THE JOURNAL OF ASIA TEFL
Vol. 17, No. 2, Summer 2020, 479-492
http://dx.doi.org/10.18823/asiatefl.2020.17.2.11.479

AsiaTEFL

The Journal of Asia TEFL

http://journal.asiatefl.org/

e-ISSN 2466-1511

Structural Analysis of Lexical-Bundles in Uncontrolled Environment of Spoken Discourse: The Same Heritage Language Groups*

Julian Chung
Sejong University, Korea

Kyoung Rang Lee
Sejong University, Korea

This study explores the lexical bundles in the speech of Korean English language users who share the same heritage language and culture, Korean, but with different proficiency in English: Korean-Americans whose first language is English (2nd-generation Korean-Americans; G2KAs), bilingual Korean-Americans who started to learn English at an early age (1.5-generation Korean-Americans; G1.5KAs), and Koreans who learned English in Korea (L2Ks). One of the big differences between native (NSs) and non-native speakers (NNSs) is that NNSs use verb-related lexical bundles dominantly unlike NSs. Thus, the structural patterns of lexical bundle uses in addition to verb uses were examined. The results were consistent to previous studies showing increasing use of lexical bundles with increase in English proficiency; NSs (G2KAs) used the lexical bundles more than G1.5KAs and L2Ks. However, in contrast to previous studies, G2KAs mostly used VP-related lexical bundles, G1.5KAs used lexical bundles in the widest range, and L2Ks used the most lexical verbs. The paper elaborates the details of the results and suggests further studies including the new group of English users, G1.5KAs, to help understand a possible influence of their heritage language and culture on English speaking.

Keywords: uncontrolled environment of spoken discourse, Korean-Americans, Korean English learners, Generation 1.5, lexical bundles, structural patterns.

Introduction

Speech reveals a great deal of information on the progression of a new or second language (L2). One distinct feature in the language of L2 learners is the inclusion of formulaic speeches. For decades, the use of formulaic speech embedded in the language of L2 learners and even native speakers (NSs) has been considered greatly by linguists due to its distinct function in language learning (Cortes, 2004; Ellis, 1996; Hyland, 2008; Wray, 2000). These formulaic sequences function as building blocks in spoken/written discourse towards fluency construction. If properly used, these recurrent and fixed phrases are critical in the acquisition of native-like competence and are vital in gaining better communicative competence (Biber, 2009; Hyland, 2008).

* This work was supported by the Ministry of Education of the Republic of Korea and the National Research Foundation of Korea (NRF-2019S1A5A2A01036649).

Previous studies have focused on these formulaic speeches and have used different labels. Wray (2000) enlisted about 50 metonyms such as *formulaic language, fixed expressions, prefabricated patterns, language chunks, ready-made utterances,* etc. Though having different brands and somewhat varieties in definition, Wray (2000) refer to all these frequently occurring language sequences functioning as ready-made units, not requiring processing by the user. In linguistics, some of these names are more recognized and received greater attention as *n-grams* (Stubbs & Barth, 2003) and *multi-word expressions* (MWEs; Siyanova-Chanturia & Martinez, 2015). The current study adapts the term, *lexical bundles*, popularly coined by Biber and Conrad (1999).

The study of lexical bundles includes written and spoken registers, at times the contrast of both (Biber, 2009; Biber & Barbieri, 2007; Conrad & Biber, 2004). On occasion, different language groups are being compared in different registers. Chen and Baker (2010) differentiated the lexical bundles of L1 and L2 academic writing, and Cortes (2004) in published and student disciplinary writing. Kim (2013) and Shin et al. (2019) compared the lexical bundles in Korean college students' essays with those in L1 students' essays. Staples et al. (2013) compared the use of lexical bundles in L2 learners' written test. Even though lexical bundles mostly in written register have been examined, those in spoken register have also gained attention. De Cock (2004) investigated the recurrent sequence diversity of native and non-native speeches. Kwon and Lee (2014) examined lexical bundles in teacher talk between non-native and native English instructors. Zipagan and Lee (2018) explored the role of lexical bundles in L2 learners' spoken language development in a controlled environment across three different proficiency levels. However, lexical bundles in an uncontrolled environment of spoken discourse have rarely been investigated.

The current study thus aims to contribute to the existing knowledge of the diversity of the lexical bundles between native (NSs) and non-native speakers (NNSs), but with the same heritage language, by investigating the structural differences of bundles among these three different speaker groups. The use of these lexical bundles in speech could affect the way they are perceived (Nekrasova, 2009), thus this study contributes to the field of language learning specifically in spoken discourse. The findings of this study are expected to direct L2 learner-speakers on the recognition and various usages of lexical bundles and verbs to raise their awareness in the organization of speech production of three different groups.

The motivation behind analyzing three groups is to enrich and diversify our understanding L2 acquisition. Language studies have conventionally compared two groups of speakers, NSs and NNSs, or groups of different proficiency within L2 learners, yet this study is one of the first attempts to examine another group of speakers who do not fit on either ends of the spectrum: 1.5-generation Korean-Americans (G1.5KAs) [1], those who acquired English after arriving in the US at a young age. This group of speakers shares the same heritage language and culture of both Korean English learners (L2Ks) and native English speaking Korean-Americans (2nd-generation Korean-Americans; G2KAs), and at the same time, can possess seemingly similar language proficiency to both groups as well. Nevertheless, comparisons either between NSs (but Korean-Americans) and G1.5KAs (usually bilinguals or English dominant) or between NNSs and G1.5KAs have little been explored.

