

## Lexical bundles in linguistics textbooks\*

Hye-Kyung Lee  
(Ajou University)

Lee, Hye-Kyung. **Lexical bundles in linguistics textbooks.** *Linguistic Research* 37(1), 121-145. In an attempt to help students in linguistics enhance their reading proficiency, this study identifies a total of 274 three- and four-word lexical bundles drawing on a 1.14-million-word corpus of linguistics textbooks. These bundles were first investigated in terms of their structural properties. Results of the structural analysis showed that NP-based and PP-based bundles accounted for almost 80 % of all bundles, consistent with previous studies showing that professional academic writing contained more intensive use of NPs than student writing (both native and non-native). Functional taxonomy of lexical bundles indicated that a high proportion of lexical bundles had referential functions (84.9%), whereas much less of them were discourse organizers (8.6%) and stance expressions (6.5%). Results of the present study add corroborative evidence to those of previous studies. The current study also produces a few findings regarding subject relatedness of yielded lexical bundles, the use of colloquial expressions (i.e., *a lot of* or pronoun-framed bundles), a higher frequency of passive constructions, and a notable number of tokens of the bundle *around the world*. Pedagogical implications of these findings are also suggested. (Ajou University)

**Keywords** lexical bundles, academic discourse, discipline-specificity, subject relatedness, linguistics textbooks

### 1. Introduction

Acquiring adequate vocabulary is pivotal in academic discourse, primarily because it provides students and professionals alike with a useful means to properly communicate and participate in relevant academic discourse communities. At the same time, the notion of appropriate lexis has undergone a gradual but significant shift to include both a set of words and recurrent word combinations (Carter 2006), since word knowledge encompasses not merely the

---

\* An earlier version of this paper has been presented in the 2019 KASELL Spring Conference. I would like to thank two anonymous reviewers for their constructive comments and suggestions. However, all remaining errors are mine.

meaning of a single word, but the meaning it can present together with its co-occurring words. Research on such word combinations has been one of the fastest-growing fields in linguistics (Hyland 2008a), permeating all areas of natural language (e.g., Biber, Leech, Conrad, and Finegan 1999; Wray 2008; Simpson-Vlach and Ellis 2010; Wang 2017). From a psycholinguistic point of view, recurrent word combinations have “a processing advantage over creatively generated language” (Conklin and Schmitt 2008: 72), in that they “stored, retrieved whole from memory at the time of use” (Wray 2002: 465). These prefabricated linguistic units can also function as a key index to language competence, while presenting major challenges especially for non-native speakers, mainly because “they are language specific and enormous in number” (Liu 2012: 25).

Tools and methods in corpus linguistics have immensely facilitated and enriched research on recurrent multi-word sequences in a wide range of areas and perspectives. Multi-word units have been proven to be able to serve the role of characterizing a text as belonging to a particular register (e.g., Biber, Conrad, and Cortes 2004; Hyland 2008a). They also turned out to be indicative of developmental differences between different groups of writers such as students and professionals (e.g., Cortes 2004; Chen and Baker 2010) and between native and non-native writers (e.g., Chen and Baker 2010; Ädel and Erman 2012; Pan, Reppen, and Biber 2016; Lee and Kim 2017). Recent years have witnessed an increasing body of literature on discipline-specific multi-word combinations (e.g., Cortes 2004; Hsu 2015; Gilmore and Millar 2018; Jhang, Kim, and Qi 2018), suggesting their teaching worthiness and useful pedagogical implications for training non-native and novice writers in specific disciplines.

Given the significant role of knowledge of multi-word combinations, the current study aims to establish a list of multi-word constructions derived from core linguistics textbooks in English commonly utilized in Korean tertiary institutions. In doing so, this research seeks to help EFL students including Korean students adequately acquire the command of multi-word combinations and to suggest practical implications on educating students in linguistics to attain adequate lexical knowledge necessary for successful reading comprehension. This study is guided by the following research questions.

- 1) What are the most frequent three- and four-word sequences in core

linguistics textbooks?

- 2) What are predominant characteristics of multi-word units typical of the discipline of linguistics?
- 3) Which lexical bundles are unique to linguistics?

## 2. Literature review

A terminological clarification is needed first. Multi-word combinations have been referred to in a wide variety of ways, some of which include recurrent word combinations (e.g., Altenberg 1998; De Cock 1998), phraseology (e.g., Altenberg 1998; Granger and Meunier 2008), lexical phrases (Nattinger and DeCarrio 1992), chunks (e.g., De Cock 2000), clusters (e.g., Schmitt, Grandage, and Adlophs 2004; Hyland 2008a), n-grams (e.g., Stubbs 2007), formulas (e.g., Simpson-Vlach and Ellis 2010), multi-word constructions (e.g., Liu 2012), and lexical bundles (e.g., Biber and his colleagues; Chen and Baker 2010; Ädel and Erman 2012; Lee and Kim 2017, to name a few). Indeed, Wray (2002) found more than 50 different terms for these kinds of constructions. The present study will adopt the term *lexical bundle*, while using other terms interchangeably when needed. Lexical bundles often do not coincide with complete syntactic phrases or self-contained semantic units, because they are automatically retrieved via a data-analyzing tool.

Much of the research attention in lexical bundles has stemmed from interest in register variations. Altenberg (1998), Moon (1998), and Biber et al. (1999) have formed the first major studies of lexical bundles. Drawing on several registers such as fictions, magazines and speeches from Oxford Hector Project corpus, Moon (1998) has carried out cross-register comparisons of lexical bundles. Without utilizing any computer program, Moon (1998) conducted a functional analysis guided by a list of lexical bundles selected from dictionaries and other references. Biber and his colleagues (e.g., Biber and Conrad 1999; Biber et al. 1999; Biber and Barbieri 2007) undoubtedly have been positioned as pioneering scholars in the area of lexical bundle studies. One major finding of note is that spoken and written registers can vary substantially in structural patterns of lexical bundles. While spoken register makes an extensive use of clausal bundles,

academic prose almost exclusively uses phrasal bundles (Biber et al. 1999: 996-997).

