Valipouri and Nassaji 2013), engineering (Hsu 2014, Mudraya 2006, Ward 2009), environment (Liu and Han 2015), 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 (applied) linguistics (Khani and Tazik 2013, Moini and Islamizadeh 2016, Vongpumivitch, Huang and Chang 2009). Whereas earlier studies examined the proportion of the domain-specific high frequency words against those in the GSL and AWL (e.g., Martinez et al. 2009, Wang et al. 2008), more recent studies (e.g., Hsu 2014) have taken the layered method by counting up the word list beyond the first two groups of 1,000 words in the BNC/COCA frequency bands to meet the vocabulary threshold often set at 95% text comprehension. When the target students were in the graduate level (Hsu 2011) or in medical science (Hsu 2013), the list was retrieved by selecting words beyond BNC 3,000 words. Given the discipline-specific nature of the text and the varied sizes and genres of corpora used across studies, the discrepancy in the amount of word lists in previous studies is quite large, which ranges from 92 to 1,260. So is the disparity in lexical coverages across word lists: The greatest lexical coverage was 19.81% in average estimated by two corpora in medical English with 819 word families by Lei and Liu (2016), while the shortest was 224 word families excluding GSL and AWL terms with the coverage of 5.07% of research articles in

linguistics by Moini and Islamizadeh (2016). Table 1 is a summary of studies on the vocabulary list with the number of discipline-specific words in each corpus specified.

Table 1. Studies on the Discipline-Specific Vocabulary List

Domain	Research	Text type	Corpus size ^a	Discipline-specific words	
				No. of words	Coverage
Agriculture	Martinez et al. (2009)	RAs	0.83	92	NA
Business	Konstantakis (2007)	TBs	0.60	560	2.79
	Hsu (2011)	TBs	7.62	426	5.66
Chemistry	Valipouri and Nassaji (2013)	RAs	4.00	390	7.23
	Mudraya (2006)	TBs	1.99	1,260	NA
Engineering	Ward (2009)	TBs	0.27	299	16.40
	Hsu (2014)	TBs	4.57	729	14.30
Environment	Liu and Han (2015)	RAs	0.86	458	15.43
Food	Esfandiari and Moein (2015)	RAs	4.65	1,090	13.00
	Vongpumivitch et al. (2009)	RAs	1.55	603	14.07
Linguistics	Khani and Tazik (2013)	RAs	1.55	773	12.48
	Moini and Islamizadeh (2016)	RAs	4.00	224	5.07
	Wang et al. (2008)	RAs	1.09	623	12.24
Medical Science	Hsu (2013)	TBs	15.02	595	10.72
	Lei and Liu (2016)	RAs	2.79	819	19.44
		TBs	3.50		20.18
Nursing	Yang (2015)	RAs	1.01	676	13.64
Social Studies	Kwary and Artha (2017)	RAs	1.04	350	12.65

Note. RAs = Research articles; TBs = Textbooks; NA = Not available

^a The total running words per million

Given the relatively longer tradition of research on English for Science and Technology (EST), earlier studies were undertaken by researchers in engineering science. Ward (1999) identified a list of 2,000 highly frequent word families, including both general and specialized vocabulary for students to adequately read a basic engineering text. His subsequent work (2009) analyzed engineering-specific 299 word types called the Basic Engineering List (BEL) that accounted for 16.40% of 271,000 running words targeted for the students with low English proficiency. Mudraya (2006), by developing a much bigger corpus of 2 million words out of compulsory engineering textbooks for Thailand university students, retrieved a rather comprehensive list of 1,260 word families as the minimum vocabulary load. Nevertheless, her study did not further analyze the most frequent vocabulary beyond the basic levels. More recent search of the *adequate* vocabulary demands of engineering English textbooks was undertaken by Hsu (2014), who created a corpus of 100 engineering textbooks comprising 20 sub-areas to retrieve an Engineering English Word List (EEWL) for Taiwanese EFL students. While the

vocabulary thresholds for engineering texts vary across subject areas, Hsu produced the most frequent 5,000 word families that cover 95% of engineering textbooks. In addition, she analysed 729 most-frequently-occurring word families beyond the first 2,000 words that accounted for 14.3% of the total words in the corpus.

Similarly productive is research on medical vocabulary given the difficulty due to the overwhelming amount of medical terms. Among others, Wang et al. (2008) derived a Medical Academic Word List (MAWL) by compiling a corpus of 1.09 million words, identifying 623 word families beyond the GSL that corresponded to 12.24% of the medical academic texts. Likewise, Hsu (2013) developed a list of 575 word families, which covered 10.72% of textbooks in medical science. More recent study by Lei and Liu (2016) compared the MAWL with a new medical academic vocabulary list (MAVL), which were derived from two large corpora, 2.7 million and 3.5 million words respectively from journal articles and textbooks in medical science.

