



## Korean English Learners' Use of Lexical Bundles in Speaking

**Maribel N. Zipagan**  
*Sejong University, Korea*

**Kyoung Rang Lee**  
*Sejong University, Korea*

This study explored the role of lexical bundles in the language development of L2 learners; the frequency, structural forms (patterns), and functions of formulaic utterances used by Korean English learners in a controlled environment across different proficiency levels were examined. Conscientious procedures were taken to identify and analyze lexical bundles of L2 learners' spoken data, since few studies have examined these, while some have compared written data. The oral response of 58 Korean English learners were transcribed and compared across three levels (novice, intermediate, and advanced). This study seemed to confirm the findings of previous studies, in that there could be an upward trend in language development as proficiency increases. It is worth noting the participants used lexical bundles in a limited manner similarly across the three proficiency levels, which was less frequent and less diverse in terms of structural and functional patterns when they had to produce oral data rather than written data. The results can be of significant help for learners to improve their English speaking, and for teachers to diagnose their students' speaking proficiency effectively, and should thus be integrated in English teaching curricula. The details of the results and implications are provided.

**Keywords:** lexical bundles; formulaic utterances; L2 learners' speaking; structural patterns; functional patterns

### Introduction

Second language (L2) learners naturally play with words and try out new phrases for better communication. These word combinations, which sound formulaic and recurrent, have been given considerable attention in different studies for decades for their importance in L2 learning (Cortes, 2004; Ellis, 1996; Hyland, 2008; Wray, 2000). Though they appear to be usually unidiomatic or structurally incomplete, their functions are important building blocks in discourse, because they serve as a basis for more advanced language learning (Biber, 2009) and are important in measuring learner development (Staples, Egbert, Biber, & McClair, 2013). Hyland (2008) claims that the naturalness of these utterances in the language signals competence, and learning to use more frequent fixed phrases can contribute to gaining better communicative competence, while in contrast, the absence of such clusters could be indicative of lack of fluency.

In the early stages of L2 learning, these formulaic sequences are indispensable because novice learners rely heavily on them. Initially restored as whole units, they are then reanalyzed and reprocessed to form more flexible constructions at later stages of development (Ellis, 1996; Wray, 2002). Wray and Perkins (2000) explained this early stage as a useful entrance point for the learner, describing it as the phrase-

book approach providing a few pre-learned utterances. More advanced learners can likewise benefit by this formulaicity to gain command of a wide range of complex lexical units which native speakers process as prefabricated patterns (Howarth, 1998).

Some notable investigations on formulaic languages were that of Biber and Conrad (1999) and subsequent studies (Biber, 2009; Biber & Barbieri, 2007; Biber, Conrad, & Cortes, 2004; Conrad & Biber, 2005; Staples et al., 2013). As expected, prefabricated patterns have been actively examined in writing of English learners and/or native English speakers. Hyland (2008) explored these sequences in writing in four disciplinary variations. Chen and Baker (2010) investigated these occurrences from different groups of language users: academic writing of native and non-native speakers. In addition, most of Biber and colleague's studies are typically comparisons between spoken and written registers: conversation vs. academic prose (Biber & Conrad, 1999), university teaching vs. textbooks (Biber et al., 2004), and university spoken vs. written registers (Biber & Barbieri, 2007).

While several learner development studies compared formulaicity to other registers, discourse genres and disciplines, its use by language learners across different proficiency levels has been rarely investigated (Staples et al., 2013). Although Staples et al. (2013) examined the frequency, function, and degree of fixedness of the use of formulaic sequences by L2 learners across three proficiency levels in the TOEFL iBT<sup>1</sup>, the focus was on writing. Even with the importance of the formulaic sequences in developing communicative competence (Hyland, 2008) and in measuring learner development (Staples et al., 2013), few studies have explored the differences in L2 learners' formulaic spoken language use across different proficiency levels.

Therefore, this study seeks to fill the gap by exploring this formulaicity in speaking based on the scored oral responses of 58 L2 learners in a simulated OPIc test<sup>2</sup>, one of the most commonly taken tests by Korean university students before and/or after graduation. Then with the results of this study, L2 learners' conversation data can be examined to further explore the role of lexical bundles in their speaking development. To this end, this study currently aims to examine whether the frequency, structural forms or patterns, and functions of these formulaic utterances used by Korean L2 learners in a controlled environment vary across different proficiency levels. The results can be helpful for English speech development in order for learners to improve fluency in speaking and for teachers to effectively diagnose their students' speaking proficiency and should thus be integrated in the English teaching curricula.

## Literature Review

### Lexical Bundles: Basic Concept and Frequency Issues

Considerable attention has been given to phraseology in the analysis of recurrent word combinations for several decades. Wray (2000) enlisted about 50 metonym variations to describe aspects of formulaicity. The most common terms include chunks, fixed expressions, formulaic language, prefabricated patterns, ready-made utterances, and so on. Among the most that have achieved a wide recognition in applied linguistics are n-grams (Stubbs & Barth, 2003), multi-word expressions (MWEs; Siyanova-Chanturia & Martinez, 2014), and lexical bundles, popularly coined by Biber et al. (1999). Though the definitions somewhat vary, all refer to frequently occurring language sequences functioning as ready-made units, not requiring processing by the user (Wray, 2000). The present study adapts the term used by Biber et al. (1999), lexical bundles.

Technically, to be considered a lexical bundle, these formulaic sequences must possess at least these

---

<sup>1</sup> The Test of English as a Foreign Language—Internet based test (TOEFL iBT) was developed to measure a test taker's ability to use and understand English at the university level, including the ability to perform academic tasks using necessary English skills (ETS, 2007).