Research efforts have also been made to develop lexical bundle lists in general academic discourse (e.g., Simpson-Vlach and Ellis 2010; Liu 2012; Gardener and Davies 2014) and in specific disciplines (e.g., Cortes 2004; Hyland 2008b; Hsu 2015; Qi 2019). As correctly pointed out by Simpson-Vlach and Ellis (2010) and acknowledged by Biber and his associates themselves, Biber and his colleagues' frequency-and-dispersion-alone approach blurs the intuitive distinction between cognitively salient lexical bundles (e.g., *on the one hand*) and incomplete bundles (e.g., *to do with the*) as long as they have equivalent frequencies.<sup>1</sup> To address this issue, Simpson-Vlach and Ellis (2010) have developed an empirically-derived method by combining two statistical measures (i.e., frequency and Mutual Information score) and ESL professions' teaching worthiness judgments.<sup>2</sup> They then successfully constructed three lists of lexical bundles: a list of core lexical bundles, a list of top 20 lexical bundles in spoken academic English, and a list of top 200 lexical bundles in written academic English. Although the new approach developed by Simpson-Vlach and Ellis (2010) could substantially reduce the number of incomplete lexical bundles, their lists still contained a small set of incomplete bundles such as *this is the* and *and this is*, as commented by Liu (2012: 27). While dealing with this problem by adopting a revised way of representing such bundles, Liu (2012) constructed an academic bundle list of 228 common items using data from academic written sub-corpora of the COCA and BNC. Liu (2012) further suggested new findings about usage patterns of written academic bundles. Of particular interest was the high frequency of "NP + action verb (e.g., *suggest*) + that", contrary to results of previous studies (e.g., Biber et al. 1999) that found passive forms to be a prevalent feature of written academic register.

In recent years, a body of research uses corpora to identify discipline-specific lexical bundles, primarily to help learners in a specific discipline to become

---

1 Biber and Conrad (1999) found that of the lexical bundles they compiled only 15% in spoken register and 5% in academic writing take the complete structure.

2 Mutual Information is a statistical procedure employed to determine the mutual dependence between the words, with a high MI score indicating a strong co-occurrence tendency. ([https://en.wikipedia.org/wiki/Mutual\\_information](https://en.wikipedia.org/wiki/Mutual_information))

familiarized with formulaic resources commonly circulated in that discipline (Cortes 2004; Hyland 2008b; Hsu 2015; Lu, Lee, and Jhang 2017; Gilmore and Millar 2018). While comparing academic writing in the disciplines of biology and history, Cortes (2004) found considerable variations between lexical bundles in these two disciplines. For example, lexical bundles in biology marked a more varied structural patterns and epistemic/stance functions. In the comparison of four-word clusters in four different disciplines, Hyland (2008b) showed that electronical engineering texts employ a wider range and a higher number of lexical bundles compared to biology, business and applied linguistics texts, proving a great disciplinary specificity of lexical bundles. Hsu (2015) developed a list of lexical bundles for EFL engineering students in Taiwan. Based on a corpus of college textbooks across twenty engineering subjects, Hsu (2015) compiled a total of 1,000 two- to six-word lexical bundles that are highly technical and common in diverse engineering disciplines. Lu et al (2017), focusing on maritime English, discussed the differences between international public and private maritime laws written in English with reference to lexico-grammatical, semantic and discourse features such as key words, key clusters and key semantic domains. Gilmore and Millar (2018) identified three- to six-word lexical bundles associated with the discipline of civil engineering and compared them with those analyzed by Hyland (2008b). Results of the comparison conducted by Gilmore and Miller (2018) indicated a higher reliance on research-oriented bundles but fewer cases of text-oriented bundles in civil engineering than the disciplines discussed by Hyland (2008b).

Research has also been rigorously conducted regarding characteristics of lexical bundles produced Korean writers (e.g., Hong 2013; Yoon and Choi 2015; Nam 2016; Lee and Kim 2017, and references therein). One characteristic attested in most studies is Korean writers' ready use of colloquial expressions, such as *have a lot/lots of*, person-based clauses (e.g. *I would like to*), contractions (e.g. *don't need to*), and combinations of these (e.g. *if we don't*). Korean learners' writing also displays less use of stance markers than native writing does.

While previous research shed light on the understanding of features of lexical bundles in different registers, groups of writers, and disciplines, no study to date has yet compiled or investigated lexical bundles characteristic of the discipline of linguistics. By analyzing structural, functional and subject-specific features of

lexical bundles in core linguistics textbooks, this study aims to add a valuable contribution to the development of lexical bundles in the field of linguistics. Another aim of this study is to assist EFL students including Korean students to adequately acquire the control and comprehension of lexical bundles.

### 3. Data and methods

#### 3.1 Data

To construct an academic word list for students in linguistics, Kim and Lee (2018, 2019) have developed Linguistics Academic Vocabulary List (LAVL henceforth) based on a corpus developed out of five major linguistics textbooks with 1.14 million running words called Corpus of Linguistics Textbooks (CLT henceforth). Compositions of the corpus are presented in Table 1.

Table 1. Compositions of CLT (modified from Kim and Lee 2019: 35)

Textbooks	Types	Tokens	Families	TTR	STTR
Book 1	16,293	245,898	4,342	6.63	35.85
Book 2	16,056	252,268	7,070	6.36	36.41
Book 3	8,651	107,391	2,894	8.06	34.61
Book 4	14,533	203,203	2,346	7.15	36.05
Book 5	15,371	333,070	6,429	4.61	32.50
Total	70,904	1,141,830	23,081	6.21	34.90

*Note.* Book 1 = *Introduction to language* (10<sup>th</sup>ed.); Book 2 = *How English works: A linguistic introduction* (3<sup>rd</sup>ed.); Book 3 = *Introducing English linguistics*; Book 4 = *Linguistics for everyone: An introduction* (2<sup>nd</sup>ed.); Book 5 = *Language files: Materials for an introduction to language and linguistics* (12<sup>th</sup>ed.)

According to Kim and Lee (2019), students in linguistics may acquire English vocabulary of the most frequent 7,000 word families as well as proper nouns, transparent compounds, abbreviations and glossary to have 95% lexical coverage of an English medium linguistics textbook. Acknowledging the challenge faced by students to learn as many as 7,000 word families, Kim and Lee (2019: 38) compiled a list of words (LAVL) of “utmost relevance and usefulness” to students in linguistics. The LAVL consists of 607 word families beyond general high-frequency two-thousand words. These two-thousand words are believed to serve as a vocabulary spring board for high school graduates.

As discussed in Introduction above, students need to possess competence of both individual words and recurrent word combination common to their disciplines for adequate reading comprehension. As pointed out by Hsu (2015), the first two-thousand words and the LAVL are still not enough for students in linguistics because the otherwise unnoticed word sequences can back up the function of individual words. With an aim to complement the LAVL, the Linguistics Academic Bundle List (LABL) is then derived from the CLT by using methodology modified from Hsu (2015).