[1] Those who came to America at a relatively young age with their own heritage language at home are referred to Generation 1.5, and they are extremely diverse in terms of English proficiency (Stanford University https://teachingcommons.stanford.edu/teachingwriting/pwr-guide/teaching-multilingual-students/generation-15-and-esl), and further research including the detailed definition of this group is being conducted. In this study, G1.5KAs is used for 1.5-generation Korean-Americans, G2KAs is for 2nd-generation Korean-Americans, and L2Ks is for Korean English learners.

Literature Review

Lexical Bundles: Concept and Definition

One of the most notable and novel studies of lexical bundles were that of Biber and Conrad. They defined lexical bundles as the most frequently occurring sequences, "of three or more words that show a statistical tendency to co-occur" (Biber & Conrad, 1999, p. 183). As to the number of combinations used, Chen and Baker (2010) remarked that there is no prohibition. The words in some bundles are strongly locked-in, only occurring in their specific combination (e.g., *kick the bucket*), while others are flexible and appear in multiple combinations (e.g., *I need to know, I want to know,* etc.). Initially identified in two major registers of conversation and academic prose in their earlier publication, lexical bundles are units that are required to occur at least 10 times per million words. Biber (1990) and colleagues (e.g., Conrad & Biber, 2004) suggested in the later studies that even data containing a few words could still produce reliable lexical bundle results through the normed rate of occurrence. 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 (Conrad & Biber, 2007, p. 268): (3/50,400)*1,000,000 = 60 per million words. The additional requirement imposed is that the bundles with a raw account of 3 must be distributed across three different texts, speakers or writers. Its frequency and occurrence in different texts, usually in at least 3-5 texts (e.g., Biber & Babieri, 2007; Cortes, 2004) or 10% of texts (e.g., Hyland, 2008), help avoid idiosyncrasies from individual writers or speakers.

Though structurally incomplete, the defining feature that distinguishes lexical bundles from other formulaic sentences is the occurrence at the phrase (e.g., *the nature of the, as a result of*) and clause boundaries (e.g., *I don't know what, I think it is*). Some shorter 3-unit or 4-unit lexical bundles can also be incorporated into longer lexical bundles (e.g., *I don't know* as part of *I don't know if*, *I don't know but, but I don't know*). Longer or shorter units of lexical bundles can be classifiable in terms of structure and discourse functions, which are described in the next section.

Structural Taxonomy

Two major taxonomies which have been widely relied upon in the studies of lexical bundles to date were classified by Conrad and Biber (2004): function in a discourse context and structural characteristics. Functional characteristics of lexical bundles are defined by discourse functions, resulting in the three functional groups of stance expression (e.g., *I don't know, I was going to, it is possible to*), discourse organizers (e.g., *what do you mean, do you know what*), and referential expressions (e.g., *in the form of, as shown in figure*). The second category, which is being highlighted in the study, is the structural characteristic classification of lexical bundles. Addressing these two taxonomies at once embraces a wider scope, and L2 learners tend to rely on structural units than functional usages when speaking (Zipagan & Lee 2018); therefore, the present study seeks to explore only the latter classification.

The structural characteristics of lexical bundles are characterized by their components (e.g., noun, verb, prepositional components, etc.). 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, 2004, p. 60). Taking into accounts the elements and the overall structure of the bundle, they are fully categorized into different structural types described by Biber et al. (2004). This structural taxonomy 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, whereas bundles like *in the case of* are from noun phrase and prepositional phrase components.

The lexical bundles in the first sub-category involve seven types of VP fragments incorporated with (connector+) first or second-person pronoun (*I'm not going to*), (connector+) third-person pronoun (*and this is a)*, discourse marker (*I mean you know*), VP with non-passive verb (*take a look at*), VP with passive verb (*can be used to*), yes-no question fragments (*are you going to*), and wh-question fragments (*what do you think*). The second sub-category are five lexical bundle types that are incorporated with dependent clause fragments with first- or second-person pronoun (*I don't know if*), wh- (*when we get to*), if (*if you want to*), (verb or adjective+) 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 +) 1st/2nd person pronoun + VP fragment	*You don't have to, I'm not going to*
1b. (connector +) 3rd person pronoun + VP fragment	*It's going to be, and this is a*
1c. Discourse marker + VP fragment	*I mean you know, you know it was, I mean I don't*
1d. VP with non-passive verb	*is one of the, have a lot of, take a look at*
1e. VP with passive verb	*Is based on the, can be used to, shown in figure N*
1f. *yes-no* question fragments	*Are you going to, does that make sense*
1g. Wh-question fragments	*What do you think, how many of you*
2. Lexical bundles that incorporate dependent clause fragments	
2a. 1st/2nd person pronoun + dependent clause fragment	*I want you to, I don't know if, you might want to*
2b. Wh-clause fragment	*What I want to, what's going to happen*
2c. *If*-clause fragment	*if you want to, if you have a, if we look at*
2d. (verb/adjective+) *to*-clause fragment	*to be able to, to come up with, want to do is*
2e. *that*- clause fragment	*That there is a, that I want to, that this is a*
3. Lexical bundles that incorporate NP and PP fragments	
3a. (connector +) NP with *of*-phrase fragment	*One of the things, the end of the, a little bit of*
3b. NP with other post-modifier fragment	*A little bit about, those of you who*
3c. Other NP expressions	*A little bit more, and stuff like that*
3d. PP expressions	*Of the things that, at the end of, at the same time*
3e. Comparative expressions	*As far as the, greater than or equal, as well as the*