<sup>2</sup> Oral Proficiency Interview by Computer (OPIc) is developed to measure a test taker's ability to speak a language by interview, which provides a simulated conversation environment (ACTFL, 2017).

basic criteria. One is word combination, usually two-, three-, four- five- or six-word units. There is no prohibition on the number of combinations used, but four-word combinations, as Chen and Baker (2010) noted, are the most researched length for writing studies probably due to its manageability in size. Four-word bundles have a more distinct array in structure than three-word units and are more substantial than five-word units. Secondly, to be considered as bundles, these sequences should have a frequency cut-off. Initially, a multi-word sequence needs to occur 10 times per million words in a corpus to be considered a bundle. In later studies, it was increased to 200. In this study, the term lexical bundle is adapted to refer to the same recurrent word-string expressions despite the relatively lower scale in corpus size, number of participants, or the controlled environment. On methodological issues as this, Biber (1990) contended that even data containing only 1,000 words could still produce reliable results. Nevertheless, additional distributional requirements are needed. To exemplify, a bundle occurring only as few as 3 times in 50,400 words would have a normed rate of occurrence of 60 per million words (Biber, 2007, p. 268):

$$(3/50,400)*1,000,000 = 60 \text{ per million words}$$

To adjust for this inflated rate of occurrence, additional restrictions are imposed for the analysis of registers. Any bundle with a raw account of three must be distributed across three different texts, speakers or writers. This criterion has indeed helped adjust for differences in representation of the sub-corpora or registers. Its frequency and occurrence in different texts, usually in at least three to five texts (e.g., Biber & Babieri, 2007; Cortes, 2004) or 10% of texts (e.g., Hyland, 2008) help avoid idiosyncrasies from individual writers/speakers. Thus, despite having a meager number of lexical bundles in a corpus, these occurrences qualify for classification using the two taxonomies proposed by Conrad and Biber (1999) and Biber et al. (2004).

## Taxonomies of Lexical Bundle

Conrad and Biber (1999) classified two major taxonomies which have been widely relied upon in the studies of lexical bundles (Cortes, 2004; Hyland, 2008): structural characteristics and function in a discourse context. Though not usually complete structural units, “lexical bundles have strong grammar correlates” (Biber et al., 2004, p. 380) and “they do fall into groups with certain structural associations” (Conrad & Biber, 1999, p. 60).

### Structural taxonomy

The structural taxonomy presented by Biber et al. (2004) lists three main categories: lexical bundles that incorporate (1) verb phrase (VP) fragments, (2) dependent clause fragments, and (3) noun phrase (NP) and prepositional phrase (PP) fragments. Each category has specific subcategories providing more detailed distinctions and examples of the structural features of lexical bundles. Bundles like *you want me to* are constructed from verb and clause components, while bundles like *in the case of* are from noun phrase and prepositional phrase components.

The lexical bundles in the first subcategory involve seven types of VP fragments incorporated with (connector+) a first or second-person pronoun (*I'm not going to*), (connector+) a third-person pronoun (*and this is a*), a discourse marker (*I mean you know*), a VP with non-passive verb (*take a look at*), a VP with a passive verb (*can be used to*), yes-no question fragments (*are you going to*), and wh-question fragments (*what do you think*). The second subcategory are five lexical bundle types that are incorporated with dependent clause fragments with a first- or second-person pronoun (*I don't know if*), wh- (*when we get to*), if (*if you want to*), (verb or adjective+) a *to*-clause (*to be able to*), and *that* (*that there is a*). The last structural type is NP and PP fragments which are incorporated with five lexical bundle types, namely: (connector+) NP with *of* (*one of the things*), NP with other post-modifier (*a little bit about*), other NP expressions (*something like that*), PP expressions (*at the end*), and comparative expressions (*as far as the*).

Table 1 lists the details of the structural taxonomy of lexical bundles.

TABLE 1  
Structural Types of Lexical Bundles (Biber et al., 2004)

Type	Example
1. Lexical bundles that incorporate VP fragments	
1a. (connector +) 1 <sup>st</sup> /2 <sup>nd</sup> person pronoun + VP fragment	<i>You don't have to, I'm not going to, well I don't know</i>
1b. (connector +) 3 <sup>rd</sup> person pronoun + VP fragment	<i>It's going to be, and this is a, and this is a</i>
1c. Discourse marker + VP fragment	<i>I mean you know, you know it was, I mean I don't</i>
1d. Verb phrase (with non-passive verb)	<i>Is going to be, is one of the, have a lot of, take a look at</i> <i>Is based on the, can be used to, shown in figure N</i>
1e. Verb phrase with passive verb	<i>Are you going to, do you want to do, does that make sense</i>
1f. yes-no question fragments	<i>What do you think, how many of you, what does that mean</i>
1g. Wh-question fragments	
2. Lexical bundles that incorporate dependent clause fragments	
2a. 1 <sup>st</sup> /2 <sup>nd</sup> person pronoun + dependent clause fragment	<i>I want you to, I don't know if, I don't know why, you might want to</i>
2b. Wh-clause fragment	<i>What I want to, what's going to happen, when we get to</i>
2c. If-clause fragment	<i>if you want to, if you have a, if we look at</i>
2d. (verb/adjective+) to-clause fragment	<i>to be able to, to come up with, want to do is</i>
2e. that-clause fragment	<i>That there is a, that I want to, that this is a</i>
3. Lexical bundles that incorporate noun phrase and prepositional phrase fragments	
3a. (connector +) Noun phrase with of-phrase fragment	<i>One of the things, the end of the, a little bit of</i>
3b. Noun phrase with other post-modifier fragment	<i>A little bit about, those of you who, the way in which</i>
3c. Other noun phrase expressions	<i>A little bit more, or something like that, and stuff like that</i> <i>Of the things that, at the end of, at the same time</i>
3d. Prepositional phrase expressions	<i>As far as the, greater than or equal, as well as the</i>
3e. Comparative expressions	

## Functional taxonomy

For the functional categorization of the lexical bundles, Biber et al. (2004) distinguished three primary classifications of lexical bundles: stance expression, discourse organizers, and referential expressions. Stance bundles are groups of words that reveal attitude, judgment, perspectives of certainty or uncertainty, and proposition or ability. Two key meanings are conveyed: epistemic and attitude/modality. Epistemic stance bundles refer to the knowledge status of the information pertaining to certainty, uncertainty, or probability/possibility like *I don't know what*, or *I don't think so*. Attitudinal/modality stance bundles express speaker attitudes towards the actions or events. There are four types of attitudinal/modality bundles: focusing on desire (*I don't want to*), obligation/directive (*you don't have to*), intention/ prediction (*I was going to*), and ability (*it is possible to*). Stance bundles are also marked by whether they convey the stance in a personal or impersonal manner: the former overtly attribute the stance to the speaker or writer or addressee, and the latter expresses similarly without being attributed to an individual like *it is possible to* or *it is necessary to*.