More relevant to the present study are word lists in linguistics and applied linguistics based on research articles. Vongpumivitch et al. (2009) identified a total of 603 word families in the Applied Linguistics Research Articles Corpus (ALC), which is composed of 1,554,032 words of 200 research articles published in five major journals: *Applied Linguistics*, *Language Learning*, *the Modern Language Journal*, *Second Language Research*, and *TESOL Quarterly*. Comparative frequency analyses revealed 475 word forms are overlapped with the AWL, which accounted for 11.2% of the ALC, with the remaining 128 non-AWL word families covering only 2.8% of the ALC. Later, by constructing a more representative data consisting of 1,553,450 running words in 240 research articles out of twelve journals, Khani and Tazik (2013) reported the most frequent 773 word types which amounted to 12.48% of the articles in applied linguistics. Whereas Vongpumivitch et al. only estimated the proportion of the AWL versus non-AWL words in ALC, Khani and Tazik provided the coverages compared against the GSL and AWL words: The first and second 1,000 band of GSL 2,000 words accounted for 76.4% of the entire corpus, whereas the AWL words corresponded to 11.96%. More recently, Moini and Islamizadeh (2016) produced the Linguistics Academic Word List (LAWL) by analyzing a corpus of 4 million words from 700 empirical studies in eleven linguistics journals. Among 1,263 word families in the list, 658 word families in the GSL represented 72.48%, while 381 word families in the AWL matched 10.18%, leaving additional 224 word families that covered 5.07% of the linguistic research articles. Although these studies displayed a fairly consistent coverage of approximately 10% by the AWL, they did not provide the vocabulary size pertaining to linguistics, nor did they

focus on the creation of the linguistics-specific word list. What is more, the linguistics-specific word lists in previous studies were aimed toward the comprehension of research articles, which are not readily accessible by undergraduate students. Therefore a new study is called for to investigate the vocabulary loads of compulsory linguistics textbooks in the foundation course for undergraduates.

3. Methodology

The present study took the layered method where the discipline-specific word list is retrieved by setting a general word list as a threshold point to select the specialist vocabulary as well as the vocabulary load. Specifically, the study derived the linguistics-specific vocabulary list by excluding the general high-frequency 2,000 words in the BNC/COCA bands, while the AntWordProfiler program (Anthony 2014) was utilized since it is more flexible in modulating the level lists.

3.1. Compilation of LTC and Development of LAVL

With reference to the foundation courses in linguistics offered at the university level in Korea, a corpus of English linguistics textbooks was developed out of five major textbooks in general linguistics that have been used as primary or secondary references in courses for undergraduate students. All the sampled linguistics textbooks were downloaded from the electronic databases. After the front matter, running heads, figures, references and indexes were removed, the whole text was included. Finally, the Linguistics Textbooks Corpus (LTC) was composed of a total of 1,141,830 running words with each sub-corpus consisting of an approximate number of 227,646 words. Table 2 displays the composition of LTC.

Table 2. Composition of LTC

Textbooks	Types	Tokens	Families	TTR
Book 1	16,293	245,898	4,342	6.63
Book 2	16,056	252,268	7,070	6.36
Book 3	8,651	107,391	2,894	8.06
Book 4	14,533	203,203	2,346	7.15
Book 5	15,371	333,070	6,429	4.61
Total	70,904	1,141,830	23,081	6.21

Note. Book 1 = *Introduction to Language* (10th ed.); Book 2 = *How English Works: A linguistic Introduction* (3rd ed.); Book 3 = *Introducing English Linguistics*; Book 4 = *Linguistics for Everyone: An Introduction* (2nd ed.); Book 5 = *Language Files: Materials for an Introduction to Language and Linguistics* (12th ed.).

The LAVL is then derived from the LTC by means of the methodology modified from Hsu (2014). The input data were processed to guarantee the optimal relevance of the lexis to the research goals, so that the words are both “common enough” and “worthwhile to learn” (Hsu 2014: 60).

3.2 The Program

The study employed the AntWordProfiler program (Anthony 2014) to retrieve the list of the word families of the LTC by uploading the twenty-five 1,000-word-family lists with four additional lists (i.e., proper nouns, interjections, transparent compounds, and abbreviations),¹ which were downloaded from the Range program (Nation 2012). Compared to the fourteen BNC word lists of the strong British nature of written English, the twenty-five BNC/COCA word-family lists include more common and recent words in the first 2,000 word families, therefore more suitable for the current study for EFL students. Since the software provides indexes for vocabulary demands measured by the increase at every 1,000 word family on the ranked BNC/COCA word lists, the minimal vocabulary required for the *adequate* comprehension of linguistics textbooks was estimated by sorting frequencies of word families ranked in an descending order of the BNC/COCA scale. Specifically, the study calculated the lexical coverage of each 1,000 word family from the first 1,000 band and advancing toward the latter 1,000 bands until the cumulative coverage reaches 95% lexical threshold point.

¹In the Range program, proper nouns, interjections, compounds and abbreviations are respectively contained in the categories of Baseword31, Baseword32, Baseword33 and Baseword34, which are dubbed ‘additional lists’ in this study.

3.3 Data Analysis

The compiled corpus, the LTC, was run on the AntWordProfiler program. The analysis returned the base words which belong to each 1,000-word-family list alongside the statistics about the frequency of each base word list, the coverage percentage of its frequency, the cumulative coverage percentage of its frequency, and the frequency of its word families, as exemplified in Table 3 below. The program also provided the words which do not belong to any of the twenty-five word-family lists and the additional lists (i.e., proper nouns, interjections, compounds, and abbreviations) as the category of 'Not in the lists.' The words contained in this category may unnecessarily inflate the number of words students ought to acquire. In recognition of this problem, the words in the 'Not in the lists' category were manually checked and listed in relevant Baseword lists downloaded from the Range program: proper nouns in the Basewrd31, interjections in the Basewrd32, transparent compounds in the Basewrd33 and abbreviations in the Basewrd34. In addition, words pertaining to linguistics were put into the newly-created category of 'Glossary'. Some of the words classified as belonging to the Glossary category are presented in (1) with relevant contexts such as (2) and (3).