Related Studies

Previous research has revealed that NSs and NNSs use lexical bundles differently (De Cock, 2000; Granger, 1998; Warga, 2005). Native language (L1) speakers have shown greater knowledge of lexical bundles than L2 learners. According to Cortes (2004, 2006), some bundles with certain functions were used more frequently than other bundles.

Chen and Baker (2010) investigated the difference of lexical bundles in L1 (published academic writing and student writing) and L2 students' academic writing. It was found that the lexical bundles in L1 academic writing exhibited the widest range and that L1 and L2 students showed similar writing patterns. Also, L1 experts' academic writing showed wider range of NP related lexical bundles while L1 and L2 students' writing showed mostly VP-related lexical bundles (Kim, 2013). In addition, high frequency expressions (e.g., *in the context of*) in published texts were underused in both student groups while the L2 writers overused certain expressions (e.g., *all over the world*) which L1 academic writers rarely used.

The results of this study seem to agree with De Cock's (2004) comparison of recurrent word combinations between NSs' and NNSs' speeches. She found that NSs' speech actually contains more recurrent word combinations of different lengths than NNSs' speech does. She added that the NSs' use of frequently recurring sequences of words displays a complex picture of overuse, underuse, or misuse of target language *(*e.g., overuse of *kind of, you know*) while for the NNSs, the use might indicate

idiosyncratic sequences (e.g., misuse of *something like that*). Erman and Lewis (2015) confirm that pragmatic markers such as *sort of* have been used ubiquitously by NSs and are under-represented by NNSs. It was one of the defining features between two groups. In general, Nekrasova (2009) stated that L2 speakers were found to be unaware of the more common, yet salient L2 chunks. In order to compensate for their lack of awareness, they often referred to L1 transfer, so they a) either modified or avoided using certain L2 constructions that did not have L1 equivalents, b) tended to overuse those L2 constructions which L1 equivalents were more common, and c) misused those constructions when L2 equivalents did not match their L1 counterparts.

Erman and Lewis (2015), in their study between a non-native Swedish group and a native English group, looked into the type to token (T/T) ratio. The types were divided into two ranges: low frequency (most frequent 1-2000 words) and high frequency words (words beyond 2000). In the low frequency words, the NNSs significantly outnumbered the NSs. Looking into the types, it revealed that NNSs recycled more words in this frequency range, implying less diversity. In the high frequency words, the token difference by the NSs to the NNSs was highly significant. This exposed that NSs use more specific vocabulary compared to NNSs.

In a Korean context, Kim (2013) found that Korean L2 writing exhibited more VP-related lexical bundles than native English counterparts while L1 writing showed a wider and more various use of lexical bundles without such a domination of a certain structural type. Moreover, Korean L2 learners' writing showed their heavier reliance on fixed expressions of spoken register than L1 writers (Yoon & Choi, 2015). When comparing three different proficiency groups' speaking within Korean L2 learners, even though they produced spoken responses to a test in a controlled environment, Zipagan and Lee (2018) observed contrasting results to the above mentioned previous studies that showed a positive relationship. In other words, novice speakers produced the most lexical bundles and advanced speakers produced the least bundles (the average normed rate of lexical bundles by 1 million words were 774 (novice), 223 (intermediate), and 190 (advanced)). However, their spoken data might yield different results from that in an uncontrolled environment of spoken discourse since most test takers in Korea tend to produce spoken responses after memorizing what they wrote in advance.

Studies have largely been binary (NSs vs. NNSs) without the inclusion of a third group of individuals whose language development and background may fall somewhere in between. In other words, little has been compared to explore the differences or similarities among *three* groups with the same heritage language and cultural background, that is, among NSs (G2KAs), NNSs who learned English in Korea (L2Ks), and those who learned English at a young age in America (G1.5KAs). This third group is assumed either to show the similar speech patterns to NSs based on their English proficiency or to Korean NNSs (L2Ks) based on their familiarity to Korean language and culture, but whether they are more like NSs or NNSs has yet to be explored.

Therefore, in order to contribute to the existing body of research on lexical bundles and describe their place within its function in the speech of NSs and NNSs in a Korean context, the present study explores the issue of structural characteristics that define the three groups and seeks to bring light to the following questions:

1) What differences or similarities, if any, are there in the frequency of lexical bundle uses among G2KAs, L2Ks, and G1.5KAs?
2) What differences or similarities, if any, are there in the structural patterns of lexical bundle use among G2KAs, L2Ks, and G1.5KAs? Do they use VP-related lexical bundles as often as L2 learners did in the previous studies?
3) What differences or similarities, if any, are there in the frequency and use of the most common lexical verbs in conversation (e.g., *get*, *go*, and *say*) among G2KAs, L2Ks, and G1.5KAs ?

