# Constructing a Vocabulary List of English Textbooks in Semantics and Pragmatics*

Hye-Kyung Lee (Ajou University)

Hyeon-Okh Kim (Ajou University)

Lee, Hye-Kyung and Hyeon-Okh Kim. 2020. Constructing a vocabulary list of English textbooks in semantics and pragmatics. *Korean Journal of English Language and Linguistics* 20, 384−410. This study aims to develop a vocabulary list of semantics/pragmatics and to complement an existing academic vocabulary list in general linguistics. For that purpose, a semantics/pragmatics corpus (SPTC) consisting of approximately 1,400,000 words was compiled. By combining both a quantitative analysis (i.e., AntWordProfiler) and a qualitative manual examination, a semantics/pragmatics vocabulary list was constructed of 409 word families out of the corpus. The compiled list was then compared with an existing introductory linguistics word list to identify lexical patterns that were common with general linguistics and unique to semantics/pragmatics. The results of the analysis showed that a set of (sub-)technical terms characterized the semantics/pragmatics texts, which suggests the need for a specialist vocabulary list for students in semantics and pragmatics. The result also showed that the semantics and pragmatics texts feature a more frequent use of specific conjunct adverbs including *hence*, which proved to be instrumental in presenting key terms and topics and organizing the flow of arguments. This might have been because semantics/pragmatics texts are richer with technical terms than are introductory linguistics texts. Pedagogical implications and directions for future research are also suggested.

**Keywords**: semantics/pragmatics vocabulary list, technical terms, academic vocabulary, linguistics vocabulary list

## 1. Introduction

Academic vocabulary refers to the words that are frequently encountered in a wide range of academic texts but are less frequently used in other genres (Coxhead and Nation 2001). For that reason, notwithstanding its pivotal nature in learning and teaching, academic vocabulary has proven to be challenging to students in specific fields, especially to novice or entry-level students who are required to study English-medium textbooks in an EFL setting. In recognition of this need, an increasing body of research has been conducted so far to establish academic vocabulary lists in either general academic (e.g., Coxhead 2000, Xue and Nation 1984) or discipline-specific (e.g., Hsu 2011, 2014, Lei and Liu 2015) texts.

In the field of linguistics, multiple researchers have attempted to construct vocabulary lists to facilitate students or non-native professionals' acquisition of adequate academic lexis based on either research articles (Khani and Tazik 2013, Moini and Islamizadeh 2016, Vongpuvititch, Huang and Chang 2009) or major linguistics textbooks (Kim and Lee 2019). However, the vocabulary type and load pertinent to each sub-discipline of linguistics can vary given the diversity of sub-areas in (applied) linguistics and the contents in each area (Kim and Lee 2019: 46). Hence, the existing word lists in general linguistics may not be sufficient for learners in linguistics who need to study textbooks in specialized or advanced linguistics courses. This paper aims to develop an English academic vocabulary list for semantics and pragmatics textbooks and to compare it with the vocabulary list in introductory linguistics textbooks. The rationale for selecting semantics and pragmatics among various sub-disciplines lies in the fact that these two subjects mainly deal with meaning-bearing linguistics units such as words or phrases. In that regard, the two contrast with other sub-disciplines such as phonetics, morphology, and syntax in which segmental parts, discrete sub-units of words, or tree structures are the primary focus. This paper also seeks practical and pedagogical implications for content instructors or material developers. The following three questions guide this study.

1) What size of vocabulary is required for students to be able to adequately comprehend their textbooks in semantics/pragmatics?
2) What kinds of words constitute a semantics/pragmatics academic vocabulary list that suffices to understand a semantics/pragmatics textbook with 95% lexical coverage?

3) What features distinguish the vocabulary in general linguistics from that in semantics/pragmatics?

## 2. Literature Review

### 2.1 Studies on Academic Vocabulary

Literature on academic vocabulary has suggested minimum lexis loads for appropriate comprehension of academic texts. Here, lexical coverage refers to "the percentage of running words in the text known by the reader (Nation 2006: 61). For example, Laufer (1989) proposes 95% lexical coverage for reasonable comprehension of authentic academic texts, while Nation (2001, 2006) puts forward 98% to gain successful comprehension of the texts. Based on further research, Laufer and Ravenhort-Kalovski (2010) bifurcate lexical coverage points: 95% for *minimal* acceptable comprehension and 98% for *optimal* comprehension of texts. Briefly, the coverage thresholds vary in different studies, and this research has adopted 95% as a computing point, because this appears to be a feasible goal for students reading introductory semantics/pragmatics textbooks.

With an aim to facilitate teaching and learning of academic lexis, researchers have to date developed a number of academic vocabulary lists. Two of the pioneering and representative lists are the University Word List (UWL) developed by Xue and Nation (1984) of 836 word families and the Academic Word List (AWL) compiled by Coxhead (2000) of 570 word families. Researchers have adopted the AWL as a reference list in assessing the proportions of discipline-specific high-frequency words in various academic disciplines and have found that the AWL covers approximately 10% of academic texts across a range of disciplines: Wang, Liang and Ge (2008) in medical studies, Matinez, Beck and Panza (2009) in agriculture, and Vongpumivitch et al. (2009) in linguistics.

Although general academic lists are useful in identifying the core academic lexis, many high-frequency words can convey different senses from discipline to discipline (Hyland and Tse 2007). To address this issue and thus to meet the needs of students in specific domains, investigators have attempted to construct domain-particular vocabulary lists, such as for business (Hsu 2011, Konstantakis 2007), chemistry (Valipouri and Nassaji 2013), engineering (Hsu 2014, Mudraya 2006), the environment

(Liu and Han 2015), (applied) linguistics (Khani and Tazik 2013, Kim and Lee 2019, Moini and Islamizadeh 2016), medical science (Lei and Liu 2016, Wang et al. 2008), nursing (Yang 2015), and social studies (Kwary and Artha 2017).

While earlier studies mainly focused on the proportions of high-frequency words against those in the GSL and AWL, later studies adopted a layered approach by counting up the word-families beyond the first and second 1,000 word-family lists in the British National Corpus and Corpus of Contemporary American English (BNC/COCA) bands (e.g., Hsu 2011, 2014, Kim and Lee 2019). The reference word-family lists used in these studies were developed by Nation (2006, 2012) and his colleagues, whose research utilized the BNC and the COCA and ranked the word-families into fourteen and twenty-five 1000 word-family lists. For example, Kim and Lee (2019) constructed a linguistics academic vocabulary list of 607 word families beyond the first two groups of 1000 word-family bands for students in linguistics to reach 95% comprehension of linguistics textbooks. However, several studies noted problems involved in excluding general high-frequency words such as GLS words or BNC/COCA first and second 1000 most-frequent words when constructing academic vocabulary lists (e.g., Cobb 2010, Gardner and Davis 2014, Lei and Liu 2016, Neufeld, Hancioglu and Eldridge 2011). Researchers pointed out that some AWL words were among the COCA/BNC high-frequency words, challenging the learning-order assumptions from high-frequency general words to low-frequency technical words. Researchers also found that some high-frequency words such as *interest* and *rate* can carry domain-specific meanings in specific academic contexts (Lei and Liu 2016: 43). In a preliminary analysis, this paper found that linguistics texts do not contain many high-frequency words that are ambiguous between general and technical meaning, and as such, this study adopted the first type of approach in which high-frequency words (i.e., the first and second 1000 word families) are excluded.