- (1) *topicalization, hyponymic, dissimilation, auxiliaries, schwa, markedness, obstruents, constative, retronymy, uvular, obstruents*
- (2) A third feature is the placement of noun phrases that typically occur in the object position of a sentence, but are also the topic of that sentence, at the beginning of the sentence, a process known as *topicalization*. (Book 5)
- (3) **Markedness** is a related concept involving binary oppositions, highlighting how we tend to consider one member of a pair of antonyms typical and the other marked. (Book 4)

These words need to be taught and learned via explicit instructions because they carry highly discipline-specific technical meanings. For this reason, a group of words initially classified as 'Not in the lists' were regrouped in the Glossary category.

Finally, the revised base word lists and the corpus were run on the AntWordprofiler program to answer Research Question 1, which is discussed in Section 4.1. For Research Question 2, it was needed to calculate the coverage percentage of the first 2,000 base

words in addition to proper nouns, transparent compounds, interjections, abbreviations and the glossary.² That percentage was then deduced from the targeted 95%, the remainder of which would be the percentage that the LAVL is expected to cover.

As for the criteria for selecting words, we adopted the method used in Hsu (2014) with some modifications, as follows:

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 3 out of 5 textbooks.
3. *Frequency*: Members of a word family occur at least 17 times across textbooks in the LTC.

As is widely accepted, the first 2,000 word families serve as a spring board for high school graduates. Hence, the words outside these lists were targeted to meet the need of developing the LAVL. The range of 3 was decided to guarantee that students exposed to different textbooks learn vocabulary relevant to their needs. The frequency of 17 was selected after several experiments to ensure that the lexical coverage of 95% was fulfilled.

4. Results and Discussions

4.1 The Vocabulary Load of Linguistics Textbooks

The results of the analysis demonstrate that knowing the most frequent 7,000 words plus proper nouns, interjections, abbreviations, transparent compounds and glossary would command 95% lexical coverage of a linguistics textbook to ensure adequate reading comprehension. This would be an answer for Research Question 1. Table 3 displays the lexical coverage of linguistics textbooks at each level of the BNC/COCA base word lists.

To repeat, students should learn at least 7,000 base words plus its families to arrive at the desired 95% lexical coverage of a linguistics textbook. However, acquiring as

²Following the tradition in the literature (e.g., Hsu 2014, Nation 2006), the lexical coverage of these additional lists were added to that of the base word lists on the amount of vocabulary necessary for adequate comprehension. For example, proper nouns require a minimum learning load because learners can easily identify the referents of proper nouns.

many as 7,000 word families must be an extremely challenging task for most EFL students. Each of the sixth to twenty-fifth 1,000 word families provides a very small increase in lexical coverage. Hence, this study strives to identify a vocabulary list which is of utmost relevance and usefulness to EFL students of linguistics. Students would better benefit from focusing on vocabulary items pertaining to linguistics textbooks rather than being exposed to words appearing frequently in general. To find a solution for this issue, we decided to address Research Question 2.

Table 3. BNC/COCA Base Word Lists in Linguistics Textbooks

BNC base word lists	Tokens (running words)	% Coverage in tokens	Cumulative % coverage in tokens	Tokens of word families
Proper noun	30,715	2.69	2.69	—
Compound	2,169	0.19	2.88	—
Interjection	2,055	0.18	3.06	—
Abbreviation	2,855	0.25	3.30	—
Glossary	571	0.05	3.35	—
1 st 1000	796,655	69.77	73.12	995
2 nd 1000	123,546	10.82	83.95	975
3 rd 1000	66,683	5.84	89.79	955
4 th 1000	20,096	1.76	91.55	819
5 th 1000	13,131	1.15	92.70	675
6 th 1000	9,135	0.80	93.50	566
7th 1000	11,304	0.99	94.49	511
8 th 1000	9,934	0.87	95.36	395
9 th 1000	2,969	0.26	95.62	318
10 th 1000	2,398	0.21	95.83	254
11 th 1000	2,458	0.21	96.04	226
12 th 1000	1,484	0.13	96.17	194
13 th 1000	1,370	0.12	96.29	146
14 th 1000	2,626	0.23	96.52	133
15 th 1000	1,028	0.09	96.61	107
16 th 1000	685	0.06	96.67	96
17 th 1000	913	0.08	96.75	83
18 th 1000	909	0.08	96.83	53
19 th 1000	457	0.04	96.87	58
20 th 1000	228	0.02	96.89	41
21 st 1000	240	0.02	96.92	39
22 nd 1000	571	0.05	96.96	41
23 rd 1000	343	0.03	96.99	27
24 th 1000	114	0.01	97.00	28
25 th 1000	211	0.02	97.02	22
Not in the lists	33,977	2.98	100.00	
Total	1,141,830	100.00		

4.2 Linguistics Academic Vocabulary List (LAVL)

For Research Question 2, we identified the most frequent words appearing between the third and twenty-fifth 1,000-word-family lists. Table 3 above demonstrates that the cumulative coverage percentage of the first and second word lists plus proper nouns, interjections, transparent compounds, abbreviations, and glossary is 83.95%. Hence establishing the LAVL involves the remaining 11.05 lexical coverage (95% – 83.95% = 11.05%). Using EXCEL, a total of 607 words were finally selected. The last included word *inhibit* with its family members occurred 17 times across 3 different textbooks. Table 4 provides the information on the number and level of the selected base words.