Methods

Participants

A number of individuals from the researchers' own social and professional networks were asked to participate in the study. Based in Seoul, prospective participants were chosen on the basis of their linguistic backgrounds and their match to one of the three study groups in this research. Eight participants agreed to participate: three G2KAs, three G1.5KAs, and two L2Ks (Table 2).

TABLE 2
Participant Groups

	Description	Female	Male	Total
L2Ks	Korean English learners (English as a foreign language)	0	2	2
G1.5KAs	Korean-Americans (English as a second language from elementary and middle)	1	2	3
G2KAs	Korean-Americans born abroad whose native language is English.	2	1	3

Participant profile, linguistic background, and language ability were determined using a questionnaire (Tables 3 and 4). All individuals were based in Seoul. Three of the participants were university students, and remaining five were working professionals. Three were American, and the remaining five were Koreans. All were born in their country of citizenship. G2KA-6 (27, female) was born in the US but spent her elementary school years in Korea, which was followed by a move back to the US. For this reason, she considered herself perfectly bilingual. All three G1.5KA participants all immigrated to the US during elementary and middle school, during their critical period. L2K-2 (29, male) briefly studied abroad in Canada for six months.

TABLE 3
General Background of Participants

	Gender	Age	Nationality / Birth Country	Education Completed	Occupation
L2K-1	Male	27	Korean	High School	University student
L2K-2	Male	29	Korean	Bachelors	Electrician
G1.5KA-3	Female	21	Korean	High School	University student
G1.5KA-4	Male	32	Korean	Masters	Sales manager
G1.5KA-5	Male	33	Korean	Bachelors	Business owner
G2KA-6	Female	27	American	Masters	English teacher
G2KA-7	Female	19	American	High School	University student
G2KA-8	Male	29	American	Bachelors	Audio engineer

TABLE 4
Language Background of Participants

	First (Heritage) Language	Ages Lived Outside Korea	Self-rated English Ability	TOEIC Score	English Speaking Frequency (Daily)
L2K-1	Korean	N/A	Beginner	715	Never
L2K-2	Korean	24	Intermediate	850	Hardly
G1.5KA-3	Korean	6-13	Native	980	Always
G1.5KA-4	Korean	13-24	Proficient	970	Often
G1.5KA-5	Korean	11-24	Proficient	920	Hardly
G2KA-6	Korean	0-5, 12-25	Native	N/A	Always
G2KA-7	Korean	0-18	Native	N/A	Hardly
G2KA-8	Korean	0-26	Native	N/A	Always

Despite their various backgrounds, the participants were all ethnically Korean, born to Korean parents. Thus, every individual indicated that Korean was their *first language* growing up in their household. In the case of American participants, not only was Korean the first language of exposure, but also it was their *heritage language,* an undeveloped home language while living in an English-dominant society.

Participants were also asked to rate their English ability on a 6-point Likert (1-Beginner, 2-Low Intermediate, 3-Intermediate, 4-Proficient, 5-Fluent, 6-Native). L2K and G2KA individuals rated themselves as less than *proficient* and *native*, respectively. G1.5KA-3 (21, female) considered herself a native English speaker despite living in the US for a shorter amount of time than G1.5KA-4 (32, male) and G1.5KA-5 (33, male), who both rated themselves as *proficient*. Due to spending her formative language learning years in the US, G1.5KA-3 (21, female) revealed that she had native English intuition during language production and that certain expressions occurred more naturally in English than Korean.

Answers were various when participants were asked how often they spoke English on a daily basis. G1.5KA-3 (21, female), G2KA-6 (27, female), and G2KA-8 (29, male) indicated that they had daily English interaction with family, friends, or significant other. L2K-1 (27, male) did not use English at all in his daily life. G1.5KA-4's (32, male) position at work as a manager international sales was the reason for speaking English nearly every day. L2K-2 (29, male) and G1.5KA-5 (33, male) both indicated that they hardly spoke English on a day to day basis. G2KA-7 (19, female), despite being an American citizen and native English speaker, indicated that she hardly spoke English in her daily life, only speaking Korean to her family at home and friends at the university she attends.

Instruments

Each participant was interviewed in person for about 40 minutes (Table 3). In addition to the comfortable environments (cafés and work offices), the researcher and participants had personal relationships, which contributed to inducing natural speech. Even the L2Ks with relatively low proficiency were able to answer the questions in English with ease, so the interviews were like interactive informal conversations rather than formal speech or prefabricated answers [2].

TABLE 5
Duration of Interview

Participant	L2K-1	L2K-2	G1.5KA-3	G1.5KA-4	G1.5KA-5	G2KA-6	G2KA-7	G2KA-8
Gender	Male	Male	Female	Male	Male	Female	Female	Male
Duration	00:39:05	00:43:53	00:45:36	00:32:57	00:32:50	00:28:51	00:48:27	00:49:31

Interviews were audio-recorded using the interviewer's cell phone with the participants' permission. Questions regarding gender roles, relationships, and marriage were selected with the intent to keep the conversation interesting, entertaining, and thought-provoking for participants. The questions were as follows:

- What are some universal qualities of men that are attractive to women? (Vice versa?)
- How do you know if someone is attracted to you? What are the signals?
- In relationships, is it necessary for men and women to follow gender roles (e.g., men should pay on dates, women must be gentle and feminine, etc.) or should men and women act completely on an equal basis?
- Are gender roles different for Americans and Koreans? If so, how?
- Is it inappropriate if a woman makes the first move, such as asking for contact or requesting a date?
- Is there a difference between Americans and Koreans in the standards of dating? If so, how?