## 2.2 Studies on Linguistics Vocabulary Lists

Previous word lists in linguistics and applied linguistics have been compiled predominantly out of research articles. For example, Vongpumivitch et al. (2009) constructed a vocabulary list of 603 word families from 200 research articles published in applied linguistics academic journals, identifying that 11.2 % of the words in their list overlap with the AWL. In more recent studies, Khani and Tazik (2013) and Moini

and Islamizadeh (2016) compiled vocabulary lists of 773 words types and 1,263 word-families respectively based on research papers in the field of (applied) linguistics. Though pioneering and fruitful, these previous studies mainly targeted vocabulary adopted in advanced and professional research articles, which are not readily accessible to students in linguistics classes. Noticing this gap, Kim and Lee (2019) investigated the vocabulary loads of linguistics textbooks in foundation courses for undergraduates in linguistics.

To help enhance students' lexical literacy in linguistics, Kim and Lee (2019) developed the linguistics academic vocabulary list (LAVL) out of the Linguistics Textbooks Corpus (LTC), which consists of major introductory linguistics textbooks. Using a layered approach discussed in section 2.1 above, Kim and Lee (2019) reported a list of 607 word families beyond the first 2,000 word families, which corresponded to 11.05% of the entire corpus. By doing so, they suggested a means of helping EFL undergraduates in linguistics to enrich their lexical competence and provided useful directions to content instructors and materials writers.

Linguistics has a range of subfields including, but not limited to, phonetics, phonology, morphology, syntax, and semantics/pragmatics. Because these subfields relate to different aspects of language and linguistic knowledge, the vocabulary types and loads must vary from area to area. Given this, the linguistics vocabulary list needs to be complemented by those required in each linguistics sub-discipline. This study pursues the goal of compiling a word list in the area of semantics/pragmatics, comparing that vocabulary with that of general linguistics.

# 3. Data and Methodology

Adopting a layered method, this study retrieved a discipline-specific word list by setting an established general word list as a threshold. Specifically, this study constructed a semantics/pragmatics-specific vocabulary list by utilizing the AntWordProfiler program (Anthony 2014) and excluding the general high-frequency 2000 words in the twenty-five BNC/COCA 1000-word-family lists.

## 3.1 Construction of the Corpus and the Vocabulary List

A corpus was constructed that comprised primary or secondary semantics/pragmatics

textbooks.[1] All selected books were downloaded from electronic databases and converted into 10 text files. The data were then manually revised by removing the front matter, running heads, figures, and back matter (e.g., references and indices). The finalized corpus was called the Semantics/Pragmatics Textbooks Corpus (SPTC henceforth) and consisted of a total of 1,369,916 running words. Table 1 presents the composition of the SPTC.

### Table 1. Composition of SPTC

| Textbooks | Types | Tokens | TTR | STTR |
|---|---|---|---|---|
| Book 1 | 9,236 | 142,296 | 6.40 | 36.20 |
| Book 2 | 5,539 | 61,403 | 9.02 | 38.10 |
| Book 3 | 5,314 | 49,903 | 10.46 | 40.67 |
| Book 4 | 14,267 | 321,210 | 4.33 | 38.31 |
| Book 5 | 12,424 | 191,781 | 6.44 | 39.09 |
| Book 6 | 5,672 | 78,563 | 7.22 | 38.03 |
| Book 7 | 5,924 | 86,295 | 7.01 | 38.17 |
| Book 8 | 6,589 | 115,560 | 5.59 | 32.99 |
| Book 9 | 7,940 | 148,036 | 5.39 | 36.40 |
| Book 10 | 10905 | 174,869 | 6.19 | 38.23 |
|  | 83,310 | 1,369,916 | 6.12 | 37.58 |

As Table 1 displays, the textbooks vary considerably in their lexical density based on the type/token ratio (TTR), which ranged from 5.39 to 10.46. As a higher TTR indicates greater lexical density, Table 1 suggests that Book 3 had the most diverse words and Book 8 the least. The TTR value is also sensitive to the length of a text, and thus for the present study, the average TTR was calculated for each textbook based on every 1,000 words, i.e., the standardized type/token ratio (STTR). Because the STTRs converged with an average of 37.58, the lexical diversity gap between textbooks was not noteworthy.

Next, the semantics/pragmatics vocabulary list (SPVL henceforth) was developed out of the SPTC. The retrieved data were then refined to ensure that the lexis was not only "common enough" but also "worthwhile to learn" for students in semantics/pragmatics classes (Hsu 2014: 60).

---

[1] To ensure the disciplinary balance, four textbooks from each sub-discipline were chosen in addition to two textbooks that cover both sub-disciplines.

## 3.2 The Procedure

As in Kim and Lee (2019), this study utilized the AntWordProfiler program (Anthony 2014) to retrieve the list of words (types, tokens and groups). AntWordProfiler offers indexes for vocabulary loads calculated by the increase at every 1,000-word family on the ranked BNC/COCA word lists. The compiled corpus, the SPTC, was run on the program, with twenty-five 1,000-word-family lists and four additional lists (i.e., proper nouns, interjections, transparent compounds, and abbreviations) uploaded as reference word lists; all these lists were donwloaded from the Range program (Nation 2012)[2]. The twenty-five BNC/COCA word-family lists provide word families based on frequency and range data. In addition, proper nouns, interjections, compounds and abbreviations are respectively assigned to the categories Baseword31, Baseword32, Baseword33, and Baseword34, which are dubbed "additional lists" in the current study.

The results of running AntWordProfiler provided various kinds of information. First, the words in the SPTC were sorted to each list according to the uploaded reference word lists. Some statistical information was also given including the frequency of the whole words (types, tokens and groups) in each level as well as the coverage percentage and cumulative percentage of the frequency, as partially exemplified in Table 2. The program simultaneously identified the words that were not assigned to any of the uploaded lists as 'not in the lists'. These unlisted words were then manually examined and sorted into relevant Baseword lists: proper nouns in Basewrd31, interjections in Basewrd32, transparent compounds in Basewrd33, and abbreviations in Basewrd34. In this process, it was noted that a considerable number of the unlisted words highly pertained to semantics/pragmatics. Hence a new category "Glossary" was added to accommodate these words (as Basewrd35). The characteristics and nature of the words regrouped in the Glossary category will be discussed in Section 4 below. A number of the unlisted words that were errors in the original texts or foreign/archaic words were removed from the list.

The SPTC and the revised base word lists were run again on the AntWordProfiler program, the output of which is presented in Table 2 below. Addressing Research

---

[2] The two terms *groups* (in AntWordProfiler) and *families* (in Range) are used interchangeably to refer to inflectional and derivational forms of a base form. For example, the three types *compensate*, *compensating*, and *compensatory* belong to the same group or family. For detailed information about the word lists in the Range program, refer to Nation (2012).