### 3.2 Procedure

Constructing a lexical bundle list involves both qualitative and quantitative approaches. To retrieve a bundle list, Wordsmith 7.0 (Scott 2018) was used. When generating a list of lexical bundles, a few key criteria have been suggested in the literature, including the length of a lexical bundle, the cut-off frequency, and the dispersion. This study will examine three- and four-word bundles because they, especially four-word bundles, are “the most researched length” and “within a manageable size for manual categorization and concordance checks” (Chen and Baker 2010: 32). Three-word bundles are included because several trials have revealed that they display a wide range of productive and relevant expressions. Frequency and dispersion thresholds vary depending on the size of the corpus and the purpose of the research. After repeated experiments with the current corpus, the frequency and dispersion thresholds for 3- and 4-word lexical bundles were set to be 20 times or more, appearing in all 5 different texts. In the literature, the cut-off frequency depends on the size of the corpus. For a large corpus, a normalized frequency threshold is usually employed ranging from 20 to 40 times per million words (e.g., Biber et al. 2004; Hyland 2008a), while for a small corpus a raw cut-off frequency ranging from 2 to 10 times is adopted. Dispersion thresholds also differ from research to research. They are usually 3 to 5 different texts (e.g., Biber and Barbieri 2007; Chen and Baker 2010) or 10 percent of a corpus (e.g., Hyland 2008a).

Automatic retrieval of lexical bundles via Wordsmith 7.0 (Scott 2018) generated a total of 1,164 entries, including 1,044 (50,715 tokens) three-word

bundles and 120 four-word bundles (4,751 tokens). These retrieved data were then manually revised using a modified methodology frequently adopted in the literature including Chen and Baker (2010), Hsu (2015) and Lee and Kim (2017). For an academic lexical bundle list to serve pedagogical purposes, bundles should be semantically and syntactically well-formed (Hsu 2015; Qi 2019). Hence meaningfulness and grammaticality guided manual refining of the bundle list. This eliminated sequences such as *is English is a* and *all over the*. One deviation from this criterion is the inclusion of structurally unacceptable sequences such as *for instance in the*, *regardless of the*, and *relationship between the*. These were retained because they contained versatile two-word sequences *for instance*, *regardless of* and *relationship between*. Their prepositions should be studied by learners.

To make the lexical list more compact, word sequences that contained proper nouns were excluded from the extracted list, following the tradition in the literature. This criterion sifted out word sequences such as *the United States*, *the New York Times*, and *in New York City*. However, some bundles such as *of the Middle English period* and *the Great Vowel Shift* were retained even though they contain proper nouns, because they were employed as key terms in linguistics texts.

The so-called free word combinations such as *of a word/language*, and *verbs adjectives and adverbs* were removed when the meaning of the bundle is compositional or transparent enough to be calculated from meanings of the component words. Nevertheless, a group of collocations deemed closely interconnected and conceptually meaningful as a whole were included in the list, such as *consonants and vowels*, *number and gender*, and *pidgins and creoles*.

Word sequences that overlap with each other were merged after concordance checks because they could inflate the number of bundles. For example, two bundles *is important to note* and *important to note that* that appeared 20 times were combined into a longer sequence *be important to note (that)*, because they both occurred as subsets of this 5-word bundle. All cases of overlapping word combinations were merged into longer subsuming ones. When a 3-word sequence turned out to be part of a 4-word sequence, only the latter with its frequency was included in the final list. This process eliminated bundles such as *be used to* ( $\rightarrow$  *can be used to*), *is possible to* ( $\rightarrow$  *it is possible to*) and *for the most* ( $\rightarrow$  *for the most part*).



Another modification was undertaken for potential variants. Lexical bundles regarded as closely related either structurally or semantically were combined into one form with accumulative frequency. For example, *a discussion of* (36 times) and *the discussion of* (35) were combined to form a single item with *the/a discussion of*. This modification resulted in items such as *a/this/the/two/different/other type(s) of*, *the meaning(s) of*, and *an/another/some example(s) of*. Numbers of bundles before and after refinement are presented in Table 2.

Table 2. Numbers of bundles before and after refinement

Corpus	Before refinement		After refinement	
	No. of types	No. of tokens	No. of types	No. of tokens
CLT	10,715	55,466	274	16,901

Although structural and semantic completeness guided the manual screening of lexical bundles, a group of incomplete bundles that were all *the*-ending bundles were included in the LAVL for the reason discussed above in this section (e.g., *for instance in the*, *regardless of the*, *relationship between the* and *be one of the*). As for the representation of bundles, a schematic method was adopted (e.g., Liu 2012). For example, the representation of the bundle *be one of the* is changed into *be one of det + NP*. As discussed by Liu (2012: 27-28), these modified representations are justified because they better reflect both linguistic description and language acquisition theories which view language as a system consisting of grammatical structures.

## 4. Results and discussion

### 4.1 Structural and functional characteristics of LABL

Finalized bundles were investigated in terms of their structural properties using the structural taxonomy modified from Biber et al. (1999). Results are presented in Table 3 below.

Table 3. Structural distribution of LABL

Category	Pattern	Percentage	Example in LABL
NP-based	1) NP+post-modifier fragment	54.0%	<i>a combination of</i>
PP-based	2) preposition+NP fragment	25.2%	<i>as a result</i>
	3) copular <i>be</i> +NP/AdjP	1.8%	<i>be one of the</i>
	4) VP with active verb	2.9%	<i>seem to be</i>
	5) anticipatory <i>it</i> + VP/AdjP + (complement clause)	2.2%	<i>it is clear</i>
	6) passive verb+PP fragment	6.5%	<i>be added to</i>
VP-based	7) (VP+) <i>that</i> -clause fragment	0%	
	8) (verb/adjective+) <i>to</i> -clause fragment	4.7%	<i>be likely to</i>
Others	9) others	2.9%	<i>may or may not</i>
	<b>Total</b>	<b>100%</b>	

As seen in Table 3, NP-based and PP-based bundles accounted for almost 80% of all bundles. This result is consistent with a tendency observed in the literature (e.g., Chen and Baker 2010; Lee and Kim 2017) showing that professional academic writing contains much more intensive use of NPs than student writing (both native and non-native) which displays a greater dependency on VP-based bundles instead. The LABL also contains a noticeable number of the so-called “fixed frames” (Biber, et al. 2003: 78), which refer to two productive frames “the + Noun + of the/a” and “in + the + Noun + of”. Some nouns used in fixed frames in the LABL that appeared more than 50 times are presented in Table 4. Examples involving these fixed frames are given in (1), (2), and (3).