[2] Korean English learners tend to write English answers in advance and provided memorized, prefabricated answers when they have to get interviewed in English.

- Is marriage timing different for men and women?
- Do you agree or disagree with the following statement?
  - Men are usually interested in physical beauty and tend to ignore other non-physical features.
  - Women usually consider economics (i.e., money, job salary, etc.) in the matters of marriage.
- Why do people cheat in a relationship? Do men and women have different reasons for cheating?
- Why do you think the divorce rate is higher than ever before universally across all countries?
- What is your best dating advice for men? What is your best dating advice for women?
- Tell me about your language history. Number of spoken languages and duration of learning.
- We are now finished! Thank you so much for your time. Do you have any last questions for me?

Data Collection Procedures

The researchers first met to specify the three groups in focus and to choose topics that the participants would answer as much as possible. First, the three groups were defined to explore the differences and similarities in their English language uses: while all groups should be of Korean background, their English proficiency levels should be diverse from NSs, bilinguals of Korean and English, and Korean NNSs.

Then, considering their interest and personality, one of the topics that they would be most interested in was selected. The questions were set in advance to use as the guidelines for the semi-structured interview.

The researchers contacted the potential participants for each group and set up the interview schedule. Each participant was informed of this study, agreed to participate, and gave permission to record the interview. The interview was conducted for about 40 minutes (Table 5) using the guideline questions.

Data Analysis Procedures

The interview data was transcribed by two experts: a teaching assistant with English major and a doctoral student majoring in English education. The teaching assistant studied in America and worked with native English speaking professors, and the doctoral student is a Filipino with near native proficiency in English. The former transcribed the data and the latter checked the first version of transcription and completed it.

The transcribed interview was filed for each person, and each participant's answers without the interviewer's parts were excerpted to explore his or her language use only.

A freeware corpus analysis toolkit, *AntConc,* was used to concordance and analyze the three groups' language uses. Since there were not many participants, obtaining vast data on lexical bundles was not expected, thus trigrams (3-word lexical bundles) were used for analysis. *Word List* [3] of *AntConc* was used to compare the three groups' lexical verb uses, and *Clusters/N-Grams* [4] were used for the trigrams. When collecting the trigrams, the minimum frequency was set to 3 times and range was set to the number of the participants of each group. In other words, since G2KAs and G1.5KAs consisted of 3 participants, the trigrams used by all 3 of each group were collected and L2Ks' trigrams used by all 2 were collected. Since only 39 trigrams were collected from the L2Ks' data, 39 trigrams from each of the other groups were compared.

[3] It provides the list of words in the corpus with the ranks and the frequencies of each word.
[4] It provides the list of lexical bundles in the corpus with the ranks, the frequencies, and the range used of each n-gram.

Results and Discussion

Frequency of Lexical Bundle Uses Across the Three Groups

In general, as Table 6 shows, the three groups did not use many trigrams in total (780 trigrams out of 29,274 words spoken, 2.66%). There was a tendency in accordance with their English proficiency levels though. L2Ks (NNSs) used the fewest trigrams while G2KAs (NSs) used the most trigrams (L2Ks 1.39%, G1.5KAs 2.27%, and G2KAs 3.53%). The normalized numbers of trigrams used by each group also showed the same trends from L2Ks' 13,866, G1.5KAs' 22,689, to G2KAs' 35,331, which is in contrast to the previous study on Korean NNSs' speeches of three different proficiency levels (novice NNSs, intermediate NNSs, and advanced NNSs), but supports the previous comparisons between NS and NNS writers. Zipagan and Lee (2018) showed that lower level English learners used lexical bundles when speaking much more than the higher levels since the novices tend to depend on the formulaic patterns (Chen & Baker, 2010; Yoon & Choi, 2015). However, it was obvious that NNSs in the current study had significantly lower proficiency than the other two groups, but they used less bundles than the others. The significantly smaller number of the words NNSs (L2Ks) spoke in total (4,760 words) than G1.5KAs (12,032 words) and G2KAs (12,482 words) might have made this contrasting result to those about Korean NNSs. Also, the different register from the previous studies might have been a possible reason; in other words, free speaking data in this study might have yielded this contrasting result from writing data as in Chen and Baker (2010) or speaking data in a test-driven environment as in Zipagan and Lee (2018). Regarding the similar results to the previous studies showing that NSs used more and more diverse lexical bundles than L2 learners (Chen & Baker, 2010; De Cock, 2004; Kim, 2013), the G1.5KAs turned out to be similar to NSs like the G2KAs. The two groups (G2KAs and G1.5KAs) seem to share more common features in speaking than L2Ks. Further studies in free speaking environment and possibly with more participants are needed, especially considering G1.5KAs' English proficiency.