Question 1 required measuring the coverage percentages of the first 2,000 word-families alongside those of the words in the four additional categories and the Glossary. In measuring the vocabulary load for adequate comprehension, lexical coverages of the additional categories have legitimately been added before those of the first 2000 words in the relevant literature (Hsu 2014, Kim and Lee 2019, Nation 2006). This is mainly because understanding the senses or referents of these words poses a minimal learning load for learners. The cumulative percentage of running words up to the first 2000 words was then deducted from the target 95% lexical coverage, which would result in the SPLV.

When selecting target words, several criteria have been adopted in the literature. For this study, the following criteria were utilized, which is based on Hsu (2014) and Kim and Lee (2019).

(i) Specialized occurrence: The word families are beyond the first 2,000 word families

(ii) Range: Members of a word family appear across at least 6 out of 10 textbooks.

(iii) Frequency: Members of a word family occur at least 66 times across textbooks in the SPTC.

To guarantee the subject relatedness of the SPLV, the words beyond the first 2000 words were targeted. This approach has been adopted in the literature because the first 2,000 word families are known to serve as a spring board for high school graduates. As for the decision of the range, the range of 6 (60%) was chosen to ensure the compatibility between the current study and the reference study (Kim and Lee 2019). Analyzing five different linguistics textbooks, Kim and Lee (2019) selected the range of 3, which covers 60% of the used textbooks. The range is usually selected to make sure that students who use different textbooks can acquire the vocabulary relevant to their learning needs. The frequency of 66 was chosen after repeated experiments to fulfill the targeted lexical coverage of 95%. The detailed discussion of this procedure will be presented in Section 4 below.

# 4. Results and Discussion

## 4.1 The SPVL

Table 2 presents the results from AntWordProfiler with the descriptive statistics including frequency and lexical coverage of the SPTC at each level of the BNC/COCA word lists.

### Table 2. BNC/COCA Base Word List in the Semantics/Pragmatics Textbooks

| LEVEL | TOKENS (running words) | TOKENS % | CUMULATIVE TOKENS % | WORD FAMILIES |
|---|---|---|---|---|
| Proper noun | 33223 | 2.46 | 2.46 | 2685 |
| Compound | 21323 | 1.58 | 4.03 | 29 |
| Interjection | 1682 | 0.12 | 4.16 | 372 |
| Abbreviation | 4177 | 0.31 | 4.47 | 274 |
| Glossary | 8322 | 0.62 | 5.08 | 153 |
| 1st | 945235 | 69.92 | 75.00 | 999 |
| 2nd | 130411 | 9.65 | 84.65 | 990 |
| 3rd | 99008 | 7.32 | 91.97 | 960 |
| 4th | 30147 | 2.23 | 94.20 | 871 |
| **5th** | **20534** | **1.52** | **95.72** | **723** |
| 6th | 10068 | 0.74 | 96.46 | 645 |
| 7th | 9332 | 0.69 | 97.15 | 514 |
| 8th | 7014 | 0.52 | 97.67 | 442 |
| 9th | 2035 | 0.15 | 97.82 | 348 |
| 10th | 2779 | 0.21 | 98.03 | 305 |
| 11th | 1485 | 0.11 | 98.14 | 254 |
| 12th | 1523 | 0.11 | 98.25 | 205 |
| 13th | 1693 | 0.13 | 98.38 | 174 |
| 14th | 1707 | 0.13 | 98.50 | 127 |
| 15th | 589 | 0.04 | 98.55 | 119 |
| 16th | 1740 | 0.13 | 98.67 | 104 |
| 17th | 1654 | 0.12 | 98.80 | 81 |
| 18th | 1024 | 0.08 | 98.87 | 73 |
| 19th | 256 | 0.02 | 98.89 | 61 |
| 20th | 339 | 0.03 | 98.92 | 36 |
| 21th | 224 | 0.02 | 98.93 | 41 |
| 22th | 396 | 0.03 | 98.96 | 33 |
| 23th | 246 | 0.02 | 98.98 | 32 |
| 24th | 95 | 0.01 | 98.99 | 33 |
| 25th | 167 | 0.01 | 99.00 | 30 |
| Not in the lists | 13516 | 1.00 | 100.00 | 5833 |
| TOTAL | 1351944 | 100.00 | | |

Regarding Research Question 1, the analysis revealed that to reach the targeted 95% lexical coverage of a semantics/pragmatics textbook, students need to acquire words up to the 5th level (5,000 word families) plus proper nouns, interjections, abbreviations, transparent compounds and the Glossary. However, as acknowledged by Hsu (2015) and Kim and Lee (2019), learning 5,000 word families on top of the first and second 1000 words must be a demanding challenge for students, in particular, for EFL learners. Furthermore, each of the fourth to twenty-fifth 1,000 words results in a very small increase in lexical coverage. To address these problems, this study adopted the methodology used by Hsu (2015) and Kim and Lee (2019) by means of identifying words that were of optimal relevance and usefulness to students in semantics/ pragmatics. This led to Research Question 2 stated above.

Table 2 shows that the cumulative coverage percentage up to the first and second thousand word lists equals 84.65%. Thus, constructing the SPVL required identifying the remaining 10.35% of lexical coverage (95% − 84.65% = 10.35%). Applying the criteria presented in section 3.2 above ((1) and (2)) and using Microsoft Excel, a total of 409 words were finally chosen. The last included word was *tautology*, which appear 66 times across 6 different textbooks. Table 3 presents frequencies and levels of the selected words. The SPVL is presented in the Appendix.

### Table 3. SPVL 409 Words across the BNC/COCA Lists

| BNC/COCA | Number of word families | BNC/COCA | Number of word families | BNC/COCA | Number of word families |
|---|---|---|---|---|---|
| 3rd | 232 | 10th | 7 | 17th | 5 |
| 4th | 53 | 11th | 2 | 18th | 1 |
| 5th | 31 | 12th | 4 | 19th | 0 |
| 6th | 22 | 13th | 5 | 20th | 2 |
| 7th | 13 | 14th | 6 | 21st | 1 |
| 8th | 14 | 15th | 2 | 22nd | 1 |
| 9th | 3 | 16th | 5 | | |

## 4.2 Comparison of the LAVL and the SPVL

### 4.2.1 The most frequent 30 words

As regards Research Question 3, the study compared the LAVL in Kim and Lee (2019) and the SPVL. The LAVL (607 words) shares 206 base words with the SPVL (409 words); these overlapping base words are presented in the Appendix.

Nevertheless, an examination of the top 30 words in both lists identifies discrepancies between the two lists. Table 4 shows the most repeatedly occurring words in each of the two lists with the range, the frequency, and the BNC/COCA level of each word. The eleven overlapping words are highlighted, which are *verb*, *linguistic*, *noun*, *category*, *lexical*, *function*, *utter*, *semantic*, *discourse*, *define*, and *description*. The underlined words in the SPVL are the ones included in the LAVL even though they do not appear in the LAVL lists in Table 4.