Table 4. LAVL 607 Word-families across the BNC/COCA

BNC/COCA	Number of word families	BNC/COCA	Number of word families	BNC/COCA	Number of word families
3 rd	322	10 th	6	17 th	2
4 th	103	11 th	11	18 th	3
5 th	51	12 th	3	19 th	3
6 th	30	13 th	5	20 th	1
7 th	22	14 th	9	21 th	2
8 th	17	15 th	4	22 th	2
9 th	5	16 th	4	23 th	2

Some of the words from the LAVL classified as belonging to the 3rd and 4th BNC/COCA levels are presented in (4) and (5), respectively.

(4) 3rd BNC/COCA level linguistics words

indirect; passive; synthesis; request; genre; declarative; literary; antonym; lax; disorder; cognate; imperative; entail; encounter; emphasis; informal; network; facial; interrogative; prototype; quantity; prose; survey; prosody; coda; relevance; genitive; transform; flap; paradigm

(5) 4th BNC/COCA level linguistics words

morpheme; discourse; corpus; complement; contrast; pidgin; relevant; nasal; creole; alveolar; code; oral; slang; fricative; subordinate; input; multiple; palate; proposition; dialogue; ambiguous; temporal; pragmatic; diphthong; boundary; global; bilabial; hierarchy; explicit; masculine; glide

Words in (3) and (4) appear to be closely related to topics and issues in linguistics textbooks. Noticeable is that most of them are sub-technical à la Hsu (2014), which is again modified from Chung and Nation (2003, 2004). This will be discussed in Section 4.3 in more detail.

In the meantime, these word lists identically point to the role of the AWL words in linguistics. Among the finalized 607 linguistics word families, a total of 208 word families were found to overlap with the AWL (see Appendix). These overlapping items proved to show 34% (in types) lexical coverage. This result not only confirms that the AWL is still meaningful in linguistics but also verifies the claims of previous research (e.g., Hsu 2014, Martinez et al. 2009, Wang et al. 2008) that the AWL can be instrumental in constructing vocabulary lists serving students' disciplinary needs.

While direct comparisons are not possible since the word lists in the previous studies are based on research articles in applied linguistics, the LAVL possesses a relatively smaller coverage of the entire corpus. The 607 word families of the LAVL in the current study yields 11.05% coverage of linguistics textbooks, whereas Vongpumivitch et al.'s (2009) list of 603 words corresponds to 14.07%, Khani and Tazik's (2013) 773 words account for 12.48% of the entire applied linguistics research articles. Taken together, the two lists amount to an average of 13.28% of texts of research articles in applied linguistics. It is partially because of the relatively smaller size of the corpus in the present study, but mainly because of the more specialized nature of research articles that are typically made up of academic lexis of greater density in previous studies.

4.3 The Nature of LAVL

In the literature of academic vocabulary, the distinction between technical and general vocabulary has been suggested (e.g., Baker 1998, Chung and Nation 2003, 2004, Hsu 2014, Yang 1986). According to Chung and Nation (2004: 252), who proposed a four-point scale to measure the technicality of vocabulary, "technical vocabulary is subject related, occurs in a specialist domain, and is part of a system of subject knowledge", while general vocabulary is defined as a word with minimal or little particular association with a specific subject. However, the division between the two is not easy to make (e.g., Hsu 2014, Lei and Liu 2016 and references therein). On the other hand, an intermediate layer between the two has been proposed under the labels such as sub-technical (e.g., Baker 1988, Cowan 1974, Hsu 2014, Yang 1986), or semi-technical (e.g., Farrell 1990). In this study, the term *sub-technical vocabulary*

will be adopted, which refers to “those words that have one or more “general” English meanings and which in technical contexts take on extended meaning” (Trimble 1985: 129). Accordingly, sub-technical vocabulary inevitably entails either polysemous or homonymous nature.³

As in other academic word lists, a wide range of sub-technical terms turned out to feature the LAVL, which are illustrated by lexemes such as *liquid* (Level 3) and *glide* (Level 4). *Liquids* and *glides* in linguistics refer to certain types of consonants by an analogy of the ways the sounds are articulated. For example, in a linguistics textbook, the definitions of liquids and glides are provided as in (6) and (7).

(6) **Liquids [l] [r]**: In the production of the sounds [l] and [r], there is some obstruction of the airstream in the mouth, but not enough to cause any real constriction or friction. (Book 1)

(7) **Glides [j] [w]**: The sounds [j] and [w], the initial sounds of you [ju] and we [wi], are produced with little obstruction of the airstream. . . After articulating [j] or [w], *the tongue glides quickly into* place for pronouncing the next vowel, hence the term glide. (Book 1)

Among the vocabulary suggested in (4) and (5) above, sub-technical terms can be easily spotted including *passive*, *transform*, *flap*, and *complement*, which all carry discipline-specific senses. One striking example is *coca*, which is placed as a BNC/COCA Level 4 word, probably with two unrelated meanings: one as the soft-drink company/product and the other as the Corpus of Contemporary American English. Thus the word *coca* itself without context can trigger an ambiguity. The excerpts (8) and (9) demonstrate the uses of the word *coca* in the same textbook.