Table 4. Nouns in “fixed frames”

LABL	(in) the + NOUN + of + (the/a)
	age, back, beginning, case, development, distribution, effect(s), end, front, grammar, history, meaning(s), middle, notion, number, process, production, pronunciation, rules, use, form, result(s)

- (1) When voiced obstruents occur at **the end of a** word or syllable, they become voiceless (Book 1)
- (2) The answer to that question depends on **the history of the** marriage. (Book 3)

- (3) **In the case of siblings**, there appears to be an age distinction only if the sibling is the same sex as the speaker (Book 5)

The functional taxonomy of lexical bundles developed by Biber et al. (2003, 2004) are mostly employed in the classification of lexical bundles with some adaptations in each study (e.g., Chen 2009; Chen and Baker 2010; Ädel and Erman 2012; Liu 2012; Lee and Kim 2017; Lee 2018). While Biber et al.'s methodology plays an influential role in relevant research, drawbacks inherent in their system have often been noted. As rightly pointed out by Liu (2012: 30), Biber et al.'s categorization involves subjectivity. In addition, one bundle may frequently serve more than one function. Despite these problems, a semantic classification of culled lexical bundles was carried out to get a sense of a general pattern of their semantic/functional properties. The methodology adopted here followed those of Chen and Baker (2010) and Liu (2012) with some modifications, which were again built on Biber et al. (2003, 2004). The system here divides bundles into three major categories: referential, discourse-organizing, and stance expressions, with each category involving several sub-categories. According to Biber et al. (2004: 384), referential bundles make reference to physical or abstract entities, the textual context, or some attribute of the entity. Stance bundles signify attitudes or evaluation of certainty about a proposition while discourse organizers mark relationship between prior and coming discourse. One major modification in the current study is the addition of the sub-category of Technical Terms under the Referential-expressions category, which is set up to accommodate technical terms pertaining to linguistics. This is described in more detail in section 4.3. Table 5 presents information about functions of lexical bundles.

Table 5. Functional classification of LABL (types)

Category	Sub-category	Percentage	Example in LABL
Referential expressions	Framing attributes	33.5%	<i>a combination of, in the meaning of</i>
	Identification	3.6%	<i>be known as, to refer to</i>
	Referential time/place	11.5%	<i>around the world, in the future</i>
	Quantify specification	7.6%	<i>the majority of, a large number of</i>
	Reporting/description	16.5%	<i>be different from, it turns out</i>
	Technical terms	12.2%	<i>manner of articulation, pidgins and creoles</i>

<b>Sub-total</b>		<b>84.9%</b>	
Discourse	Linking	7.4%	<i>on the other hand, as well as</i>
organizers	Inferential	1.1%	<i>DET result of, in the sense</i>
<b>Sub-total</b>		<b>8.6%</b>	
Stance	Epistemic stance	4.0%	<i>seem to be, be (more) likely to</i>
expressions	Attitudinal/modality	2.5%	<i>be important to note, be able to</i>
<b>Sub-total</b>		<b>6.5%</b>	
<b>Total</b>		<b>100%</b>	

Results of the analysis revealed that a high proportion of the lexical bundles in the LABL served referential functions (84.9%) whereas much less of them were stance discourse organizers (8.6%) and stance expressions (6.5%). Results presented here add corroborative evidence to those of several previous studies (e.g., Liu 2012; Hsu 2015), suggesting that referential bundles made up the largest proportion in academic written texts.

#### 4.2 Subject relatedness of LABL

The current study produced a few findings regarding subject relatedness of the yielded lexical bundles, which addresses Research Question 3 suggested in Introduction. Pertaining to the literature of academic words, a division between technical and non-technical vocabulary has been made (e.g., Yang 1986; Baker 1988; Chung and Nation 2003, 2004; Hsu 2015). For example, Chung and Nation (2004: 252) have proposed a four-point rating scale of technicality, according to which “technical vocabulary is subject related, occurs in a specialist domain, and is part of a system of subject knowledge”, whereas general vocabulary is dubbed words with minimal/little association with a specific discipline. At the same time, an intermediate category between the two has been put forward under such terms as sub-technical (e.g., Cowan 1974; Yang 1986; Baker 1988; Hsu 2015), semi-technical (e.g., Farrell 1990), and lay-technical (e.g., Fraser 2009). In the present study, the label ‘sub-technical lexical bundles’ will be chosen to denote those that have one or more “general” English meanings which in technical contexts take on “extended meaning” (Trimble 1985: 129). Technical lexical bundles refer to those that “reflect strictly technical knowledge and behave as highly-specialized ones” (Hsu 2015: 125). Still, it should be acknowledged that

the demarcation between sub-technical and technical vocabulary is not as easy as it seems in many cases. Hsu (2015: 125), for example, claims that multi-word sequences such as *immobilized cell bioreactor* and *dielectrically controlled resolution* are unique to a specific engineering discipline as they are employed almost exclusively in the engineering context. Contrary to Hsu's argument, however, the technicality of these sequences appears to stem from the technicality of individual words rather than from word sequences. Noticing the problem involved in demarcating sub-technicality and technicality, the current study is going to subsume these two categories into one.

A notable number of linguistics-specific (sub)-technical lexical bundles were identified. Table 6 presents linguistics-specialized bundles with their frequencies.

Table 6. (Sub)-technical bundles in LABL

<b>Bundle</b>	<b>Freq.</b>	<b>Bundle</b>	<b>Freq.</b>	<b>Bundle</b>	<b>Freq.</b>
DET noun phrase	161	the oral cavity	50	a prepositional phrase	28
place(s) of articulation	138	the vocal cords	50	maxim of quantity	25
in Old English	131	the alveolar ridge	48	a speech act	25
DET verb phrase	91	consonants and vowels	48	maxim of quality	24
the vocal track	87	the cooperative principle	43	number and gender	24
Early Modern English	84	the present tense	41	pidgins and creoles	24
DET direct object	75	an auxiliary verb	32	a nasal consonant	21
manner of articulation	73	the Middle English period	32	of universal grammar	21
third person singular	55	the plural morpheme	28		

A set of bundles display compositional meanings. If meanings of component words of a bundle are known, the meaning of the bundle can be inferred (i.e., they can be dubbed transparent lexical bundles). For example, in the case of the bundles *DET noun/ verb/ prepositional phrase*, the senses of the bundles directly draw on senses of involved words. Since they are unique to linguistics contexts, students should acquire them for appropriate comprehension of the texts. Indeed, the bundle *DET noun phrase* ranks the 16<sup>th</sup> most frequent bundle in the LABL.