### Table 4. Top 30 Words in the LAVL and SPVL

| | LAVL | | | | SPVL | | |
|---|---|---|---|---|---|---|---|
| word | range | frequency | level | word | range | frequency | level |
| **verb** | 5 | 2725 | 5 | **semantic** | 10 | 4847 | 5 |
| **linguistic** | 5 | 2623 | 4 | **utter** | 9 | 3716 | 4 |
| **noun** | 5 | 2112 | 7 | context | 10 | 3124 | 3 |
| phrase | 5 | 1930 | 3 | **verb** | 10 | 2677 | 5 |
| vowel | 5 | 1901 | 7 | **linguistic** | 10 | 2421 | 4 |
| dialect | 5 | 1395 | 6 | theory | 10 | 2368 | 3 |
| clause | 5 | 1185 | 3 | pragmatic | 9 | 2329 | 5 |
| **category** | 5 | 1107 | 3 | proposition | 10 | 2221 | 4 |
| morpheme | 5 | 1080 | 14 | logic | 10 | 2019 | 3 |
| **lexical** | 4 | 1072 | 8 | concept | 10 | 1910 | 3 |
| **function** | 5 | 1044 | 3 | structure | 10 | 1895 | 3 |
| syllable | 5 | 950 | 6 | interpret | 10 | 1859 | 3 |
| plural | 5 | 959 | 4 | distinct | 10 | 1723 | 3 |
| variety | 5 | 853 | 3 | **define** | 7 | 1372 | 3 |
| **utter** | 5 | 813 | 4 | abandon | 6 | 1336 | 3 |
| phonetic | 5 | 799 | 9 | principle | 10 | 1328 | 3 |
| **semantic** | 5 | 738 | 5 | predicate | 10 | 1327 | 7 |
| theory | 5 | 675 | 3 | presuppose | 9 | 1322 | 6 |
| pronoun | 5 | 663 | 8 | infer | 10 | 1297 | 4 |
| suffix | 5 | 593 | 11 | **discourse** | 10 | 1275 | 3 |
| construct | 5 | 589 | 3 | **noun** | 10 | 1272 | 7 |
| linguist | 5 | 576 | 8 | analyse | 6 | 1248 | 3 |
| adverb | 5 | 562 | 3 | **phrase** | 10 | 1188 | 3 |
| relative | 5 | 556 | 3 | **category** | 10 | 1160 | 3 |
| **discourse** | 4 | 550 | 3 | **lexical** | 10 | 1144 | 8 |
| syntax | 5 | 546 | 8 | **function** | 10 | 1133 | 3 |
| **define** | 5 | 542 | 3 | **description** | 10 | 1091 | 3 |
| **description** | 5 | 483 | 3 | communicate | 10 | 1055 | 3 |
| acquire | 5 | 462 | 3 | aspect | 10 | 1022 | 3 |
| corpus | 4 | 454 | 6 | metaphor | 10 | 1014 | 4 |

Noticeable is that the SPVL in Table 4 contains many more terms pertaining to semantics and pragmatics than does the LAVL, which is predictable from the natures of the two corpora. With overlapping words excluded, the SPVL in Table 4 is characterized by words such as *context, interpret, presuppose, infer, communicate* and *metaphor*, which are highly relevant to the subfields of semantics and pragmatics as well as to general linguistics. The result of the comparison indicates that the LAVL alone does not suffice to meet the needs of students in semantics/pragmatics classes.

To further verify subject-relatedness of the SPVL, a keyword analysis was carried out using the KeyWords function of another word-processing program, the WordSmith (Smith 2016a). The KeyWords function allows for identifying key words in a given text by means of comparing the words in the target text with a reference set of words. Any word that is frequent in the target text is considered key (Scott 2016b: 6). The key words of the SPTC were calculated by comparing the words in the SPTC with those in the LTC. The result of the comparison showed that the top 12 keywords of the SPVL include *sense, truth, semantics, logical, interpretation, pragmatics, theory, relations, implicature, pragmatic, proposition*, and *utterance*.[3] This finding also supports the above-mentioned claim regarding the need for a vocabulary list that is more specialized in semantics/pragmatics. The list can supplement a general linguistics vocabulary list.

### 4.2.2 (Sub-)Technical terms

A terminological clarification is needed before this discussion continues. In the literature on academic words, several researchers have attempted to draw a distinction between technical and general words (e.g., Baker 1998, Chung and Nation 2003, 2004, Hsu 2014, Lei and Liu 2016, Yang 1986). For example, using a four-point scale to measure the technicality of words, Chung and Nation (2004: 252) explained that "technical vocabulary is subject related, occurs in a specialist domain, and is part of a system of subject knowledge," whereas general vocabulary refers to words without any particular association with a specific discipline. However, demarcating these two categories is not straightforward in the absence of clear criteria. Acknowledging this challenge, a group of researchers has suggested an intermediate category between the two under the label of sub-technical (e.g., Baker 1988, Cowan 1974, Hsu 2014, Yang

---

[3] Top 10 keywords refer to 12 words that have the highest keyness scores automatically calculated by the WordSmith program.

1986), or semi-technical (e.g., Farrel 1990). Following Kim and Lee (2019), this study will employ the term sub-technical to refer to "those words that have one or more "general" English meaning and which in technical context take on extended meaning" (Trimble 1985: 129).

For a similar reading comprehension effect, the LTC requires words up to the $7^{th}$ level (Kim and Lee 2019), whereas the SPVL needs words up to the $5^{th}$ level, as shown in Table 2 above. This difference is no doubt due to the much higher occurrences of compounds and glossary words in the SPTC. In the LTC, these two categories account for 0.19% and 0.05% of the whole occurrences respectively, whereas in the SPTC the portions of the same categories reach 1.68% and 0.62% respectively. This discrepancy means that in semantics/pragmatics textbooks, compounds and technical terms appear much more repeatedly than in introductory linguistics textbooks. Of these two categories, technical terms, in particular, draw our attention because students certainly need to acquire such technical terms pertaining to semantics/pragmatics for successful reading comprehension and content learning.

The (sub-)technical terms in the two lists differ in terms of not only the volume, but also types. (Sub)-technical terms can come from the BNC/COCA lists (from the $3^{rd}$ to the $25^{th}$) and the Glossary category. Kim and Lee (2019) identified such sub-technical terms as *liquid* and *glide* in the LAVL that are associated with the subfields of phonetics and phonology. These words qualify as sub-technical terms in that they refer to specific types of consonants in linguistics, whereas the same words are used as common words in other contexts.

In the SPVL, a set of sub-technical terms are identified such as *construction*, *failure*, *false*, *ostensive*, and *satisfactory* that solely occur in the SPVL. These words have general meanings, but in linguistics they are employed technically. Examples in (1) and (2) illustrate the technicality of these words.