(8) McDonaldization, or **coca** colonization (a humorous derivation from the beverage, cited in 1960), is often linked with globalization, particularly by those who feel more anxious or negative about the phenomenon. (Book 2)

³Even though it needs to be acknowledged that the distinction between polysemy and homonymy is not straightforward to draw, the consensus has emerged that at least two criteria can be applied to demarcate the two:

- i) the lexicographer’s knowledge of the historical derivation of words, and
- ii) unrelatedness vs. relatedness of meaning (Lyons 1977: 550–551)

- (9) Setting an entirely new bar in terms of scale is the recent Corpus of Contemporary American English (COCA), which is over 400 million words of spoken and written American English. (Book 2)

Such sub-technical terms need to be taught through explicit instructions, preferably with the definitions and the contexts in which they occur, because they possess meanings non-inferable from their literal senses.

The appearance of everyday common nouns can also be a characterizing property of the LAVL, which includes such words as *cot* (Level 7), *spoon* (Level 4) and *couch* (Level 4). It is definitely because language and thus linguistics are fundamentally and tightly interrelated with human being's mundane life (Linguistics Society of America, 2019). Neither being technical nor being sub-technical, they are mainly employed for introducing language data to explain specific topics. In (10), the second speaker's repetition of the word *spoon* is presented in order to demonstrate how a grown-up speaker teaches a young learner. In (11), on the other hand, the lexical items *couch* and *sofa* are utilized to explicate one of the pivotal topics in linguistics, synonymy.

- (10) Conversation between a father and a child

Child: *Want other one **spoon**, Daddy.*

Father: *You mean, you want the other **spoon*** (Book 5)

- (11) It has been said that there are no perfect synonyms. . . Still, the following two sentences have very similar meanings:

*He's sitting on the **sofa**. / He's sitting on the **couch**.* (Book 1)

In addition, the LAVL includes an archaic form *thou* (Level 4) and its family members *thy* and *thee*, which is hardly witnessed in other academic word lists. The examples in (12) and (13) present the relevant excerpts from the LTC.

- (12) As Blake (1992: 536) notes, although usage was inconsistent, during the Middle English period, **thou** and **thee** were used to denote "intimacy or contempt," while **ye** and you "were neutral and polite." (Book 1)⁴

⁴ Although the word *ye* is another archaic form, it is not included in the LAVL due to its lack of range and frequency.

(13) Early Modern ENGLISH (text 1611)

Our father which art in heaven, hallowed be **thy** Name. **Thy** kingdome come.
(Book 5)

From (12) and (13), the use of such archaic English pronouns can be easily understood, as they are indispensable to the discussions in historical linguistics.

Of equal interest is that the LAVL includes conjunct adverbials such as *nevertheless*, *moreover*, *likewise* and *hereby*. Existing academic lists such as EEWL (Hsu 2014) and MAVL (Lei and Liu 2016) also involve such adverbials: the former contains *moreover*, *nevertheless*, *furthermore*, *thereby* and *despite*, while the latter has *thereby* and *thereafter*. Conjunct adverbials allow the writer to connect informational contents in a specific respect and, at the same time, to assess how he/she views the connection between linguistic units (Quirk et al. 1985: 632). However, one adverb in the LAVL, which is *hereby*, serves a rather different metalinguistic purpose than connecting propositional contents or modifying the action depicted by the verb/sentence. (14), (15) and (16) demonstrate the contexts where the adverb *here* is utilized in the LTC.

(14) We take the word **hereby** and insert it before the potentially performative verb:

(15) I **hereby** promise I will help you with your project this week.

(16) #John **hereby** promises he will help you with your project this week. (Book 5)

The frequent occurrences of *hereby* in linguistics textbooks are attributable to a linguistic phenomenon called the performative speech act, in which the particular event described by the verb is achieved in the performance of the speech act itself. One of the tests to determine whether a sentence contains a performative verb is to insert *hereby* before the main verb, as explained in (14). The concordance checks of all of the occurrences of *hereby* in LCT reveal that the word *hereby* is used for explicating the topic of performative verbs rather than for relating neighboring ideas or modifying the described event.

In passing, another noticeable characteristic of the LAVL is the presence of a number of exclamations and spoken interjections, even though they do not appear between Level 3 and Level 25 along the BNC/COCA word frequency scale. Academic word lists usually exclude exclamations or spoken interjections, the reason for which is commented by Hsu (2014: 59) that “the engineering texts were written, formal in nature.” Nevertheless, in

the LAVL, exclamations or interjections serve a couple of linguistics-specific functions: for elaborating on certain linguistic phenomena or for presenting language data. In the excerpt (17), the interrelatedness between language and the social factors such as gender and class is discussed. The excerpt (18) displays how exclamations are utilized in presenting language data.

(17) She [Eliza Doolittle in *My Fair Lady*] says words like **'ow** and **garn**, both of them supposedly subliterate exclamations. Higgins claims that such speech determines more about her social status than does either her shabby clothing or poor hygiene. (Book 2) (Brackets added)

(18) Here is a patient with Broca's aphasia trying to explain how he came to the hospital for dental surgery:

Yes ... **ah** ... Monday ... **er** ... Dad and Peter H ... (his own name), and Dad ... **er** ... hospital ... and **ah** ... Wednesday ... Wednesday, nine o'clock ... and **oh** ... Thursday ... ten o'clock, **ah** doctors ... two ... an' doctors ... (Book 4)

To sum up, the linguistics text is further characterized by a wide array of sub-technical terms, while everyday vocabulary is utilized as examples in the linguistics data. Likewise, the occurrence of archaic forms and the use of adverbials with metalinguistic references add unique features to the academic vocabulary of linguistics texts.