Several constructions take the syntactic form of *A and B*, including bundles such as *consonants and vowels*, *number and gender*, and *pidgins and creoles*. As discussed in section 3 above, these bundles were included in the LABL because they were viewed to be closely interconnected and conceptually meaningful as a whole rather than to be free word combinations. Some component words of these bundles are included in the LAVL (Kim and Lee 2018, 2019) while some are not. The LAVL contains *vowel*, *pidgin* and *creole*. This means that the LAVL alone cannot accommodate the fact that the word *pidgin* fairly frequently collocates with the word *creole* or the fact that the word *vowel* frequently collocates with the word *consonant*. These bundles also show that these words are juxtaposed in this specific order rather than in the other way, making such juxtaposed bundles worth mentioning.

Equally noticeable is the high frequency of opaque bundles, in which the sum of meanings of the individual words of a lexical bundle fails to render the meaning of the relevant lexical bundle as these bundles are employed highly technically. Opaque bundles encompass terms such as *the cooperative principle*, *maxim of quantity/quality* and *a speech act*. Learners might be able to construe meanings of these bundles without any instructions on these terms. However, as they are employed discipline-specifically, their definitions should be targeted for explicit instructions. Indeed, most textbooks present definitions of such terms before these terms are further explored or discussed with reference to language data. Excerpts (4) and (5) display initial occurrences of the two bundles (i.e., *the cooperative principle* and *speech act*) where technical senses of these terms are explicitly explained.

- (4) The philosopher H. Paul Grice proposed **the cooperative principle** to explain how conversation involves a certain level of “cooperation” among communicants: Our talk exchanges do not normally consist of a succession of disconnected remarks, and would not be rational if they did. They are characteristically, to some degree at least, cooperative efforts. (Book 3)
- (5) We call this underlying purpose of the utterance, be it a reminder, a warning, a promise, a threat, or whatever, the illocutionary force of a

speech act. (Book 2)

### 4.3 More findings

More results featuring the LABL were found. A few colloquial expressions made the LABL, including *a lot of* or pronoun-framed bundles. Personal pronouns are proven to dominate the conversation register bundle list in the most frequent four-word bundles from registers of the British National Corpus Baby edition (Hyland 2008a: 45). The use of *a lot* and *lots of* are also believed to be related to colloquial or informal styles (Quirk, Greenbaum, Leech, and Svartvik 1985). Even though *a lot of* and its variants (i.e. *lot of*, *lots of*) were deemed characteristic of colloquial register, they actually made some preceding academic bundles lists: *a lot of* in Liu (2012) and *a lot (of)* and *lots of* in Hsu (2015). Hence, it is open to debate whether bundles containing *lot(s)* characterize colloquial and informal registers. Still, academic bundle lists seldom feature colloquial pronoun-framed bundles, as academic texts belong to formal and mostly written register (c.f. Liu (2012) for a few exceptions). Concordance checks of relevant bundles indicate why they feature the LABL as opposed to lexical bundles lists of other disciplines (e.g., Liu 2012; Hsu 2015). The two pronoun-framed bundles are utilized to present language data, which is clearly indispensable in linguistics textbooks (Kim and Lee 2018, 2019). Consider excerpts in (6) and (7).

- (6) Because implicatures result from violations of one or more maxims, they can be easily cancelled by providing further, clarifying information. For example:

Dad: Very nice girl. **What do you think**, Hon?

Mom: The turkey sure was moist. Toni basted it every ten minutes.

The additional remark cancels, or at least weakens, the implicature that Mom dislikes Toni. (Book 1)

- (7) Gephardt provides a very lengthy answer: The electors are going to elect George W. Bush to be the next president of the United States, and I believe on January 20, not too many steps from here, he's going to be sworn in as the next president of the United States. **I don't know** how

you can get more legitimate than that. (Book 3)

The bundles in (6) and (7) serve the function of presenting made-up conversations or quoting someone's utterances. Because exploring spoken data, either constructed or quoted, plays a key role in the discussion of linguistics-related issues, occurrences of pronoun-framed constructions can be easily accommodated as one pivotal feature of the LABEL.

Another finding deserving attention is the issue surrounding the preference between passive constructions and active constructions in academic texts. It should be cautiously mentioned that utilizing passive constructions is highly complex due to its multi-faceted lexical, semantic and contextual factors in play (e.g., Quirk et al. 1985; Hinkel 1997). In the literature such as Biber et al. (1999), passive forms such as *it has been shown/suggested that* display higher frequencies. Still, passive forms are considered to be characteristic of Anglo-American academic writing (Biber 1988; Atkinson 1991; Ädel and Erman 2012), while, a later study by Liu (2012) demonstrates that active use of the construction such as "NP + linguistic action verb (e.g., *suggest*) + that" dominates his list. Drawing on his observation, Liu (2012: 31) further states that passive forms "should be high on the list of the MWCs (multi word constructions) for ESL/EFL writers to acquire." As opposing results concerning the debate may hinge on varying data selections or research methodologies, it is implausible to confirm which perspective better explicates the phenomenon under discussion. Nevertheless, the current study documents a high frequency of passive constructions, including the bundles in (8).

- (8) be (closely) related to/ be referred to as/ be added to/ be based on/ be composed of/ be derived from/ be exposed to/ be followed by/ be known as/ can be described/ can be traced back

Indeed, as many as 18 bundles (6.5%) were couched in passive constructions whereas eight bundles (2.9%) took active forms as shown in Table 3 above. The choice between passive and active constructions delicately correlates with a wide range of constraints. Hence, rather than presenting either form to students first, instructions should be accompanied by authentic examples where each type of



constructions is preferred. Excerpts in (9) and (10) display uses of the verb *refer to* in an active and passive forms.

- (9) Accent refers to the characteristics of speech that convey information about the speaker's dialect. . . . People in the United States often **refer to** someone as having a British accent or an Australian accent; in Britain they **refer to** an American accent. (Book 1)
- (10) Forbidden acts or words reflect the particular customs and views of the society. . . . In the world of Harry Potter, the evil Voldemort is not to be named but **is referred to** as "You Know Who (Book 1).

The contexts in these excerpts suggest why one is preferred to the other. The subject of the sentence must have been chosen in accordance with the author's intention to develop the argument. Then, either the active or passive construction is selected accordingly.