(1) a. Such an ability is basic not only to language usage but also to model-building of all sorts, from map-making to the ***construction*** of theories (Book 7)
b. Much recent work on the meaning of conditional ***constructions*** builds on the similarities between conditionals and modals (Book 1)
(2) a. The direct objects of such verbs are referentially ***opaque***, meaning that substitution of a co-referential NP can affect the truth value of a sentence. (Book 1)

b. In a second game, O'Neill had younger children enlist parental help to retrieve stickers dropped into one of two identical ***opaque*** containers, out of reach. (Book 4)

The two quotes in (1) contain the word *construction*. The usage in (1a) delivers the general meaning of the word *construction* (i.e., *the way a thing is constructed*), whereas the one in (1b) has a discipline-specific meaning (i.e., *the arrangement of two or more forms in a grammatical unit*). In a similar vein, the two occurrences of the word *opaque* in (2) show two distinct meanings of the word: the technical meaning in (2a) (i.e., *substituting a coreferential NP affecting the truth value of a sentence*) and the general meaning in (2b) (i.e., *not transparent*). The extent of a word's technicality quite varies from word to word. The technical meaning of the first word *construction* is indeed registered in most dictionaries, presumably because that meaning is widely known to ordinary speakers. In contrast, the technicality of the word *opaque* is rarely dealt with in dictionaries, indicating that some technical terms should be explicitly presented and instructed to students in semantics/pragmatics courses.

The Glossary category is of special attention because a substantial number of (sub-)technical terms originate from that category. As discussed in Section 3.2, words belonging to this category are not listed on any of the BNC/COCA word lists. Among them, words judged to have strong technical flavor are grouped into this category. Table 5 presents a total of 20 words in the Glossary category that meet the range (i.e., 6 different textbooks) and frequency (i.e., more than 66 times). Words that also appear in the LAVL Glossary category are highlighted for comparison.

**Table 5. Words in the SPVL Glossary Category with Range 6 and Frequency Over 66**

| Word | Range | Frequency | Word | Range | Frequency |
|------|-------|-----------|------|-------|-----------|
| implicature | 10 | 1775 | lexicalized | 8 | 108 |
| performative | 7 | 246 | illocution | 6 | 102 |
| quantifier | 7 | 437 | compositionality | 8 | 99 |
| hyponym(y) | 7 | 431 | aspectual | 7 | 93 |
| NPs | 8 | 197 | perlocutionary | 8 | 92 |
| componential | 8 | 155 | stative | 6 | 81 |
| telic | 6 | 126 | locutionary | 7 | 76 |
| deontic | 6 | 119 | selectional | 9 | 71 |
| **perfective** | 8 | 115 | denotational | 6 | 47 |
| **performatives** | 6 | 115 | disambiguation | 6 | 69 |

A large number of the words in Table 5 are highly technical including *implicature, (il/per)locution, deontic*, and *hyponym*, because these words mainly convey semantics/pragmatics-related senses. The most frequent word is *implicature*, which indeed appears across all 10 textbooks. Because this word is highly subject-specific, it needs to be explicitly defined preferably in its initial occurrence, which is exemplified in (3).

> (3) Within the intentional-inferential framework initiated by Grice, one of the principal ways of thinking of the interrelations between meaning and context has been in terms of the notion of conversational ***implicature***. The theory of conversational ***implicature*** was developed by Grice in a famous series of lectures delivered in 1967. (Book 5)

In excerpt (4), the definition of the lexeme *illocution* is given. In a similar fashion, the technical sense of this term is clearly defined.

> (4) The ***ILLOCUTIONARY ACT*** (or simply the ***ILLOCUTION***) carried out by a speaker making an utterance is the act viewed in terms of the utterance's significance within a conventional system of social interaction. One way to think about the illocutionary act is that it reflects the intention of the speaker in making the utterance in the first place. (Book 8)

When the words in the Glossary categories in both lists (i.e., LAVL and SPVL) are compared, the lists differ in both volume and characteristics. Table 6 presents the words in the Glossary category of the LAVL and the overlapping words are highlighted.

### Table 6. Words in the LAVL Glossary Category

| Word | Range (out of 5) | Frequency | Word | Range (out of 5) | Frequency |
|---|---|---|---|---|---|
| **predicator** | 1 | 48 | dissimilation | 1 | 10 |
| infix | 3 | 26 | intonational | 3 | 9 |
| **ditransitive** | 3 | 24 | uvular | 2 | 9 |
| palatography | 1 | 22 | **topicalization** | 2 | 8 |
| suprasegmental | 2 | 21 | obstruents | 2 | 8 |
| protolanguage | 2 | 21 | **performatives** | 3 | 5 |
| **perfective** | 2 | 19 | glottalized | 1 | 4 |
| **morphosyntactic** | 2 | 16 | **markedness** | 1 | 3 |
| alphabetism | 1 | 15 | retronymy | 1 | 3 |
| schwa | 3 | 14 | decreolozation | 1 | 2 |
| articulator | 2 | 14 | sociolinguist | 1 | 2 |
| exophoric | 2 | 14 | infinitival | 1 | 2 |
| UG | 3 | 13 | interdentals | 1 | 1 |
| **constative** | 1 | 11 | unrounding | 1 | 1 |

In compiling the LAVL, the range and frequency thresholds were set at three different textbooks and over 17 times across 3 textbooks respectively. Under these thresholds, only two words in Table 6 would pass the criteria: *infix* and *ditransitive*. This presents a stark contrast with the large number of technical terms found in the SPVL. The comparison demonstrates that many technical terms in the semantics/pragmatics textbooks cannot be accommodated in existing word lists such as the BNC/COCA lists. This again suggests that the technical terms pertaining to semantics/pragmatics should be presented to students in an explicit and clear way for their adequate comprehension of their textbooks.

### 4.2.3 More features

Kim and Lee (2019) further demonstrated that linguistics textbooks are characterized by the use of everyday vocabulary, archaic forms and the conjunct adverbials. The SPVL shows both similarities and differences in these additional characteristics.

Among the everyday common words, Kim and Lee (2019) identified words such as *cot* (level 7), *spoon* (level 4), and *couch* (level 4) that were primarily used in language data to explicate pivotal topics. Similarly, the SPVL features common nouns and verbs, albeit different in kind and more in type. Common words in the SPVL include *mammal* (level 4), *bachelor* (level 6), *ward* (level 4), *bald* (level 5), and *sibling* (level 5). Of these words some deserve special attention because they are

prototypically used for suggesting language data to explicate specific topics in semantics/pragmatics. The excerpts in (5) and (6) present the classic example in which the word *bald* occurs:

(5) An account of reference has to deal with cases where there is no referent to fit the definite description, as in Bertrand Russell's famous example: The King of France is **bald**, or where the referent is not real, for example the man in the iron mask or the wizard of Oz. (Book 10)

(6) Use of the notion of presupposition is intended to explicate the pre-theoretical intuition that certain sentences or utterances take something for granted. Here is a well-known example: The King of France is **bald**. In the terminology of presupposition, this has been argued to presuppose that there is a king of France. (Book 6)

In both quotes, Russell's classic sentence, *The King of France is bald*, is mentioned to discuss presupposition. Indeed almost all occurrences of the word *bald* in the corpus are in this specific sentence in the context of presupposition, although other presupposition-related sentences could have been utilized. Presenting language needs to help learners in a specific discipline comprehend key topics. At the same time, it also should aid students in becoming familiar with and competent in subject-specific resources commonly circulated in that discipline (cf. Cortes 2004, Hyland 2008, Hsu 2015, Gilmore and Millar 2018 for discipline-specific lexical bundles). Students in semantics/pragmatics can be acquainted with recurrently circulated example sentences like *The King of France is bald* by virtue of being repeatedly exposed to such examples. Two other common nouns, *bachelor* and *sibling*, are chiefly used to explain topics such as the sense/reference distinction or symmetric sense relationship, as shown in excerpts (7) and (8).