Discourse organizers, as the name suggests, help to compose and structure the text itself, including topic introduction, elaboration and clarification. Topic introduction/focus bundles provide overt signals that a new topic is introduced or being focused on as *do you know what*. Topic elaboration/clarification bundles serve to add more information to a topic such as *what do you mean*.

Referential expressions are bundles that directly identify an entity, relate to a given attribute, or refer to time, place, and text. Referential expressions are classified into these four: identification/focus, imprecision, specification of attributes, and time/place/text references. Identification/focus bundles

identify an entity or part of it as noteworthy (*one of the most*). Imprecision bundles connect with previous discourse which is imprecisely expressed (*or something like that*). Specification of attributes bundles emphasize some particular attribute of the entity, like quantities (*per cent of the*), tangible attributes (*in the form of*), and intangible attributes (*in the absence of*). Time/place/text/ references denote those areas or are also multifunctional (*in the end of the*).

Special conversation functions occur in conversation: politeness routines (*thank you very much*), simple inquiries (*what are you doing*), and reporting clauses (*I said to him*).

A thorough list with examples is shown in Table 2.

TABLE 2  
*Functional Types of Lexical Bundles (Biber et al., 2004)*

Type	Example
<b>I. Stance Expressions</b>	
A. Epistemic Stance	
Personal	<i>I don't know what</i>
Impersonal	<i>the fact that the</i>
B. Attitudinal/Modality Stance	
B1) Desire	<i>if you want to</i>
B2) Obligation/Directive	
Personal	<i>you don't have to</i>
Impersonal	<i>it is necessary to</i>
B3) Intention/Prediction	
Personal	<i>I was going to</i>
Impersonal	<i>it's going to be</i>
B4) Ability	<i>it is possible to</i>
<b>II. Referential Expressions</b>	
A. Topic Introduction/Focus	<i>what do you think</i>
B. Topic Elaboration/Clarification	<i>on the other hand</i>
<b>III. Discourse Organizers</b>	
A. Identification/Focus	<i>one of the most</i>
B. Imprecision	<i>or something like that</i>
C. Specification of Attributes	
C1) Quantity Specification	<i>there's a lot of</i>
C2) Tangible Framing Attributes	<i>the size of the</i>
C3) Intangible Framing Attributes	<i>in the case of</i>
D. Time/Place/Text Reference	
D1) Place reference	<i>in the United States</i>
D2) Time reference	<i>at the same time</i>
D3) Text deixis	<i>as shown in figure</i>
D4) Multi-functional reference	<i>in the middle of</i>
<b>IV. Special Conversational Functions</b>	
A. Politeness	<i>thank you very much</i> <i>what are you doing?</i>
B. Simple Inquiry	<i>I said to him/her</i>
C. Reporting	

## Studies on Lexical Bundles

While few studies have explored the use of bundles in the speech of L2 learners across different proficiency levels, Biber et al. (2013) investigated these bundles in the written responses of 480 TOEFL iBT test-takers across three proficiency levels. As reported, there was not much variability in the scores of different proficiency levels in terms of functional use or the degree of fixedness of the bundles. The study claimed that there may be a developmental sequence for some aspects of formulaic language use and that the test takers may not have yet developed the skills necessary, e.g., for referring to abstract entities, a common function within academic writing.

Chen and Baker's (2010) study contrasting the use of lexical bundles in the writings of three groups (student native speakers, student non-native speakers, and native expert writers) found a gap in the use of

lexical bundles between native expert academic writing and university student writing (native and non-native alike). Structurally and functionally, the use of lexical bundles in non-native student essays was surprisingly similar to that of native student essays. They both contained more VP-based bundles and discourse organizers than native expert writings, which was a sign of immature writing. Native professional writers, on the other hand, exhibited a wider range of NP-based bundles and referential markers. The test-takers in Biber et al. (2013) used even proportionally fewer referential bundles than the two groups of student writers in Chen and Baker's (2010).

In the Korean context, Kim (2013) compared the four-word lexical bundles of Korean college students' English essays (258 essays with 63,475 words) with those used by native students' (784 essays with 326,320 words). The results specified that Korean learners underused the lexical bundles that their native counterparts most frequently used, and that the former exhibited a lavish use of the pronoun "I". This was reportedly due to the lack of Korean learners' proficiency and proper literacy. Regarding the structures and functions, Korean learners predominantly used VP-based bundles and stance expressions, while their native counterparts did not have a serious domination of certain types of structural and functional bundles. Similarly, Yoon and Choi (2015) examined the four-word lexical bundles in Korean university students' EFL compositions and equated the register and use with those from a native speakers' corpus. Yielding comparable results to Kim's (2013), the Korean university students, in contrast to the native students, heavily relied on the bundles widely used in speech registers, such as bundles containing personal pronouns and contractions and expressing stance. This raises the question of whether Korean L2 learners use those bundles in spoken registers when actually speaking, and if so, what kinds of bundles they use. The researchers further suggested that adult writers need to shift from highly repetitive use of personal pronouns in the limited set of phrases and patterns they feel safe to use in writing to proper exposure or explicit learning of genre features appropriate to writing, specifically that of argumentative writing.

While the former studies regarding Korean learners' use of lexical bundles focused mainly on writing, this study aimed to fill the gap by identifying and analyzing the lexical bundles from a spoken register of Korean English language learners.

The study specifically aims to address the following questions:

1. What differences, if any, are there in the frequency of lexical bundle use across proficiency levels?
2. What differences, if any, are there in the structural patterns of lexical bundle use across proficiency levels?
3. What differences, if any, are there in the functional patterns of lexical bundle use across proficiency levels?