It is interesting to see the actual differences between L2Ks and G1.5KAs as well as differences between G1.5KAs and G2KAs. Even though G1.5KAs seemed to speak English just like NSs, or NNSs seemed to speak English similarly to G1.5KAs, each group showed their own distinctive features of how often they used formulaic expressions when speaking.

TABLE 6
Frequency and Normalized Number of Trigrams

	L2Ks	G1.5KAs	G2KAs	Total
No. of words spoken	4,760	12,032	12,482	29,274
No. of top 39 trigrams used	66	273	441	780
(% of the total number of words)	(1.39%)	(2.27%)	(3.53%)	(2.66%)
Normalized no. of the trigrams / 1 million words	13,866	22,689	35,331	26,645

Structural Patterns of Lexical Bundle Use

As presented above, NNSs (L2Ks) used the least trigrams and NSs (G2KAs) used the most trigrams. Differences or similarities in the structural patterns of lexical bundle use among the three groups should be examined and whether they used VP-related lexical bundles as often as L2 learners in the previous studies.

As shown in Table 7, lexical bundles with VP fragments used by L2Ks and G2KAs consist of the major part of all instances (L2Ks 75.76% and G2KAs 68.85%), which is more than G1.5KAs (56.78%). However, L2Ks did not seem to be able to use the lexical bundles with dependent clause fragments so frequently as the other groups with higher proficiency due to the structural difficulties of the expressions using wh- and if-clause fragments. Follow-up interviews with this result should be conducted to understand how they chose the expressions when speaking.

TABLE 7
Structural Types of Trigrams

Structural Types	NNS	1.5 NNS	NS	Total	Example
PHRASAL					
1. LB with VP fragments	50 (1.05%) <75.76%>	155 (1.29%) <56.78%>	332 (2.66%) <75.28%>	537 (1.83%) <68.85%>	
1a. (connector +) 1st/2nd person pronoun + VP fragments	29 (0.61%)	72 (0.60%)	157 (1.26%)	158 (0.88%)	*I don't, you have to*
1b. (connector +) 3rd person pronoun + VP fragments	13 (0.27%)	28 (0.23%)	85 (0.68%)	126 (0.43%)	*they need to, it's not*
1c. Discourse marker + VP fragment			6 (0.05%)	6 (0.02%)	*mean it's*
1d. VP (with non-passive verb)	5 (0.11%)	47 (0.39%)	49 (0.39%)	101 (0.35%)	*don't think, don't know*
1f. *yes-no* question fragments	3 (0.06%)	8 (0.07%)	18 (0.14%)	29 (0.10%)	*do you understand, don't you*
1g. Wh-question fragments			17 (0.14%)	17 (0.06%)	*what do you*
2. LB with dependent clause fragments	6 (0.13%) <9.09%>	61 (5.07%) <22.34%>	66 (0.53%) <14.97%>	133 (0.45%) <17.05%>	
2b. Wh-clause fragments		31 (0.26%)		31 (0.11%)	*that's what, when I was*
2c. *If*-clause fragments		30 (0.25%)	9 (0.07%)	39 (0.13%)	*if you're*
2e. *that* clause fragments	6 (0.13%)		57 (0.46%)	63 (0.22%)	*I think that, that's the*
3. LB with NP and PP fragments	10 (0.21%) <15.15%>	52 (0.43%) <19.05%>	43 (0.34%) <9.75%>	105 (0.36%) <13.46%>	
3a. (connector +) NP with *of*-phrase fragments		33 (0.27%)	43 (0.34%)	76 (0.26%)	*a lot of, kind of like*
3b. NP with other post-modifier fragments		16 (0.13%)		16 (0.05%)	*something like that, high school and*
3c. Other NP expressions	3 (0.06%)			3 (0.01%)	*no problem uh*
3d. PP expressions	7 (0.15%)	3 (0.02%)		10 (0.03%)	*in my case, in the recording*
CLAUSAL					
Independent clauses		5 (0.04%) <1.83%>		5 (0.02%) <0.64%>	*How about you*
Total	66 (1.39%) <100%>	273 (2.27%) <100%>	441 (3.53%) <100%>	780 (2.66%) <100%>	

Lastly, G1.5KAs produced a short sentence, "*how about you?,*" incorporating only three words. Similar results of producing short sentences with only four words were found in Zipagan and Lee's (2018) study as well, which they said was not often observed in other previous studies. The question, "*how about you?,*" seeks for the interviewer's opinions unlike the other two groups, which may mean that the G1.5KAs considered the interviewer like their friends more than the other groups. Also, being that the interviewer was an English professor, L2Ks might have consciously performed to impress or show formality, causing them to hesitate to have a comfortable back and forth with the interviewer. It should be further studied as to why the other groups did not ask for the interviewer's opinions as well as why G1.5KAs asked for it while being interviewed. Also, why NSs in this study, G2KAs, dominantly used VP-related lexical bundles, unlike NS writers in the previous studies (e.g., Kim, 2013) should be further studied qualitatively as well as quantitatively. They are NSs but share the same heritage language, Korean, to the other groups, which might have caused the similar uses of VP-related lexical bundles. It should be further explored what made Koreans, regardless of being NSs (G2KAs in this study) or NNSs (all the three groups of Zipagan & Lee (2018)) of English, use VP-related lexical bundles more than the other types of lexical bundles.