(7) That '**bachelor**' should be semantically related in this way to 'married' is part of its sense; and it is part of the sense of 'married' that it should be related in a certain way to '**bachelor**'. (Book 9)

(8) For instance, if 'X is married to Y' is true so is 'Y is married to X' and vice versa, hence, "is married to" is a symmetric relation. Further examples of symmetric relations include: is sitting next to, is the same age as, is a **sibling** of. (Book 3)

These common words therefore turn out to serve metalinguistic rather than referential purposes.[4]

The LAVL includes a limited set of conjunct adverbs (i.e., *nevertheless, moreover, likewise*, and *hereby*), and the SPVL also contains a group of conjunct adverbs including *nevertheless, hereby, moreover, hence, unlike, furthermore,* and *despite*. Conjunct adverbs are defined as having the function of conjoining linguistic units, while delivering the speaker's assessment of how he/she perceives the connection between the units (Quirk et al. 1985: 632). Given that, conjunct adverbs must be highly useful for organizing the flow of information as well as for expressing writers' evaluation about the relationship between conjoined linguistic units. (cf. Blakemore 1987, 1989, Fraser 1990, 1999, Schiffrin 1987, for discourse connectives/markers). Among the conjunct adverbs in the SPVL, the first rank was held by the adverb *hence* (level 3), which is not included any previous academic lists. *Hence* is often viewed to be akin to adverbs such as *therefore* (level 2) and *thus* (level 2). However, a closer look at the data reveals subtle differences between the uses of these three adverbs. *Hence* is often followed by a noun phrase that conveys what can be reached via preceding arguments. Excerpts (9) and (10) exemplify this point:

(9) In the case of overlapping antonyms, the scales partially overlap, **hence their name**. For instance, in the case of good: bad, zero on the scale of badness corresponds to the point representing 'neither good nor bad' on the scale of goodness. (Book 3)

(10) The claim is that the implicated premise or premises of an utterance, together with the explicatures, logically warrant the implicated conclusions: **hence the use of the terms 'premise' and 'conclusion'**. (Book 6)

The NP in (9), *the name*, expresses what can be arrived at from the forgoing discussion. The NP introduced by *hence* in (10) also can be construed in a similar manner. That is, based on the explanation about implicated premises and implicated conclusions in the first sentence of (10), the author justifies the use of the terms "premise" and "conclusion". Nouns that occur in an NP after the adverb *hence* include *name, term, notion, category*, and *label*, which predominantly provide the use or

---

[4] The adjective metalinguistic "applies to situations in which language is used to comment on language. The distinction between the language in use, the 'object language', and the language used to describe it, the 'metalanguage', comes from modern formal logic" (Allott 2010: 120).

meaning of technical terms or notions. The frequent uses of *hence* in the semantics/pragmatics texts relate to the fact that these subjects deal with meaning-bearing linguistics units such as words or phrases as opposed to other sub-disciplines in which segmental units, discrete sub-parts of words, or tree structures are primarily focused. The content contained in such NPs must be highly instrumental in introducing topics worth teaching in a compact and effective way (Ni 2003). To validate this finding, a search of *hence*-based clusters was conducted in COCA and the search indeed revealed that the top five NP clusters following *hence* were *the name, the need, the term, the title*, and *the reason*.[5] As opposed to *hence*, the other two synonymous adverbs do not allow NPs to occur after them. Instead, the adverb *therefore* is strongly associated with negation, in that the most frequent four-word clusters in the SPVL are *therefore it is not, therefore there is no, therefore he didn't,* and *therefore he must not*. In contrast, the adverb *thus* repeatedly co-occurs with a full sentence or precedes the word *far*, forming the fixed phrase *thus far*.

Kim and Lee (2019) noted that one adverb *hereby* in the LAVL carries a metalinguistic rather than conceptual function because the adverb is exclusively used to explain the topic of performative sentences in which the situation conveyed by the sentence takes an effect by virtue of the sentence's being uttered. A similar tendency is observed in the SPVL; Entire occurrences of the adverb *hereby* involve either explanations of performative sentences or presentations of performative sentences, as exemplified in (11) and (12):

(11) Performatives can optionally be modified by the performative adverb ***hereby***; this adverb cannot be used with non-performative statements. (Book 1)
(12) Indicate whether the following sentences are odd or not odd.
　　(1) I ***hereby*** warn you that you will fail.
　　(2) They ***hereby*** warn her that she will fail.
　　(3) I ***hereby*** promised him that I would be at the station at three o'clock. (Book 8)

The adverb *hereby* is seldom employed in everyday spoken language. However, because it is disproportionally favored in linguistics texts, students in linguistics,

---

[5] The data were retrieved on May 29, 2020, from https://www.english-corpora.org/coca/.

especially those in semantics/ pragmatics, should learn the highly discipline-specific function of this adverb.

Some adverbs were also determined to co-occur quite recurrently, relaying a non-compositional meaning. One representative example is *strictly and literally*, which appear 7 times in the corpus.[6] For comparison, this cluster does not occur in the LAVL at all. Examples in (13) and (14) illustrate the uses of this cluster.

(13) Coupling that with the widely held assumption that sentence semantics is propositional, and so truth-conditional, leads to the endorsement of a principle along the following lines, where "what is said" is to be understood as the proposition ***strictly and literally*** expressed by an utterance. (Book 4)

(14) This distinguishes them from metaphors and exaggerations which are not lies even though they might be ***strictly and literally*** false and known to be so by the speaker. (Book 6)

The cluster *strictly and literally* was the most frequent three-word cluster involving *strictly*. The sense conveyed by this cluster does not result from the sum of meanings of the component adverbs. Rather the first adverb *strictly* must intensify the meaning of the second adverb *literally*, emphasizing the full extent of literalness. However, the same cluster does not make the *strictly*-based cluster list of the SPVL and nor of COCA, which suggests that this cluster is strongly associated with semantics/ pragmatics subfields. The cluster under discussion is frequently employed to explain the distinction between literal (i.e., proposition or what is said) and non-literal meaning (i.e, what is implicated or tropes including metaphor) meaning, which is deemed one of the most pivotal topics in semantics and pragmatics. In addition to this cluster, another frequent word combination was noticed (i.e., *conventionally associated* (11 times)). This cluster is also very useful in explicating the topic of conventional implicature, in which the meaning of *conventional* is used discipline specifically. The topic of semantics/pragmatics word clusters can be pursued elsewhere to supplement the SPVL for a fuller vocabulary inventory of semantics/pragmatics texts.