## **Methodology**

### **Participants**

The speech production came from 58 adult Korean EFL students (16-40 years old) attending OPIc classes in an English language academy in central Seoul, Korea. They voluntarily registered to the academy to improve their English proficiency. The diagnostic test results showed little difference in English proficiency by age. Their oral proficiency, pre-determined by the academy, were divided into three levels: 20 novice, 19 intermediate, and 19 advanced learners. There were in total 36 female and 22 male participants.

TABLE 3  
*Participants Profile*

	Novice	Intermediate	Advanced	Total
Female	14	10	12	36
Male	6	9	7	22
Total	20	19	19	58

## Instruments

One of the standardized speaking tests most frequently administered in Korea is the Oral Proficiency Interview by computer (OPIc), a test that was developed to measure a test taker's ability to speak a language by interviewing him or her, to help the test taker speak in a simulated conversation environment (ACTFL, 2017). Many Korean university students and those wanting to work for a company have taken the OPIc to show their English speaking proficiency.

The participants in their OPIc classes had to take simulated OPIc tests monthly provided by the language academy, for one of which every student has two minutes to state their answer to each of the following six questions:

- a. Please tell me about yourself.
- b. Tell me about your house. Please describe in detail.
- c. Tell me about your typical day from start to finish.
- d. What kind of technology do you use the most these days: a cellular phone, a laptop computer, or any handheld device? How do you use it these days?
- e. How has technology changed from past to present? What was the technology you used in the past? How was it different from the technology you use today?
- f. I enjoy travelling to many countries. Please ask me three or four questions about the vacation I have recently taken.

## Data Collection Procedure

The participants took an OPIc simulation test every month and their monthly test production data were saved on the academy website. The test production data of the first month were extracted in order to exclude the repetition effect caused by monthly tests and/or instruction effects.

The participants' oral production data from the first simulated OPIc test were grouped by their pre-determined levels (Novice, Intermediate, and Advanced). Each data set was transcribed and then coded with their corresponding structural (Table 1) and functional (Table 2) tags.

## Data Analysis Procedure

### Transcription

To facilitate transcription of the recordings, a utility software called *Voice Walker 2.0* was used. This software offers some degree of accuracy and more reliable transcription, and is indispensable for deciphering rapid speech, especially for recognizing words when two speakers talk simultaneously.

### Coding and tagging

Two computer software programs were operated to identify word bundles, tag codes, and assist in analyzing the corpus. *UAM Corpus Tool 3.3*, a set of text annotation tool, was used manually and semi-automatically for convenience in code tagging and retrieval. *AntConc 3.4*, a freeware concordance, was

vital specifically in identifying one or multi-unit N-grams, with each sequence stored separately.

### **Operationalization**

To qualify as a lexical bundle in this study, the four-unit sequence must be used by multiple speakers from different proficiency levels, and not simply as an individual style. Frequency was analyzed with *AntConc 3.4*. Lexical bundles, to note, do not need to be complete syntactic or semantic units. They may link two syntactic units such as *I think what you* and are not separated by clause and phrase boundaries. Contractions are considered two words. In the bundles, contractions such as *don't* in *I don't know* were considered two separate words and counted similarly as with their corresponding non-contracted forms, *do* and *not*.

The refinement procedure of the initially identified four-unit bundles discarded the ones that include repetitions (e.g., *I I, the the*), hesitation items (e.g., *uhm, uh*), topic-specific bundles (e.g., *my English name is*) and context-indicating bundles (e.g., *live in South Korea*).

### **Classification of formulaic language**

In this study, the formulaic language patterns were classified according to Biber et al.'s (2004) structural (Table 1) and functional (Table 2) types. Although bundles had incomplete syntactic structures, they are classifiable by their structural associations. Three structural types (with VP fragments, with dependent clause fragments, and with NP and PP fragments) were subcategorized into 17 types according to the specific structural patterns each bundle incorporates. For the functional classification, the bundles were placed into four groups unified by similar discourse functions: stance expressions, discourse organizers, referential expressions, and special conversation functions. These four function groups were subcategorized into 11 types according to the specific functions each bundle performs. The frequency, structural types, and functional types were compared among the three groups of different speaking proficiency levels.

## **Results and Discussion**

### **Frequency of Lexical Bundle Use across Proficiency Levels**

The refining process which eliminates the bundles with traces of repetitions, hesitation items, topic-specific and context-indicating bundles left a set of 158 from the 244 raw bundles, divided into 21 for Novice, 52 for Intermediate, and 85 for Advanced. As seen in Table 4, the 21 lexical bundles comprise 2.44% of every four words of the 3,447 total word count of the Novice level. The Intermediate level had 52 bundles, 1.74% out of 11,980 words, while the Advanced level had 85, or 1.71% out of 19,935 words. In total, bundles used by the three levels was 158, 1.79% out of 35,362 words, and were normalized to the cut off of one million. The results showed that 774 was the average normed rate of lexical bundles for the Novice per million words. The Intermediate level had a rate of 223 per million words while the Advanced level had 191.



TABLE 4  
Lexical Bundle Frequency and Normalized Rate

	Novice	Intermediate	Advanced	Total
No. of Words used in the OPIC test	3,447	11,980	19,935	35,362
Lexical bundles token frequency after refinement/ (% of the total number of words/level)	21 (60.92%)	52 (43.41%)	85 (42.64%)	158 (44.68%)
Average Normed rate of lexical bundles / 1 million words	774	223	190	

The results demonstrate that despite the amount of words and the number of raw bundles used, the participants with the highest proficiency level used relatively fewer bundle tokens than the other lower groups in terms of percentage and normalized rate per one million. This outcome illustrates a pattern of bundle use by different proficiency levels. The higher the level, the lower the dependence on the use of lexical bundles. This finding supports the previous studies showing that formulaic language was an essential device for lower level learners (Chen & Baker, 2010; Kim, 2013; Yoon & Choi, 2015). Ellis (1996) asserted that novice language learners relied heavily on formulaic sequences, which were initially stored as whole units, but then reanalyzed and reprocessed to form more flexible constructions at later stages of development. Nevertheless, the number of lexical bundles and the variations manifested in the succeeding sections seemed to move in an upward trend from lower to higher levels of proficiency.