Of equal interest are a notable number of tokens of the bundle *around the world* in the LABL. Certain time/place deictic expressions such as *all over the world* and *for the first time* are regarded as typical of non-academic texts. The bundle *all over the world* is argued to indicate a general pattern of non-native writers to over-generalize and be categorial (Chen 2009; Chen and Baker 2010). Given that, the occurrence of the bundle *around the world* in the LABL is rather striking in that the LABL is developed from academic written texts. However, a close look at the concordance lines manifests why that specific bundle appears so frequently in the LABL.

- (11) Nevertheless, children **around the world** acquire language in much the same way (Book 1)
- (12) So the fact that many languages **around the world** make use of many of the same sounds in their phonological inventories is at least in part due to the similarity of the apparatus we all use to make those sounds. (Book 5)
- (13) As English spreads **around the world**, new varieties of English are born. (Book 2)
- (14) As a result of colonization, shipping, and trade, many creoles grew up

in coastal areas all **around the world**. (Book 4)

Excerpts in (11) and (12) elaborate on the topic of Universal Grammar or Language Universal, which refers to “parts that pertain to all languages” (Fromkin, Rodman, and Hymes 2014: 13). As supporting evidence for Universal Grammar, the common properties observed in the stages of language acquisition or across different languages are repeatedly suggested in linguistics textbooks as illustrated in (11) and (12). At the same time, language is an indispensable aspect of human being in general, because “we live in a world of language” (Fromkin et al. 2014: 1). Hence, making generalizations concerning socio-linguistic phenomena is in need, so is the use of the bundle *around the world* as illustrated in (13) and (14). Arguably, the use of bundles such as *around the world* does not necessarily mark non-native writers’ tendency to be over-generalize, contrary to Chen (2009) and Chen and Baker (2010). Rather, such deictic bundles in academic writing can serve an effective discipline-specific purpose as different disciplines require students to have different sets of lexical bundles at their disposal for successful reading comprehension.

## 5. Summary and conclusion

In an attempt to develop a list of lexical bundles pertaining to linguistics, this study identified a total of 274 lexical bundles of three and four words employing both quantitative and qualitative methods. The finalized bundles were first investigated in terms of their structural properties using the structural classification modified from Biber et al. (1999). Results of a structural analysis showed that NP-based and PP-based bundles dominated the LABL with up to 80% coverage. This finding is aligned with a tendency observed in some previous studies showing that expert academic writing contains more intensive use of NPs than student writing. The analysis of functional taxonomy of bundles revealed that referential bundles were dominant (84.9%) compared to discourse organizers (8.6%) and stance bundles (6.5%), consistent with results of several preceding studies (e.g., Liu 2012; Hsu 2015). This study also documented a number of linguistics-specific (sub)-technical lexical bundles, the presence of

colloquial expressions (i.e., *a lot of* or pronoun-framed bundles), a high frequency of passive constructions, and occurrences of the bundle *around the world*.

Despite an increasing interest in discipline specificity, the paucity of discipline-specific research needs to be acknowledged. In particular, few studies have investigated lexical bundles frequently employed in the discipline of linguistics. The current study thus hopes to provide a useful means for students in linguistics when they study their core textbooks. It also expects to offer EAP/ESP (English for academic/specific purposes) instructors with pedagogical insights and practical guidelines in preparing teaching materials and designing a curriculum.

As this study utilized a limited number of texts, incorporating more textbooks can be pursued in future research. Given that different sub-disciplines of linguistics resort to different sets of vocabulary, investigating lexical variations among sub-disciplines of linguistics is needed in further research. In particular, it should be carefully investigated how these identified lexical bundles are suggested to students. That is, these bundles can be targeted for explicit attention and/or presented in authentic contexts for students themselves to infer their senses empirically.

## References

- Ädel, Annelie and Britt Erman. 2012. Recurrent word combinations in academic writing by native and non-native speakers of English: A lexical bundles approach. *English for Specific Purposes* 31: 30-49.
- Altenberg, Bengt. 1998. On the phraseology of spoken English: The evidence of recurrent word-combinations. In A. P. Cowie (ed.), *Phraseology: Theory, analysis and applications*, 101-122. Oxford: Oxford University Press.
- Atkinson, Dwight. 1991. Discourse analysis and written discourse convention. *Annual Review of Applied Linguistics* 11: 57-76.
- Baker, Mona. 1988. Sub-technical vocabulary and the ESP teacher: An analysis of some rhetorical items in medical journal art. *Readings in a Foreign Language* 4(2): 91-105.
- Biber, Douglas. 1988. *Variation across speech and writing*. Cambridge: Cambridge University Press.
- Biber, Douglas and Federica Barbieri. 2007. Lexical bundles in university spoken and writ-

- ten registers. *English for Specific Purposes* 26: 263-286.
- Biber, Douglas and Susan Conrad. 1999. Lexical bundles in conversations and academic prose. In Hilde Hasselgård and Signe Oksefjell (eds.), *Out of corpora: Studies in honour of Stig Johansson*, 181-190. Amsterdam: Rodopi.
- Biber, Douglas, Susan Conrad, and Viviana Cortes. 2003. Lexical bundles in speech and writing: An initial taxonomy. In A. Wilson, G. Leech, P. Rayson, and T. McEnery (eds.), *Corpus linguistics by the Lune: A festschrift for Geoffrey Leech*, 71-93. Frankfurt: Peter Lang.
- Biber, Douglas, Susan Conrad, and Viviana Cortes. 2004. If you look at ...: Lexical bundles in university teaching and textbooks. *Applied Linguistics* 25(3): 371-405.
- Biber, Douglas, Stig Johansson, Geoffrey Leech, Susan Conrad, and Edward Finegan. 1999. *Longman grammar of spoken and written English*. Longman: London.
- Carter, Ronald. 2006. What is advanced-level vocabulary? The case of chunks and clusters. In E. Coombe (ed.), *Words matter: The importance of vocabulary in English language teaching and learning*, 23-41. Alexandria, Va.: TESOL.
- Chen, Yu-Hua. 2009. *Lexical bundles across learner writing development*. Unpublished Ph.D. Thesis. Lancaster: Lancaster University.
- Chen, Yu-Hua and Paul Baker. 2010. Lexical bundles in L1 and L2 academic writing. *Language Learning and Technology* 14(2): 30-49.
- Chung, Teresa M. and Paul Nation 2003. Technical vocabulary in specialised texts. *Reading in a Foreign Language* 15(2): 103-116.
- Chung, Teresa M. and Paul Nation. 2004. Identifying technical vocabulary. *System* 32(2): 251-263.
- Conklin, Kathy and Norbert Schmitt. 2008. Formulaic sequences: Are they processed more quickly than nonformulaic language by native and nonnative speakers? *Applied Linguistics* 29(1): 72-89.
- Cortes, Viviana. 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.
- De Cock, Sylvie. 1998. A recurrent word combination approach to the study of formulae in the speech of native and non-native speakers of English. *International Journal of Corpus Linguistics* 3(1): 59-80.
- Farrell, Paul. 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, Simon. 2009. Breaking down the divisions between general, academic and technical vocabulary: The establishment of a single, discipline-based word list for ESP learners. *Hiroshima Studies in Language and Language Education* 12: 151-167.
- Fromkin, Victoria, Robert Rodman, and Nina Hyams. 2013. *An introduction to language*. (10<sup>th</sup>