Frequency and Use of the Most Common Lexical Verbs in Conversation

Considering that VP-related lexical bundles were mostly used by all the three groups in this study, the kinds of lexical verbs used were also explored. Tables 8 and 9 show the kinds of the most common lexical verbs in conversation each group used and how frequently. Unlike the total number of trigrams the three groups used, the uses of the most common lexical verbs in conversation (*get, go, say, know, think, see, want, come, mean, take, make,* and *give*[5]) showed somewhat different trends. The verb variations of language users are assumed to increase as their proficiency level increases (Chen & Baker, 2010; Kim, 2013). However, it was not L2Ks but G1.5KAs that used the fewest types of the verbs (4 types: *get, go, think,* and *want*) least frequently (224 out of 12,032 words, 1.86%). However, as expected, G2KAs (with the highest proficiency) used five verbs (*get, say, know, think*, and *mean*) most frequently (414 out of 12,482, 3.32%).

The three groups used *think*, the fifth verb in the list, the most frequently in common. It is possible that the topic made them use the verb often to emphasize their own opinion on their ideal types of boy or girlfriends. If they were asked a different topic, for example, regarding the information of their jobs or majors, then the results might have been different. Further studies on the relationships between the topics and the verbs used should be conducted.

Although the differences in the types of the verbs used were not that big, it should be noted that L2Ks used the most various verbs (7 types: *get, say, know, think, see, want*, and *make*) of all. It is in contrast to the previous observations reporting that those with higher proficiency tend to use various verbs than those with lower proficiency (Chen & Baker, 2010; Kim, 2013; Staples et al., 2013). Furthermore, comparing *all* the verbs each group used, and not only the 12 verbs commonly used in conversation, yielded the same results. The normalized numbers of the verbs in Table 7 show that L2Ks used more types of verbs (16) more frequently (53,992) than the other two groups. Considering the top-100 word list, L2Ks used more lexical verbs than the other two groups despite being the least proficient, while the other two groups used a wider range of word types (verbs, nouns, adverbs, etc.). With more interview data, Korean NNSs group's (L2Ks) heavy dependence on verbs when speaking needs to be further explored and the diversity of word uses should be compared to see the distinctive features in word uses of each group.

The results of the G1.5KA group are also noteworthy. On one hand, the total amount of the words G1.5KAs used was closer to that of natives (G2KAs) than non-natives (L2Ks), but on the other hand, G1.5KAs used the least variety of lexical verbs compared to the two groups. This group should be further studied in future research. Also, as mentioned earlier, they are too diverse to easily group or classify, so more studies should be conducted to give insight to the Korean English users, both NSs or NNSs including G1.5KAs.

TABLE 8
Frequency of the Most Common Lexical Verbs in Conversation

		L2Ks	G1.5KAs	G2KAs
1.	*Get*	17 <13.18%>	36 <16.07%>	46 <11.11%>
2.	*go*		36 <16.07%>	
3.	*say*	12 <9.30%>		29 <7.00%>
4.	*know*	19 <14.73%>		77 <18.60%>
5.	*think*	40 <31.01%>	86 <38.39%>	206 <49.76%>
6.	*see*	9 <6.98%>		
7.	*want*	17 <13.18%>	66 <29.47%>	
8.	*come*			
9.	*mean*			56 <13.53%>
10.	*take*			
11.	*make*	15 <11.63%>		
12.	*give*			
	Total	129 <100%>	224 <100%>	414 <100%>

[5] The 12 most common lexical verbs used in conversation by native English speakers were collected using the corpus (Georgia State University).

TABLE 9
The Most Common Lexical Verbs in Conversation vs. All the Lexical Verbs Used

		L2Ks	G1.5KAs	G2KAs
The most common lexical verbs in conversation	No. of the verbs	129	224	414
	(% of the total number of words)	(2.71%)	(1.86%)	(3.32%)
	Normalized no. of the verbs / 1 million words	27,101	18,617	33,168
	No. of the types of the verbs	7	4	5
All the lexical verbs used (Among the top 100 words)	No. of the verbs	257	491	672
	(% of the total number of words)	(5.40%)	(4.08%)	(5.38%)
	Normalized no. of the verbs / 1 million words	53,992	40,808	53,838
	No. of the types of the verbs	16	11	11

Conclusion

This study explored the characteristics of trigrams and lexical verbs in speaking of three different English speaking groups of ethnic Koreans. Previous studies have compared the lexical bundle uses between NSs and NNSs (Chen & Baker, 2010; De Cock, 2004; Erman & Lewis, 2015; Kwon & Lee, 2014; Shin et al., 2019) or among NNSs with different proficiency levels (Kim, 2013; Yoon & Choi, 2015; Zipagan & Lee, 2018). This study tried to reveal the differences among Korean English users with the novel inclusion of another proficiency group: Korean-Americans whose first language is English (G2KAs), bilingual Korean-Americans who started to learn English at an early age (possibly more English dominant than Korean dominant; G1.5KAs), and Koreans who learned English in Korea (first language being Korean; L2Ks). Since they share the same heritage language and culture, but with different proficiency in English, it was expected to show different patterns in language use from the previous studies.