### Structural Patterns of Lexical Bundle Use Across Proficiency Levels

Premised on the structural framework of Biber et al. (2004) for lexical bundles, Fig. 1 below represents the general distribution of bundles within different structural categories, and Table 5 shows the distribution of bundles across subcategories of grammatical structure.

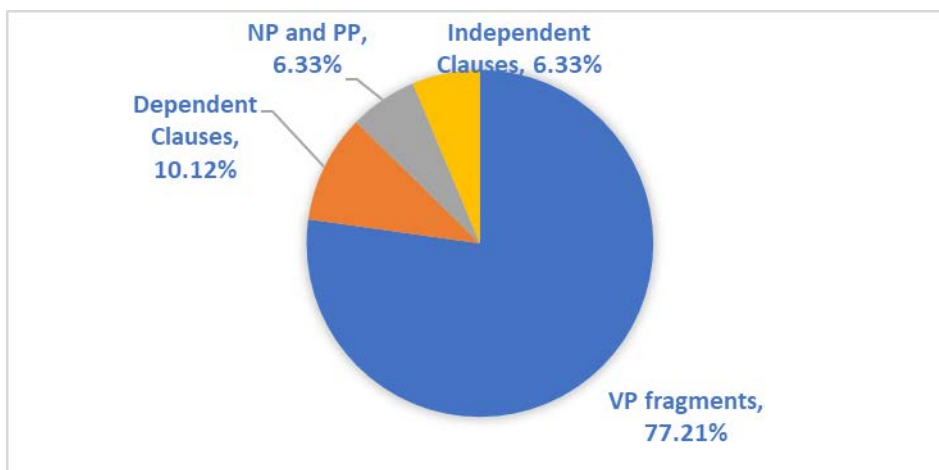


Figure 1. Distribution of bundles across different structural types.

There was an extensive difference in the use of structural categories (Fig. 1). VP fragments consisted of 77.21% of bundles, followed by dependent clause fragments at 10.12%, and then NP and PP fragments (6.33%) and independent clause (6.33%). In contrast to Table 1, independent clauses were identified. Korean learners seemed to produce short sentences that incorporated only four words, which was not so frequently observed previously as was in this study. Additional studies are needed to explore whether other L2 learners or native English speakers would also produce four-word sentences in a similar situation as in this study.

Most of the lexical bundles used by the three levels had a verb attachment in the bundles. This finding was consistent with Conrad and Biber’s (2005) study that conversation tended to have more verbs, more

personal pronouns, and more questions, and that structures typical of conversation were used, such as *I don't know* and other fragments using personal pronouns specifically incorporating first-person reference. The first-person pronoun *I* occurred 99 times (62.66%) in the bundles, while *my*, *we*, and *you* appeared 19 times (7.4%). The greatest portion of each proficiency level was for VP fragments that incorporated the first- or second-person pronoun (see Table 5).

Similar Korean studies by Kim (2013) and Yoon and Choi (2015) pointed out to a similar pattern. This, according to the former, is due to the lack of Korean learners' proficiency and proper literacy. Chen and Baker (2010) also postulated that participants with higher proficiency, e.g., a near-native-like level, exhibited a wider range of NP-based bundles and referential markers. Correspondingly, participants with lower proficiency in this study showed limited use of NP and PP-based bundles in contrast to the diverse and even wider distribution in the other two groups.

TABLE 5  
*Structural Types of Lexical Bundles*

<i>Structural Types</i>	Novice	Intermediate	Advanced	Total	Example
<b>PHRASAL</b>					
<b>1. LB with VP fragments</b>	<b>18 (11.39%)</b>	<b>39 (24.68%)</b>	<b>65 (41.14%)</b>	<b>122 (77.21%)</b>	
1a. (connector +) 1 <sup>st</sup> /2 <sup>nd</sup> person pronoun + VP Fragment	14 (8.86%)	26 (16.50%)	43 (27.22%)		<i>I don't know, I don't understand</i>
1b. (connector +) 3 <sup>rd</sup> person pronoun + VP Fragment		5 (3.17%)	5 (3.17%)		<i>it's like a , it's easy to</i>
1c. Discourse marker + VP fragment			2 (1.27%)		<i>And after that, After that I go</i>
1d. Verb phrase (with non-passive verb)	3 (1.90%)	7 (4.43%)	12 (7.60%)		<i>have a lot of, go back to my</i>
1f. <i>yes-no</i> question fragments			1 (0.63%)		<i>Could you tell me</i>
1g. Wh-question fragments	1 (0.63%)	1 (0.63%)	2 (1.27%)		<i>What is the best, where did you go</i>
<b>2. LB with dependent clause fragments</b>	<b>1 (0.63%)</b>	<b>7 (4.43%)</b>	<b>8 (8.86%)</b>	<b>16 (10.12%)</b>	
2b. Wh-clause fragment	1 (0.63%)	5 (3.17%)	6 (3.80%)		<i>When I was young, when I want to</i>
2c. <i>If</i> -clause fragment		1 (0.63%)	1 (0.63%)		<i>If you want to</i>
2e. <i>that</i> clause fragment		1 (0.63%)	1 (0.63%)		<i>That's why I</i>
<b>3. LB with NP and PP fragments</b>	<b>1 (0.63%)</b>	<b>4 (2.53%)</b>	<b>5 (3.17%)</b>	<b>10 (6.33%)</b>	
3a. (connector +) Noun phrase with <i>of</i> -phrase fragment		1 (0.63%)	1 (0.63%)		<i>a lot of things</i>
3b. Noun phrase with other post-modifier fragment			1 (0.63%)		<i>A little bit and</i>
3c. Other noun phrase expressions	1 (0.63%)	1 (0.63%)	1 (0.63%)		<i>But these days I</i>
3d. Prepositional phrase expressions		2 (1.27%)	2 (1.27%)		<i>in my case</i>
<b>CLAUSAL</b>					
Independent clause	1 (0.63%)	2 (1.27%)	7 (4.43%)	10 (6.33%)	<i>We have a lot</i>
<b>Total</b>	<b>21 (13.30%)</b>	<b>52 (32.91%)</b>	<b>85 (53.80%)</b>	<b>158 (100%)</b>	

Another interesting feature to note in the dense use of VP is the restricted use of verb variations. Table 6 lists the limited use of verbs by the different levels. Based on the list, the participants seemed to have

not yet widened their vocabulary command. The most common verb that was proportionally used by all three levels were the verbs *do*, *have*, and *go*. *Do* was mostly accompanied by the negative marker *not* (e.g., *I don't + V*) such as *I don't want* or *I don't know*. *Have* was mainly incorporated with a connector and infinitive *to* such as *So I have to* or *Because I have to*. *Go* was also linked with the infinitive *to*. Aside from *have* and *go*, some other verbs on the list were also extensively used with the infinitive *to* such as *like to*, *want to* and *need to*.