- edition). Boston: Wadsworth.
- Gardner, Dee and Mark Davies. 2014. A new academic vocabulary list. *Applied Linguistics* 35: 305-327.
- Gilmore, Alexander and Neil Millar. 2018. The language of civil engineering research articles: A corpus-based approach. *English for Specific Purposes* 18: 1-17.
- Granger, Synviane and Fanny Meunier. (eds.) 2008. *Phraseology: An interdisciplinary perspective*. Amsterdam: John Benjamins.
- Hinkel, Eli. 1997. Indirectness in L1 and L2 academic writing. *Journal of Pragmatics* 27: 361-386.
- Hong, Sinchul. 2013. An n-gram analysis of Korean English learners' writing. *Korean Journal of English Language and Linguistics* 13(2): 313-336.
- Hsu, Wenhua. 2015. The most frequent formulaic sequences in college engineering textbook. *Corpus Linguistics Research* 1: 109-149.
- Hyland, Ken. 2008a. Academic clusters: Text patterning in published and postgraduate writing. *International Journal of Applied Linguistics* 18(1): 41-62.
- Hyland, Ken. 2008b. As you can see: Lexical bundles and disciplinary variations. *English for Specific Purposes* 27: 4-21.
- Jhang, Se-Eun, Sungkuk Kim, and Yilian Qi. 2018. Lexical bundles in ESP writing. Marine accident investigation reports. *Linguistic Research* 35(Special Edition), 105-135.
- Kim, Hyeon-Okh and Hye-Kyung Lee. 2018. Identifying the vocabulary size for Korean EFL students in linguistics. Paper presented at the 4th Asia Pacific Corpus Linguistics Conference. Takamatsu, Japan.
- 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): 27-53.
- Lee, Hye-Kyung. 2018. Stance expressions in Korean students' writing in English. *Discourse and Cognition* 25(4): 95-116.
- Lee, Hye-Kyung and Hyeon-Okh Kim. 2017. Features of lexical bundles in academic writing by non-native vis-à-vis native speakers of English. *English Language and Linguistics* 23(2): 67-88.
- Liu, Dilin. 2012. The most frequently-used multi-word constructions in academic written English: A multi-corpus study. *English for Specific Purposes* 31: 25-35.
- Lu, Wenyu, Sung-Min Lee, and Se-Eun Jhang. 2017. Keyness in Maritime institutional law texts. *Linguistic Research* 34(Special Edition): 51-76.
- Moon, Rosamund. 1998. *Fixed expressions and idioms in English: A corpus-based approach*. Oxford: Clarendon Press.
- Nam, Daeyun. 2016. Lexical bundle structure in argumentative writing of Korean and American university students. *The Jungang Journal of the English Language and Literature* 58(4): 347-370.

- Nattinger, James and Jeanette DeCarrio. 1992. *Lexical phrases and language teaching*. Oxford: Oxford University Press.
- Pan, Fan, Randi Reppen, and Douglas Biber. 2016. Comparing patterns of L1 versus L2 English academic professionals: Lexical bundles in telecommunications research articles. *Journal of English for Academic Purposes* 21: 60-71.
- Qi, Yilian. 2019. *Syntactic and semantic patterns of domain-specific multiword units in marine accident investigation reports*. Ph.D. Dissertation. Busan: Korea Maritime and Ocean University.
- Quirk, Randolph, Sidney Greenbaum, Geoffrey Leech, and Jan Svartvik. 1985. *A comprehensive grammar of the English language*. London: Longman.
- Schmitt, Nobert, Sarah Grandage, and Svenja Adolphs. 2004. Are corpus-derived recurrent clusters psycholinguistically valid? In Nobert Schmitt (ed.), *Formulaic sequences: Acquisition, processing and use*, 127-152. Amsterdam: John Benjamins Publishing.
- Scott, Mike. 2018. *Word Smith Tools 7.0*. Stroud: Lexical Analysis Software.
- Simpson-Vlach, Rita and Nick Ellis. 2010. An academic formulas list: New methods in phraseology research. *Applied Linguistics* 31(4): 487-512.
- Stubbs, Michael. 2007. An example of frequent English phraseology: Distribution, structures and functions. In Roberta Faccinetti (ed.), *Corpus linguistics 25 years on*, 89-105. Amsterdam: Radopi.
- Trimble, Louis. 1985. *English for science and technology: A discourse approach*. Cambridge: Cambridge University Press.
- Wang, Ying. 2017. Lexical bundles in spoken academic EFL: Genre and disciplinary variation. *International Journal of Corpus Linguistics* 22(2): 187-211.
- Wray, Alison. 2002. *Formulaic language and the lexicon*. Cambridge: Cambridge University Press.
- Wray, Alison. 2008. *Formulaic language: Pushing the boundaries*. Oxford: Oxford University Press.
- Wright, Heidi. 2019. Lexical bundles in stand-alone literature reviews: Sections, frequencies, and functions. *English for Specific Purposes* 54: 1-14
- Yang, Huizhong. 1986. A new technique for identifying scientific/technical terms and describing science texts. *Literary and Linguistic Computing* 1: 93-103.
- Yoon, Choongil and Ji-Myoung Choi. 2015. Lexical bundles in Korean university students' EFL compositions: A comparative study of register and use. *Modern English Education* 16(3): 47-69.