5. Conclusions and Implications

This study attempts to provide a way of assisting EFL university students in Korea in enriching their lexical competence needed to grasp primary textbooks. For the purpose, the study compiled a linguistics corpus (LTC) consisting of 1,141,830 running words and developed an academic vocabulary list (LAVL) for undergraduate students who study linguistics. The utilization of the AntWordProfiler uploaded with the 25,000 BNC/COCA word lists revealed that EFL college students should master the minimum vocabulary load of 7,000 word families plus proper nouns, interjections, abbreviations, glossary, and transparent compounds in order to reach 95% lexical coverage of the linguistics textbook. Beyond the first 2,000 word lists, the study further identified a total of 607

word families that accounted for 11.05% of the entire corpus. While the knowledge in the first two 1,000 high-frequency bands in BNC/COCA constitutes 80.59% of the tokens of the entire linguistics corpus, students of linguistics should build up the second language lexicon by adding 607 high frequency linguistics-specific words to reach 95% comprehension of their textbooks. Among the field-specific 607 word families, 208 words were overlapped with the AWL. While this finding was corroborative with the previous studies where the AWL contributes to building up academic literacy across disciplines, the results also demonstrated that additional set of technical words ought to be acquired to meet the vocabulary threshold of linguistics study.

The findings suggest important pedagogical implications. First, the vocabulary demands for EFL university students should be estimated and taught in a systematic way. For students in linguistics, for example, knowledge of the most frequent 7,000 word families is not a realistic goal to achieve in a short period of time. While their entry level vocabulary size may vary, the English textbooks in the secondary level are developed on the basis of the 3,000 Korea Basic English Word list (Shin 2014). In other words, several sessions of instruction may be in need to help students build up the specialist academic lexis as part of the curriculum. Meanwhile, the students should be familiarized with the genre-specific features of technical words to better understand the text for their academic study. They may be encouraged to search textual examples in a well-developed corpus with each genre properly labeled (Jung and Wharton 2012). Second, the study brings materials writers' attention to the useful information regarding the minimum amount of lexis and the comprehensive range of key vocabulary for the students' successful academic performance. As academic corpora and their applications may highlight authentic features of language use within specific disciplines, the corpus analyses may contribute to students' understanding of the differences between specialist vocabulary and general English (Hong and Jhang 2010). It also provides specific directions to ESP instructors regarding *how much* and *what level* of vocabulary in addition to how to design tests and develop materials with a more useful list of words and expressions. Consequently, EFL students could better benefit from focusing on words of imminent need and relevance. By doing so, the study demonstrates the usability of corpus analysis in ELT methodologies, assessment, and materials development with practical implications regarding the importance of corpora in language teaching and learning.

The future directions of research may include the following. First, given the relatively small size of the corpus in the current study, the precision of the vocabulary load in

linguistics can be made with more replications by incorporating a wider range of textbooks. Next, the word list and amount pertaining to each domain of linguistics could be explored. Given the diversity of areas in linguistics and applied linguistics and the depth of the content in each area, the word list of the present study may not be sufficient for students who have to read more sophisticated textbooks in advanced courses in linguistics. In addition, the most frequent lexical bundles in linguistics textbooks should be identified, as the domain-specific formulaic expressions and word list are complementary as basic building blocks for advanced academic discourse. Since there are a range of differences and preferences in the use of lexical bundles even among the English native speakers' writing in literature, linguistics and education (Chung and Song 2012), lexical combinations within and across disciplines should be identified. Finally, as an extended line of research on the list of words and lexical bundles in reading, the lexical load of the students' writing at the tertiary level should also be investigated to ensure the balanced development of both the receptive and productive vocabulary in academic discourse.

References

- Anthony, L. 2014. AntWordProfiler (Version 1.4.1) [Computer Software]. Tokyo, Japan: Waseda University. Available from <http://www.laurenceanthony.net/software>
- Baker, M. 1988. Sub-technical vocabulary and the ESP teacher: An analysis of some rhetorical items in medical journal articles. *Reading in a Foreign Language* 4, 91–105.
- Chung, S.-W. and K. Song. 2012. A corpus-based analysis of lexical bundles in academic articles on English literature, linguistics, and education written by English native speakers and Koreans. *Korean Journal of English Language and Linguistics* 12(4), 947–975.
- 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.
- Cowan, J. R. 1974. Lexical and syntactic research for the design of EFL reading materials. *TESOL Quarterly* 8, 389–399.
- 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.
- Farrell, P. 1990. *Vocabulary in ESP: A Lexical Analysis of the English of Electronics and a Study of Semi-Technical Vocabulary*. Dublin: Centre for Language and Communication Studies.
- Gee, P. 2004. *Situated Language and Learning: A Critique of Traditional Schooling*. London: Routledge.
- Hong, S. C. and S. E. Jhang. 2010. The compilation of a Maritime English Corpus for ESP learners. *Korean Journal of English Language and Linguistics* 10(4), 963–985.
- 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.
- Jung, C. K. and S. Wharton. 2012. Finding textual examples of genres: Issues for corpus users. *Korean Journal of English Language and Linguistics* 12(1), 129–148.
- 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.
- 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.
- 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 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.
- 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.