In sum, as proficiency level increases, more verb variations are included. This is consistent with previous studies (Chen & Baker, 2010; Kim, 2013) positing that limited use of lexical bundles was caused by L2 learners' lack of proficiency. It is possible that the participants' limited use of the verbs might be due to a lack of opportunities to teach vocabulary explicitly and separately in Korean English classes. It is recommended that teachers teach various words and how to use them in diverse contexts. Another possibility is that the relatively old participants in this study tended not to take risks in using new words, which should be further explored with more participants of the similar age.

TABLE 6  
*Verb Variations in LB with VP Fragments*

	Novice	Intermediate	Advanced	Total
<i>do</i>	8	12	13	33
<i>have</i>	4	9	12	25
<i>go</i>	1	2	10	13
<i>want</i>	2	2	5	9
<i>know</i>	1	3	3	7
<i>use</i>		3	4	7
<i>can</i>	1	2	4	7
<i>like</i>	2	1	2	5
<i>take</i>	2	2		4
<i>think</i>		2	2	4
<i>need</i>		1	1	2
<i>live</i>			2	2
<i>tell</i>			2	2
<i>understand</i>	2			2
<i>plan</i>			1	1
<i>is</i>		9	11	20
<i>am</i>		4	3	7

## Functional Patterns of Lexical Bundle Use Across Proficiency Levels

### Stance bundles

Bundles expressing the speaker's stance were overall more frequent than organizing discourse or expressing reference, which showed similar results to Kim's (2013) Korean English learners rather than Chen and Baker's (2010) Chinese English learners. All the stance bundles used were for attitudinal/evaluative stance, which functioned to express the participants' opinions about the topic (Staples et al., 2013). Covering more than the half of the bundles (Fig. 2 and Table 7), stance expressions consisted of 55.06%, much more than discourse organizers (23.41%) and referential expressions (18.35%). Stance expressions were greatly employed by the participants to convey their personal expressions.

Of the two kinds of stance expressions, attitudinal/modality stance were extensively used. Attitude modality stance bundles (38.61%) were proportionately distributed in the three sub-categories of desire, obligation/directive, and ability. The first attitudinal stance focused on the participants' desire (12.66%) such as *I really want to* or *I don't like to*. The second conveyed the speaker's attitude about propositions (obligation/directive), for example, *so I have to*, or *don't need to*. If the speaker's judgment to do something was involved, they were grouped under ability (e.g., *so we can use* or *is very easy to*). The Novice did not have any tokens of obligation/directive stance. A probable explanation is that Novice

learners might not have yet developed the skills necessary to create this aspect of formulaic language, as Biber et al. (2013) proposed.

On the other hand, epistemic stance, which expressed the speaker’s evaluation in terms of certainty or uncertainty, consisted of 16.46%. *I don’t know* and *I don’t understand* were the most common epistemic bundles in the study, especially for lower proficiency levels.