## Appendix

### The Linguistics Academic Bundle List (LABL)

**Note:** The LABL includes 274 three- and four- word lexical bundles. The number is given after each bundle to mark its frequency.

a combination of(29)	a dialect of(22)	a difference in(34)
a group of(83)	a hyponym of(23)	a large number of(32)
a lot of(70)	a matter of(36)	a member of (42)
a nasal consonant(21)	a pair of(21)	a piece of(21)
a prepositional phrase(28)	a range of(22)	a second/first language(120)
a sentence such as(24)	a series of(64)	a set of(113)
a speech act(25)	a speech sound(26)	a variety of(101)
a way of(25)	a word such as(27)	an auxiliary verb(32)
appear to be(60)	around the world(56)	as a result (of)(185)
as a whole(25)	as long as(20)	as opposed to(77)
as part of(47)	as soon as(21)	as well as (in)(371)
at some point(25)	at the beginning (of)(272)	at the end (of)(145)
at the same time(80)	at this state(25)	back of the tongue(23)
be added to(21)	be associated with(54)	be based on(46)
be composed of (21)	be derived from (20)	be determined by (25)
be (quite) different from (60)	be difficult to (25)	be exposed to (20)
be followed by(48)	be important to note (that)(20)	
be known as(44)	be (more) likely to(136)	be part of(90)
be pronounced as(50)	be (sometimes) referred to as(61)	
be (closely) related to(46)	be similar to(32)	be used as(25)
be used in(76)	between the two(66)	borrowed into English(25)
by means of(43)	can be described(22)	can be traced back(55)
can be used to(44)	can function as(24)	change over time(30)
consonants and vowels(48)	DET direct object(75)	DET discussion of(71)
DET example(s) of(184)	DET form of(144)	DET/ADJ kind(s) of(349)
DET noun phrase(161)	DET of these(124)	DET/DJ part(s) of(125)
DET result(s) of(158)	DET speaker(s) of(113)	DET state of(47)
DET system of(68)	DET/ADJ type(s) of(184)	DET verb phrase(93)
early Modern English(84)	exactly the same(55)	for instance in the(24)
for the most part(33)	from another language(23)	from one another(32)
have more than one(23)	I don't know(34)	in addition to(120)

- in contrast to(31)                      in DET case(s)(233)                      in DET (first) example(113)  
 in DET/ADJ way(s)(109)                      in each of these(20)                      in front of(32)  
 in Old English(131)                      in order to(321)                      in other words(174)  
 in place of(27)                      in relation to(32)                      in response to(29)  
 in terms of(237)                      in the ase of(70)                      in the form of(23)  
 in the future(25)                      in the history of(36)                      in the meaning of(30)  
 in the middle of(39)                      in the mouth(58)                      in the next section(28)  
 in the past(53)                      in the pronunciation(20)                      in the right(23)  
 in the sense(21)                      in the south(45)                      in the world(158)  
 in this book(54)                      in this context(28)                      in this section(57)  
 in words such as(43)                      it is clear(27)                      it is difficult to(25)  
 it is important to(72)                      it is necessary to(20)                      it is possible to(58)  
 it turns out(23)                      languages of the world(33)                      manner of articulation(73)  
 maxim of quality(24)                      maxim of quantity(25)                      may or may not(29)  
 more or less(49)                      more than one meaning(22)  
 native American languages(51)                      native speakers of English(20)  
 number and gender(24)                      of language acquisition(66)                      of speech sounds(65)  
 of the English language(25)                      of the past(23)                      of the same language(25)  
 of the world(154)                      of Universal Grammar(21)                      on the basis of(32)  
 on the context(21)                      on the first syllable(27)                      on the Internet(25)  
 on the other hand(200)                      one of the most(47)                      our knowledge of(56)  
 pidgins and creoles(24)                      place(s) of articulation(138)                      process by which(42)  
 regardless of the(24)                      relationship between the(48)                      seem to be(58)  
 speakers of DET/ADJ language(s)(75)                      spoken and written(47)  
 tend to be(43)                      the ability to(76)                      the act of(28)  
 the addition of(46)                      the age of(55)                      the alveolar ridge(48)  
 the application of(36)                      the back of(58)                      the basis of(41)  
 the beginning of(176)                      the case of(89)                      the category of(27)  
 the concept of(25)                      the context in which(26)                      the context of(42)  
 the Cooperative Principle(43)                      the creation of(45)                      the description of(21)  
 the development of(87)                      the difference between(124)                      the distinction between(42)  
 the distribution of(50)                      the effect(s) of(56)                      the end of(208)  
 the existence of(29)                      the extent to which(29)                      the focus of(29)  
 the formation of(29)                      the front of(51)                      the function of(23)  
 the grammar of(76)                      the history of(124)                      the history of English(44)  
 the idea of(35)                      the importance of(23)                      the introduction of(26)  
 the invention of(20)                      the language(s) of(151)                      the list of(21)  
 the loss of(33)                      the main clause(26)                      the majority of(37)  
 the meaning(s) of(454)                      the Middle English period(32)                      the middle of(53)



the name of(60)	the nature of(60)	the need to(20)
the notion of(70)	the number of(106)	the object of(30)
the oral cavity(50)	the order of(34)	the organization of(23)
the origins of(38)	the past tense(143)	the place of(35)
the plural morpheme(28)	the position of(35)	the present tense(41)
the process of(95)	the production of(67)	the pronunciation of(97)
the relationship between(83)	the rest of(49)	the role of(39)
the roof of(40)	the rules of(110)	the same is true(21)
the same meaning(55)	the same thing(68)	the sounds of(81)
the speech of(71)	the spread of(22)	the status of(34)
the structure of(116)	the study of how(30)	the subject of(69)
the syntax of(25)	the system of(24)	the top of(28)
the truth of(30)	the use of(300)	the vocal cords(50)
the vocal tract(87)	the word order(29)	third person singular(55)
to account for(38)	to be able to(34)	to be used(25)
to communicate with(34)	to determine whether(36)	to distinguish between(21)
to each other(58)	to one other(43)	to refer to(118)
to talk about(31)	to the end of(22)	to what extent(21)
true or false(44)	used to describe/indicate(56)	used to refer to(30)
varieties of (American) English(107)	whether or not(27)	what do you think(39)
what is means(24)	with the subject(23)	with one another(39)
with respect to(101)		with the tongue(30)
yes no questions(37)		

**Hye-Kyung Lee**

Professor

Dept. of English

Ajou University

206, World cup-ro, Yeongtong-gu

Suwon, 16499, Rep. of Korea

E-mail: hklee@ajou.ac.kr

Received: 2019. 11. 05.

Revised: 2019. 12. 29.

Accepted: 2020. 01. 12.