---

[6] The adverb *strictly* itself does not make the SPVL. Instead, as the BNC/COCA word list system allows, the base word *strict* with its family member occurs 191 times, among which the adverb *strictly* accounts for 119 times.

## 5. Conclusion

This study aimed to develop a vocabulary list for semantics/pragmatics as a way of helping university students in linguistics, particularly in semantics/pragmatics, enhance their lexical competence in these subfields of linguistics and to complement the existing academic vocabulary list in introductory general linguistics. For that aim, this study compiled a semantics/pragmatics corpus consisting of approximately 1,400,000 words out of 10 major semantics and pragmatics textbooks. Using the AntWordProfiler program (Smith 2016) uploaded with the BNC/COCA word lists and manual vetting adopted by Hsu (2015) and Kim and Lee (2019), this study identified a total of 409 word families (SPVL) of optimal relevance and immediate usefulness, which accounted for 10.35% of the entire corpus. Then the derived list was compared with the existing introductory linguistics word list (LAVL).

Among the 409 word families, 206 words overlapped with the LAVL, which showed that the LAVL also relates to the subfields of semantics and pragmatics to a great extent. Many semantics/pragmatics-specific words were found to be unaccommodated by the BNC/COCA words lists, which rendered a greater number of Glossary words in the SPVL than those in the LAVL. At the same time, it was revealed that additional and different words were needed for adequate comprehension of semantics and pragmatics texts. Specifically, comparing the two word lists showed that the SPVL contained a considerable number of semantics/pragmatics-specific technical terms that did not make the LAVL. In addition, the SPVL further featured the presence of certain conjunct adverbs including *hence* and field-specific word clusters such as *strictly and literally*, which demonstrated textual features semantics/pragmatics texts.

There are some pedagogical implications to this study's results. First, the vocabulary load of university students in a specific field can be more optimally evaluated in the way that is explored in this study by focusing on vocabulary of relevance and needs. That is, rather than merely calculating the types and size of words, which normally results in an unrealistically large number of words, this line of study can better identify words that are closer and thus more relevant to students' needs in a specific field. Second, this study can also provide useful information and guidelines for instructors and material writers in semantics/pragmatics when preparing teaching materials. This is because academic corpora developed out of common textbooks in a specific discipline can present authentic language uses that circulate frequently in that discipline. That way, students in a specific field can better benefit from concentrating

on words that merit learning rather than simply increasing their vocabulary size in a random way (Kim and Lee 2019).

Future studies can address the following issues. As discussed in Section 4.2.3, high－frequency lexical clusters in semantics/pragmatics texts can be identified and thus complement the vocabulary list for an ideal vocabulary inventory in this discipline, because individual words alone do not suffice to account for the lexical features entertained in a discipline. Second, the ways words are presented to students need to be investigated. Either the words alone can be taught in separate sessions prior to content classes or the words can be indirectly presented to students along with the contexts in which they are used.

# References

Allott, N. 2010. *Key Terms in Pragmatics*. New York: Continuum.

Anthony, L. 2014. *AntWordProfiler* (Version 1.4.1). [Computer Software]. Retrieved 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.

Blakemore, D. 1987. *Semantic Constrains in Relevance*. Oxford: Blackwell.

Blakemore, D. 1989. Denial and contrast: A relevance theoretic analysis of *but*. *Linguistics and Philosophy* 12, 15－38.

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

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

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

Cortes, V. 2004. Lexical bundles in published and student disciplinary writing: Examples from history and biology. *English for Specific Purposes* 23, 397－423.

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, 213－263.

Coxhead, A. and P. Nation. 2001. The specialized vocabulary of English for academic

purposes. In J. Flowerdew and M. Peacock, eds., *Research Perspectives on English for Academic Purposes*, 252–267. Cambridge: Cambridge University Press.

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.

Fraser, B. 1990. An approach to discourse markers. *Journal of Pragmatics* 14, 383–395.

Fraser, B. 1998. Contrastive discourse markers in English. In A. H. Jucker and Y. Ziv, eds., *Discourse Markers: Descriptions and Theory*, 301–326. Amsterdam: Benjamins.

Gardner, D. and M. Davis. 2014. A new academic vocabulary list. *Applied Linguistics* 35, 305–327.

Gilmore, A. and N. Millar. 2018. The language of civil engineering research articles: A corpus-based approach. *English for Specific Purposes* 18: 1–17.

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

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

Hsu, W. 2015. The most frequent formulaic sequences in college engineering textbook. *Corpus Linguistics Research* 1, 109–149.

Hyland, K. 2008. As you can see: Lexical bundles and disciplinary variations. *English for Specific Purposes* 27, 4–21.

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 S. 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. *English for Specific Purposes* 22, 42-53.

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

Matinez, I., S. C. Beck and C. B. Panza. 2009. Academic vocabulary in agriculture research articles: A corpus-based study. *English for Specific Purposes* 28, 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, 235-256.

Nation, P. 2001. *Learning Vocabulary in Another Language*. Cambridge: Cambridge University Press.

Nation, P. 2006. How large a vocabulary is needed for reading and listening? *The Canadian Modern Language Review* 63(1), 59-82.

Nation, 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

Neufeld, S., N. Hancioglu and J. Eldridge. 2011. Beware the range in RANGE and the academic in AWL. *System* 39, 533-538.

Ni, Y. 2003. Noun phrase in media text: A quantificational approach. In J. Aitchinson and D. M. Lewis, eds., *New Media Language,* 159-168. London: Routledge.

Quirk, R., S. Greenbaum, G. Leech and J. Svartvik, J. 1985. *A Comprehensive Grammar of the English Language*. London: Longman.

Schiffrin, D. 1987. *Discourse Markers*. Cambridge: Cambridge University Press.

Scott, M. 2016a. *WordSmith Tools 7*. Stroud: Lexical Analysis Software.

Scott, M. 2016b. *WordSmith Tools Manual*. Stroud: Lexical Analysis Software.

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. Hwang and R. Y. Chang. 2009. Frequency analysis of the words in the Academic Word List (AWL) and non-AWL content words in applied linguistics research papers. *English for Specific Purposes* 28, 33-41.