- Linguistics Society of America. 2018. *Linguistics Society of America*. Retrieved from <https://www.linguisticsociety.org/content/linguistics-everyday-life>
- 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.
- Lyons, J. 1977. *Semantics*. Cambridge: Cambridge University Press.
- Martin, A. V. 1976. Teaching academic vocabulary to foreign graduate students. *TESOL Quarterly* 10, 91–97.
- 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.
- 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 York: 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. *The BNC/COCA Word Family Lists*. Retrieved from https://www.victoria.ac.nz/_data/assets/pdf_file/0004/1689349/Information-on-the-BNC_COCA-word-family-lists-20180705.pdf
- 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.
- Quirk, R., S. Greenbaum, G. Leech and J. Svartvik. 1985. *A Comprehensive Grammar of the English Language*. London: Longman.
- 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.
- Shin, D. 2014. What English vocabulary are we learning? *Journal of Foreign Studies* 30, 64–95.
- Swales, M. M. 1990. *Genre Analysis: English in Academic and Research Settings*. Cambridge: Cambridge University Press.
- Trimble, L. 1985. *English for Science and Technology: A Discourse Approach*.

Cambridge: Cambridge University Press.

- 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.
- Vongpumivitch, 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. 1999. How large a vocabulary do EAP engineering students need? *Reading in a Foreign Language* 12(2), 309–323.
- 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, H. 1986. A new technique for identifying scientific/technical terms and describing science texts. *Literary and Linguistic Computing* 1, 93–103.
- 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

Hyeon-Okh Kim
Professor, Department of English
Ajou University
206 Worldcup-ro, Youngtong-gu,
Suwon 16499, Korea
Email: hokim67@ajou.ac.kr

Hyeon-Okh Kim·Hye-Kyung Lee

*The academic vocabulary list in linguistics
for EFL university students*

Hye-Kyung Lee
Professor, Department of English
Ajou University
206 Worldcup-ro, Youngtong-gu,
Suwon 16499, Korea
Email: hkleee@ajou.ac.kr

Received: January 30, 2019

Revised: March 10, 2019

Accepted: March 19, 2019

Appendix

The Linguistics Academic Vocabulary List

Note: The LAVL includes the 607 word families beyond the BNC/COCA 2K words. A number is given after each word to mark the level in the BNC/COCA base word list. Among them, 208 items are identified to belong to the AWL, which are in bold.

abbreviate_7	absence_3	abuse_3	academy_3	accommodate_3
accompany_3	accomplish_3	accurate_3	accusative_22	achieve_3
acquire_3	activate_4	adjust_3	adverb_10	advocate_3
affirm_3	affix_9	alike_4	allophone_18	alongside_3
alternate_3	alveolar_14	ambiguity_5	ambiguous_5	analogous_6
analogy_5	analyse_3	anatomy_5	ancient_3	angle_3
animate_4	antonym_17	apology_3	appropriate_3	approve_3
approximate_3	apt_5	archaic_8	architecture_3	articulate_4
artificial_4	aspire_4	assert_3	assess_3	asterisk_11
attribute_3	audience_3	author_3	authority_3	axe_4
babble_8	ban_3	bantu_19	baseball_4	behave_3
belief_3	bi_11	bias_3	bicycle_4	bilabial_22
bilingual_7	binary_16	blend_3	bold_4	border_3
boundary_3	broadcast_3	budget_3	bundle_4	callosum_18
candidate_3	capacity_3	cartoon_4	categorical_7	category_3
cell_4	chart_3	chin_4	cite_3	clarify_3
classify_4	clause_3	cleft_9	clerk_4	click_4
cluster_3	coca_11	coda_12	code_3	cognate_13
cognition_8	coherent_4	coin_3	colleague_3	colloquial_11
comparative_4	compatible_4	competence_4	complement_4	comprehend_4
compress_4	comprise_3	con_4	conceive_3	concept_3
conclude_3	concrete_3	conduct_3	confine_3	confirm_3
conflict_3	conform_4	connotation_10	conquest_6	consequence_3
constituency_4	construct_3	consult_3	contraction_6	contrast_3
controversy_3	convenient_4	convention_3	conversion_4	coordinate_3
copula_21	cord_4	core_3	corpus_6	correlate_3
corrupt_3	cot_7	couch_4	counterpart_4	creole_7
critic_3	criticise_3	criticism_3	crucial_3	crystal_3
cue_5	cute_6	data_3	database_3	deaf_4
debate_3	declarative_13	declare_3	decline_3	decode_7
deficit_3	define_3	deictic_16	delete_5	deliberate_3
demonstrate_3	denote_6	dental_6	description_3	determiner_16
device_3	devour_7	diachronic_14	dialect_6	dialogue_3
differ_3	differentiate_4	digital_3	dimension_3	diminutive_9
diphthong_15	dis_9	disagree_3	discourse_3	discriminate_3
disorder_3	display_3	dispute_3	distant_3	distinguish_3
distribute_3	division_3	document_3	domain_4	domestic_3
dominance_5	dominate_3	dual_4	duration_4	dynamic_4