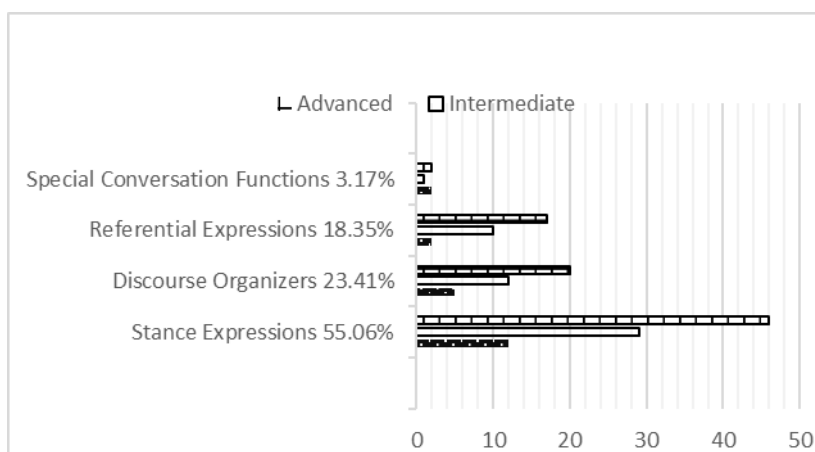


Figure 2. Distribution of lexical bundles across functional types

TABLE 7  
Functional Types of Lexical Bundles

	Novice	Intermediate	Advanced	Total
<b>I. Stance Expressions</b>	<b>12 (7.60%)</b>	<b>29 (18.35%)</b>	<b>46 (29.11%)</b>	<b>87 (55.06%)</b>
A. Epistemic Stance	5 (3.17%)	11 (6.96%)	10 (6.33%)	26 (16.46%)
Personal	3 (1.90%)	8 (8.86%)	7 (4.43%)	
Impersonal	2 (1.27%)	3 (1.90%)	3 (1.90%)	
B. Attitudinal/Modality Stance	7 (4.43%)	18 (11.39%)	36 (22.79%)	61 (38.61%)
B1) Desire	4 (2.53%)	5 (3.17%)	11 (6.96%)	20 (12.66%)
B2) Obligation/Directive		6 (3.80%)	8 (8.86%)	14 (8.86%)
Personal		5 (3.17%)	7 (4.43%)	
Impersonal		1 (0.63%)	1 (0.63%)	
B3) Intention/Prediction	2 (1.27%)	5 (3.17%)	10 (6.33%)	17 (10.76%)
Personal	2 (1.27%)	5 (3.17%)	8 (8.86%)	
Impersonal			2 (1.27%)	
B4) Ability	1 (0.63%)	2 (1.27%)	7 (4.43%)	10 (6.33%)
<b>II. Discourse Organizers</b>	<b>5 (3.17%)</b>	<b>12 (7.60%)</b>	<b>20 (12.66%)</b>	<b>37 (23.41%)</b>
A. Topic Introduction/Focus	1 (0.63%)	3 (1.90%)	7 (4.43%)	11 (6.96%)
B. Topic Elaboration/Clarification	4 (2.53%)	9 (5.70%)	13 (8.23%)	26 (16.46%)
<b>III. Referential Expressions</b>	<b>2 (1.27%)</b>	<b>10 (6.33%)</b>	<b>17 (10.76%)</b>	<b>29 (18.35%)</b>
A. Identification/Focus		3 (1.90%)	3 (1.90%)	6 (3.80%)
C. Specification of Attributes	2 (1.27%)	6 (3.80%)	13 (8.23%)	21 (13.29%)
C1) Quantity Specification	2 (1.27%)	5 (3.17%)	9 (5.70%)	
C3) Intangible Framing Attributes		1 (0.63%)	4 (2.53%)	
D. Time Reference		1 (0.63%)	1 (0.63%)	2 (1.27%)
<b>IV. Special Conversational Functions</b>	<b>2 (1.27%)</b>	<b>1 (0.63%)</b>	<b>2 (1.27%)</b>	<b>5 (3.17%)</b>
A. Politeness			1 (0.63%)	
B. Simple Inquiry	1 (0.63%)	1 (0.63%)	2 (1.27%)	
<b>Total</b>	<b>21 (13.30%)</b>	<b>52 (32.91%)</b>	<b>85 (53.80%)</b>	<b>158 (100%)</b>

### Discourse organizers

Following the stance bundles in frequency were discourse organizers, which were used to structure texts. Of the two types, topic elaboration/clarification (e.g., *that’s why I* and *in my case I*) outnumbered

topic introduction/focus (e.g., *when I was young* and *you tell me about*) with 16.46% to 6.96%. This can be explained by the context in which the output was extracted. The topic was overtly introduced and prompted, and thus no need for reiteration. Topic elaboration and clarification were more frequent for the reason that the speakers were asked to expound the topic. Similar to that of Staples et al. (2013), a majority of the bundles were related to the specific topics used in the prompts, and test takers used very few bundles that referenced the textual context in a general way, rather than framing a specific topic within the production. Therefore, conversational data in a more natural environment than the simulated conversation in this study should be also examined.

### Referential expressions

Referential expressions were characterized by the function of attribute specification. Reaching 18.35% in total for referential expressions, the most common type of attribute specification was the quantifying specification (e.g., *have a lot of* and *there are so many*). Some identification/focus expressions (e.g., *yeah that's it* and *it is very*) were also used (3.80%), along with a few samples of time reference (1.27%; e.g., *at that time*).

### Special conversation functions

Although not as frequently used as other functions, a few special conversation bundles were identified: for politeness, *could you tell me* and for simple inquiry, *what is the best*. Only 3.17% consisted of special conversation functions, again in accordance with Biber et al.'s (2013) proposition that there may be a developmental sequence in formulating sequences and that some speakers, especially those with lower proficiency, have not yet developed the necessary skills. Perhaps politeness appears later in the developmental sequence, as it was not yet used among the lower proficiency levels. This appeared only in the advanced level.

Since the participants answered questions given by computer, they might not have used any polite expressions. Facing a physical interviewer with different ranks and roles (professors to students, friends to friends, or students to professors), the participants might have engaged in different patterns and frequencies of special conversation bundles. Again, further studies are needed with oral production data in a natural conversation environment in addition to a standardized test.

Comparing the three levels of proficiency, there seemed to be not much difference in the trend for structural and functional patterns. However, as the level progressed, the variation and frequency of lexical bundles also increased. Korean L2 learners in this study showed a developmental sequence in acquiring lexical bundle skills as Staples et al. (2013) proposed, and a heavy reliance on formulaic patterns of low-level non-native speakers as in Myles et al.'s (1998) study. The participants are expected to move toward self-constructed language as proficiency increases (Myles et al., 1998), and teachers are encouraged to help L2 learners begin with memorization and one-to-one form-function mapping so that they can move in the direction of more native-like production of language as Ellis (2006) proposed.

## Conclusion

The present study explored the characteristics of four-unit lexical bundles in the speech production of Korean L2 learners taking a mock OPIc test. In doing so, the study underwent conscientious procedures in the identification and analysis of lexical bundles. Since few studies have examined L2 learners' spoken data and other studies comparing L2 learners' writing test production data, this study aimed to explore the frequency, structural patterns, and functional patterns of L2 learners' lexical bundle use in speaking. The present study seemed to confirm the findings of the previous studies, where there might be an upward trajectory in the language development as proficiency increases (Ellis, 2006; Myles et al., 1998). It should

be noted that the participants used lexical bundles in a limited manner similarly across the three proficiency levels, which was less frequent and less diverse in terms of structural and functional patterns when they had to produce oral data than when they had to produce written data.

The findings in this study carry important implications for English speaking test takers and English curricula developers, specifically in speaking and vocabulary learning. The present study revealed that adult Korean L2 learners lack vocabulary and hedges in their speech production. The underuse of hedging devices prompted learners to be categorical and to over-generalize. Without explicit learning exposure to linguistic features (diverse word choice and hedging), Yoon and Choi (2015) reasoned that learners resort to using excessive personal pronouns and repetitive usage of a limited set of phrases and patterns they feel safe using in their production. Proper guidance, for instance, in making fewer first-person references such as *I think* and *I want*, and more hedging devices for signifying probability can be practiced instead, such as *would*, *may*, and *likely to*. The most overused bundle for the three proficiency groups were *I don't know* and *I don't understand*, lingo popular on the Internet as *idk* and *idu*. This study asserts that this overt manifestation shows deficiency in linguistic skills and lower proficiency level of the learners. *Idk* and *Idu* can be indirectly stated as *I have no idea* or *I have no knowledge* to avoid re-using and abusing these common expressions. Regarding the over-usage of the restricted and general verbs such as *have*, one can employ other similar expressions such as *possess* or *own*. Then again, teachers should plan the lessons to help L2 learners reach a proficiency level in order to produce a more sophisticated communicative pattern. This emphasizes that English learners should be given more opportunities for familiarizing themselves with expressions that go beyond the usual, ordinary, and general by engaging in formal training including this specific linguistic feature.