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

Xue, G. and P. Nation. 1984. A university word list. *Language Learning and Communication* 3, 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

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

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

# Appendix

## The Semantics and Pragmatics Vocabulary List

*Note*: The SPVL consists of 409 base words beyond the BNC/COCA 2K words. A number is given after each word to indicate the the frequency of each words. Among these words, 206 items also belong to the LAVL, which are presented in bold.

| | | | | |
|---|---|---|---|---|
| abandon_1336 | **absence_91** | abstract_387 | **accommodate_74** | accomplish_134 |
| **accurate_91** | **achieve_246** | **acquire_211** | adequate_78 | adjective_426 |
| adopt_198 | **adverb_310** | aim_139 | allege_68 | alternative_268 |
| **ambiguity_437** | **ambiguous_323** | **analyse_1248** | analytic_107 | anaphora_333 |
| **animate_106** | anomaly_131 | antecedent_189 | **antonym_342** | **apology_75** |
| applicable_82 | **appropriate_475** | **approximate_68** | arbitrary_143 | arise_274 |
| aspect_1022 | **assert_679** | assign_308 | assumption_532 | asymmetry_66 |
| **attribute_258** | **audience_97** | **author_243** | bachelor_115 | bald_132 |
| **behave_448** | **belief_301** | **binary_98** | bound_94 | **boundary_169** |
| calculus_66 | **candidate_76** | canonical_77 | capture_173 | **cagetory_1160** |
| cation_69 | causal_122 | characteristic_254 | **cite_110** | **clause_852** |
| **cognition_128** | cognitive_673 | coherence_180 | collocation_92 | com_241 |
| communicate_1055 | **compatible_195** | **competence_73** | **complement_195** | complex_519 |
| component_624 | compose_198 | compound_86 | **comprehend_75** | **con_523** |
| **concept_1910** | **conclude_473** | **concrete_134** | conjunction_234 | **consequence_151** |
| consequent_151 | considerable_140 | consist_209 | consistent_207 | constitute_461 |
| constrain_414 | **construct_989** | construe_95 | content_858 | context_3124 |
| contradict_308 | contrary_85 | **contrast_842** | **controversy_113** | **convention_756** |
| convey_450 | cooperative_132 | **core_87** | **correlate_154** | correspond_628 |
| criteria_181 | **criticism_95** | **crucial_192** | cue_66 | **data_138** |
| **declarative_179** | default_91 | **define_1372** | **deictic_441** | deixis_352 |
| **demonstrate_297** | denote_406 | derive_520 | **description_1091** | depite_86 |
| **determiner_121** | **device_113** | diagram_104 | **dialogue_142** | dictionary_460 |
| **differ_307** | **dimension_221** | **dis_190** | **discourse_1275** | disjunction_109 |
| distinct_1723 | **distinguish_730** | **distribute_85** | **division_159** | **domain_557** |
| **dynamic_133** | elaborate_67 | **element_280** | elliptical_214 | embed_144 |
| **emphasis_66** | **emphasize_107** | empirical_130 | **encode_433** | **entail_989** |
| **entity_722** | entry_299 | epistemic_231 | equivalent_357 | **essential_331** |
| et_401 | **etc_809** | **evaluate_70** | evoke_108 | ex_117 |
| **exception_103** | **exclude_88** | exclusive_108 | exemplify_66 | existential_171 |
| experiment_116 | **explicit_490** | **explore_75** | extension_385 | **extent_134** |
| external_141 | factor_252 | failure_103 | false_855 | felicity_265 |
| figurative_69 | **focus_627** | **formal_612** | former_151 | **formula_478** |
| framework_200 | frequent_194 | **function_1133** | **fundamental_145** | furthermore_117 |
| gap_99 | gender_80 | gnerate_138 | generic_101 | **gesture_122** |
| grammar_415 | grammatical_887 | gricean_247 | hence_221 | **hereby_86** |
| **hierarchy_242** | homonym_102 | **hypothesis_171** | **hypothetical_71** | ideal_95 |
| **identical_146** | **idiom_177** | illocutionary_430 | **imperative_139** | implicate_661 |

| implicit_194 | importance_143 | impose_69 | inclusion_78 | inclusive_67 |
|---|---|---|---|---|
| independent_336 | indirect_409 | infer_1279 | informative_238 | inherent_128 |
| initial_123 | input_140 | intension_130 | inter−160 | interact_315 |
| internal_161 | interpret_1859 | interrogative_114 | intonation_129 | intransitive_105 |
| intuition_127 | intuitive_190 | irony_135 | isolate_87 | italic_79 |
| label_160 | latter_216 | lexeme_660 | lexical_1144 | lexicon_208 |
| linguist_395 | linguistic_2421 | link_276 | literal_592 | literature_201 |
| logic_2019 | mammal_77 | maxim_723 | mechanism_155 | ment_120 |
| mere_191 | metalanguage_146 | metalinguistic_110 | metaphor_1014 | method_116 |
| metonymy_208 | minimal_95 | modal_280 | modality_298 | modify_126 |
| moreover_81 | morphology_225 | motion_280 | motivate_169 | mutual_167 |
| necessity_76 | negate_569 | negative_430 | neutral_71 | nevertheless_99 |
| nominal_144 | notation_192 | notion_991 | noun_1272 | novel_84 |
| objected_69 | oblige_125 | obtain_67 | omit_71 | optimal_77 |
| oriented_111 | origin_88 | ostensive_78 | outline_83 | overall_75 |
| overlap_71 | overt_121 | paradigm_151 | parallel_130 | parameter_66 |
| paraphrase_188 | participant_446 | particle_119 | passage_72 | passive_126 |
| perception_68 | perspective_163 | phenomenon_426 | philosophy_567 | phonology_181 |
| phrase_1188 | plausible_106 | plural_159 | polar_210 | polysemy_377 |
| pose_108 | postulate_212 | potential_227 | pragmatic_2329 | pre_229 |
| precede_109 | precise_270 | predicate_1327 | predict_352 | premise_141 |
| preposition_145 | presence_121 | preserve_84 | presuppose_1322 | primary_267 |
| principle_1328 | prior_173 | proceed_76 | pronominal_69 | pronoun_574 |
| proposition_2221 | proto_80 | prototype_364 | psychology_291 | quantify_105 |
| quantity_186 | rational_95 | recipient_71 | recognition_107 | redundant_67 |
| reflect_202 | reject_153 | relative_506 | relevance_729 | relevant_709 |
| request_401 | resemble_131 | resolve_81 | respective_186 | response_221 |
| restrict_438 | reveal_112 | review_93 | rhetoric_161 | salient_166 |
| scalar_243 | schema_139 | scheme_72 | scholar_99 | scope_404 |
| semantic_4847 | semiotic_160 | senses_610 | sensitive_83 | sentential_75 |
| sequence_238 | sibling_72 | simultaneous_76 | sincere_83 | singular_165 |
| sion_72 | sole_88 | solution_155 | source_354 | spatial_170 |
| status_235 | stimulus_109 | straightforward_78 | strategy_201 | strict_191 |
| structure_1895 | subordinate_94 | subsequent_134 | substance_81 | substitute_104 |
| suceed_80 | sufficient_105 | summary_197 | superordinate_98 | symbol_314 |
| synonym_394 | syntactic_665 | syntagmatic_98 | syntax_331 | target_234 |
| task_208 | tautology_66 | taxonomy_194 | technical_209 | temporal_220 |
| terminology_139 | text_370 | theme_407 | theoretical_456 | theory_2368 |
| thesis_86 | tic_75 | token_196 | transitive_213 | translate_335 |
| trigger_136 | typology_76 | ultimate_90 | un_69 | underlie_220 |
| unique_170 | universe_534 | unlike_143 | usage_229 | utter_3716 |
| vague_129 | valid_187 | variety_242 | verb_2677 | via_100 |
| vice_82 | violate_163 | virtue_183 | vocabulary_255 | vocal_69 |
| volume_170 | vs_400 | ward_101 | yield_93 | |