effective_3	element_3	elevate_3	eloquent_6	emerge_3
emphasis_3	emphasise_3	emphatic_6	enable_3	encode_6
encounter_3	enforce_3	enhance_3	ensure_3	entail_5
entity_4	essay_3	essence_4	essential_3	estimate_3
etc_3	ethnic_3	evaluate_3	eve_4	evident_3
evolution_3	exceed_3	exception_3	excerpt_7	exclude_3
explicit_3	exploit_3	explore_3	explosion_3	extent_3
extinct_5	extraordinary_3	facial_5	facilitate_3	faculty_3
federal_3	filter_3	finite_6	flap_4	flexible_3
fluent_6	focus_3	forensic_6	formal_3	formation_3
formula_3	frequent_3	fricative_19	function_3	fundamental_3
gap_3	garbage_6	genetic_3	genie_11	genitive_19
genre_4	geography_3	gesture_3	giraffe_10	glide_4
global_3	globe_5	glottal_17	gradual_3	habitual_7
hawk_5	heel_3	heroby_10	hierarchy_4	highlight_3
hint_3	hive_7	homophone_14	horizontal_4	host_3
humour_3	hypothesis_3	hypothetical_6	icon_5	id_5
identical_4	idiom_7	imitate_4	immigrant_4	impact_3
imperative_5	implicit_4	incorporate_3	independence_3	independent_3
index_3	indirect_4	indo_23	inevitable_3	infinite_4
inflect_8	informal_4	informative_6	inhibit_4	initial_3
initiate_3	input_3	inquire_3	insert_3	insight_3
instinct_3	intelligence_3	intelligible_5	interact_3	interchange_6
intermediate_5	internal_3	international_3	interrogative_14	intimate_3
intuition_6	intuitive_6	inventory_5	ipa_18	jargon_7
journal_3	jutes_23	kitty_11	knight_5	label_3
lateral_5	lateralise_15	lax_8	layer_3	lecture_3
legislate_3	legitimate_3	lexical_8	lexicography_13	lexicon_8
likeness_3	likewise_4	linear_4	linguist_8	linguistic_4
link_3	liquid_3	literacy_4	literary_3	literature_3
lobe_8	logic_3	lyric_4	majority_3	mammal_4
mandarin_8	manipulate_3	manuscript_4	masculine_5	mechanism_3
media_3	medium_3	ment_14	mere_3	metaphor_4
mid_4	millennium_5	mimic_5	minimal_4	misconception_7
mitigate_5	modify_3	monolingual_15	monster_3	moral_3
moreover_3	morpheme_14	mosquito_5	motion_3	motive_3
multiple_3	myth_3	nasal_7	negate_7	negative_3
network_3	neuter_11	neutral_3	nevertheless_3	nim_21
nominal_5	notion_3	noun_7	null_7	numerous_3
oblige_3	obstruct_4	obtain_3	occupation_3	ocean_3
offence_3	omission_6	omit_4	onomatopoeic_16	oral_3
organ_3	oriented_3	origin_3	originate_4	orthography_11
overlap_4	overview_5	overwhelm_3	ox_8	pa_5
palate_8	palm_3	paradigm_5	parallel_3	parameter_4
parrot_6	parse_9	participant_3	participate_3	participle_14
passage_3	passive_4	patients_3	patron_4	pea_5
peculiar_4	perceive_3	perception_3	periphery_4	permissible_8

permit_3	persist_3	perspective_3	persuasion_5	philosophy_3
phonetic_9	phrase_3	pidgin_15	pit_3	plausible_5
plural_4	poke_4	posit_8	pragmatic_5	praise_3
precede_3	precise_3	predicate_7	predominant_4	prefix_11
preposition_12	prescribe_4	prescription_4	presence_3	preserve_3
prestige_4	primary_3	principal_3	principle_3	prior_3
proceed_3	professor_3	proficient_7	prominence_6	promote_3
prompt_3	pronoun_8	proportion_3	proposition_4	prose_5
prosody_14	proto_20	prototype_5	proximity_5	publication_3
publish_3	puzzle_3	qualitative_5	quantity_5	radical_3
random_3	rap_4	recognition_3	recursive_11	reflect_3
reflex_6	regulate_3	reinforce_3	reject_3	relative_3
relevance_4	relevant_3	renaissance_5	request_3	resemble_3
resolve_3	resonance_6	resource_3	respective_3	restore_3
restrict_3	resume_3	retain_3	reverse_3	revise_3
revive_3	rhyme_6	rhythm_3	ridge_4	rival_3
rural_3	sample_3	scandinavia_7	scholar_3	scholarship_4
scope_3	scrabble_9	semantic_5	sensitive_3	sensory_6
sequence_3	severe_3	shortly_3	sibling_5	sickle_9
significant_3	simultaneous_3	sincere_4	slang_8	sociolinguistic_13
sofa_5	software_3	sole_3	solution_3	solve_3
source_3	spatial_4	spit_4	spontaneous_4	spoon_4
static_5	status_3	stigmatise_10	stipulate_6	straightforward_4
strategy_3	strive_5	subjunctive_14	subordinate_4	subsequent_3
substantial_3	substitute_3	substrate_7	succeed_3	sufficient_3
suffix_11	superior_3	survey_3	swahili_11	syllable_6
sympathy_3	synchronic_16	syndrome_4	syntax_8	synthesis_4
talent_3	task_3	technical_3	temporal_5	terminology_6
territory_3	theory_3	thou_4	tick_4	tissue_3
tract_5	transcript_4	transform_3	transition_3	transmit_3
trauma_4	tremendous_3	tribe_3	turtle_5	un_5
underlie_3	underline_5	uniform_3	unique_3	universe_3
utter_4	vague_3	variety_3	vast_3	verb_5
vernacular_8	versa_13	versus_3	vervet_19	via_3
vibrate_4	vice_3	virtue_3	visible_3	vision_3
vocal_4	vocative_19	volume_3	vowel_7	whisper_3
witch_5	yiddish_10			