Limitations of the study include the relatively small number of participants and smaller corpus size. Nevertheless, Biber (1990) postulated that a meager amount of 1,000 words of the data could still produce reliable results. Tribble (1977), as additional support for Biber, claimed that as long as the register is specialized, a small corpus would still be adequate to provide meaningful results.

Finally, pertaining to the implications and the limitations mentioned, this study suggests that future studies can investigate similar types of research on a larger scale and/or in a more natural conversation environment. Also, future researchers can explore the effect of formal training on the use of lexical bundles in reference to the results of this study. This paper hopes to contribute to the existing knowledge of lexical bundles and understand its characteristics for language teaching and learning.

## The Authors

*Maribel N. Zipagan* is a Ph.D. candidate in the Dept. of English Language and Literature of Sejong University in Seoul, Korea. She has been teaching English to Korean English learners for more than 10 years in Korea. Currently, she is interested in English learners' errors as well as the use of formulaic sequences in order to help them improve their English.

Department of English Language and Literature  
Sejong University  
Seoul, 143-747, Korea  
Tel: +82-2-3408-3302  
Email address: zipbelle@gmail.com

*Kyoung Rang Lee* (corresponding author) is Associate Professor at Sejong University, Seoul, Korea. She is interested in individual differences in teaching and learning English, including strategies of both teachers and students. Currently, she is devoted to better understanding Korean English learners' lexical bundle use in terms of their sociolinguistic backgrounds.

Department of English Language and Literature

Sejong University  
Seoul, 143-747, Korea  
Tel: +82-2-3408-3118  
Email address: kranglee@sejong.ac.kr

## References

- American Council on the Teaching of Foreign Languages [ACTFL]. (2017). Oral proficiency assessments (Including OPI & OPIc). Retrieved September 14, 2017 <https://www.actfl.org/professional-development/assessments-the-actfl-testing-office/oral-proficiency-assessments-including-opi-opic>
- Biber, D. (1990). Methodological issues regarding corpus-based analyses of linguistic variation. *Literary and Linguistic Computing*, 5, 257-269.
- Biber, D. (2009). A corpus-driven approach to formulaic language: Multi-word patterns in speech and writing. *International Journal of Corpus Linguistics*, 14, 275-311.
- Biber, D., & Barbieri, F. (2007). Lexical bundles in university spoken and written registers. *ESP*, 26, 263-286.
- Biber, D., & Conrad, S. (1999). Lexical bundles in conversation and academic prose. In H. Hasselgard, & S. Oksefjell (Eds.), *Out of corpora: Studies in honour of Stig Johansson* (pp. 181-190). Amsterdam: Rodopi.
- Biber, D., Conrad, S., & Cortes, V. (2004). *If you look at...: Lexical bundles in university teaching and textbooks*. *Applied Linguistics*, 25, 371-405.
- Chen, Y., & Baker, P. (2010). Lexical bundles in L1 and L2 academic writing. *Language Learning & Technology*, 14(2), 30-49.
- Conrad, S., & Biber, D. (2004). The frequency and use of lexical bundles in conversation and academic prose. *Lexicographia*, 20, 56-71.
- Cortes, V. (2004). Lexical bundles in published and student disciplinary writing: Examples from history and biology. *ESP*, 23(4), 397-423.
- Ellis, N. (1996). Sequencing in SLA: Phonological memory, chunking, and points of order. *SSLA*, 18, 91-126.
- Ellis, N. (2006). Cognitive perspectives on SLA. The associative cognitive CREED. *Association Internationale de Linguistique Appliquée Review*, 19, 100-121.
- Educational Testing System [ETS]. (2017). What is TOEFL iBT? Retrieved September 14, 2017 from <http://www.ets.org/toefl/ibt/about>
- Howarth, P. (1998). Phraseology and second language proficiency. *Applied Linguistics*, 19(1), 24-44.
- Hyland, K. (2008). As can be seen: Lexical bundles and disciplinary variation. *ESP*, 27, 4-21.
- Kim, J. (2013). Lexical bundles in Korean college students' English essays: A corpus-based comparative study. *English Language and Literature Teaching*, 19(3), 157-179.
- Myles, F., Hooper, R., & Mitchell, R. (1998). Rote or rule? Exploring the rule of formulaic language in classroom foreign language learning. *Language Learning*, 48(3), 323-363.
- Siyanova-Chanturia, A., & Martinez, R. (2015). Idiom principle revisited. *Applied Linguistics*, 36(5), 549-569.
- Staples, S., Egbert, J., Biber, D., & McClair, A. (2013). Formulaic sequences and EAP writing development: Lexical bundles in the TOEFL iBT writing section. *Journal of English for Academic Purposes*, 12, 214-225.
- Stubbs, M., & Barth, I. (2003). Using recurrent phrases as text-type discriminators. A quantitative method and some findings. *Functions of Language*, 10, 1, 61-104.
- Tribble, C. (1997). Improvising corpora for ELT: Quick and dirty ways of developing corpora for language teaching. In B. Lewsadowska-Tomaszczyk, & J. Melia (Eds.), *PALC '97 (Practical Applications in Language Corpora)* (pp. 106-117). Frankfurt: Peter Lang.
- Wray, A. (2000). Formulaic sequences in second language teaching: Principle and practice. *Applied*

*Linguistics*, 21(4), 463-489.

Wray, A., & Perkins, M. (2000). The functions of formulaic language: An integrated model. *Language and Communication*, 20, 1-28.

Yoon, C., & Choi, J. (2015). Lexical bundles in Korean university students' EFL compositions: A comparative study of register and use. *Modern English Education*, 16(3), 47-69.