The results were consistent with previous studies; in other words, this study observed an increase in language production with increase in English proficiency like the previous studies (Ellis 2006; Staples et al. 2013), but in contrast to the Korean L2 learners' decrease in production with increased proficiency (Zipagan & Lee, 2018). Further studies to expand this topic could be the differences in lexical bundles amongst different genders, ethnicities, or age groups.

Another interesting result was found in terms of the structural types of the trigrams used by the three groups. The G1.5KAs used verbs and VP-related trigrams less than the other two groups, but they used more types of the trigrams more frequently than the other two groups. In other words, G2KA and L2K groups used mostly VP-related trigrams while G1.5KAs used more lexical bundles with dependent clause fragments (22.34% > L2Ks' 9.09% and G2KAs' 14.97%), lexical bundles with NP and PP fragments (19.05% > L2Ks' 15.15% and G2KAs' 9.75%), and independent clauses (1.83% > 0 % of L2Ks and G2KAs). When considering the previous studies and the total frequency of this study that showed upward trends in correlation to proficiency level increases, G1.5KAs' wider use than G2KAs and L2Ks might have an implication toward possible benefits to be bilinguals. Or when considering the results of this study about the number of the types of the lexical verbs used, what made L2Ks with the lowest proficiency use the more types of the lexical verbs than the others (in other words, why NNSs depend mostly on verbs when speaking) should be further explored.

Although this study included a small number of the participants, contrasting results with new groups may offer more avenues of comparing the language user groups, not just with NSs but NSs with the same heritage language (G2KAs in this study). In other words, in this study, grouping and comparing English language users in a different way from the previous studies, NSs with NNSs, or NNSs with different proficiency (advanced, intermediate, and beginner), were attempts to better understand a possible influence of their heritage language and culture on English speaking. Possible future studies in relation to the results were suggested. It is hoped that this study would contribute to the understandings of the different groups' language uses in terms of lexical bundles and lexical verbs. Teachers are encouraged to teach lexical verbs but in a variety of contexts so that learners can extract the precise meaning and

appropriate situation for the verbs to be used in conversation. Also, they should teach the lexical bundles with other speech parts than verbs. L2Ks used mostly lexical bundles with VP fragments, so it will benefit them from using lexical bundles with NP fragments and/or lexical bundles with dependent clause fragments such as wh- and if-clauses. Lastly, the VP-related lexical bundles used by G2KAs should be thoroughly compared with those used by L2Ks to see what made G2KAs used them, unlike NSs in other studies that used lexical bundles in a wider range.

The Authors

Julian Chung is an assistant professor in the General English Program at Sejong University, Seoul, Korea. He received his M.A. in TESOL at Columbia University, Teachers College and B.A. in Linguistics at the University of California, Riverside. His interests include language attrition and bilingualism, and is dedicated to the unearthing the processes or retrieval and inhibition in language learning.

General English Program
Sejong University
Gunja-dong 98, Gwangjin-gu, Seoul, 143-747, Korea
Phone: +82-2-3408-3615
Fax: + 82-2-3408-3443
Email: chung.julian@gmail.com

Kyoung Rang Lee is a 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 the differences and similarities among Korean English users: Korean-Americans whose first language is English, bilingual Korean-Americans who started to learn English at an early age, and Koreans who learned English in Korea.

Department of English Language and Literature
Sejong University
Gunja-dong 98, Gwangjin-gu, Seoul, 143-747, Korea
Tel: +82-2-3408-3118
Fax: +82-2-3408-4302
Email: kranglee@sejong.ac.kr

References

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. *English for Specific Purposes, 23,* 397-423.

Cortes, V. (2006). Teaching lexical bundles in the discipline: An example from a writing intensive history class. *Linguisctics and Education, 17,* 391-406.

De Cock, S. (2000). Repetitive phrasal chunkiness and advanced EFL speech and writing. In C. Mair & M. Hundt (Eds.), *Corpus linguistics and linguistic theory* (pp. 51-68). Amsterdam: Rodopi.

De Cock, S. (2004). Preferred sequences of words in NS and NNS speech. *Belgium Journal of English and Literatures (BELL),* New Series *2,* 225-246.

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 Apliquée Review, 19,* 100-121.

Erman, B., & Lewis, M. (2015). There is nothing like native speech: A comparison of native and very advanced non-native speech. In P. Shaw, B. Erman, G. Melchers, & P. Sundkvist (Eds.), *From clerks to corpora: Essays on the English language yesterday and today* (pp. 349-366). Stockholm: Stockholm University Press.

Granger, S. (1998). Prefabricated patterns in advanced EFL writing: Collocations and formulae. In A.H. Cowie (Ed.), *Phraseology: Theology, analysi, and applications* (pp. 145-160). Oxford: Claredon Press.

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 & Literature Teaching, 19*(3), 157-179.

Kwon, Y.-E., & Lee E.-J. (2014). Lexical bundles in the Korean EFL teacher talk corpus: A comparison between non-native and native English teachers. *The Journal of Asia TEFL*, *11*(3), 73-103.

Nekrasova, T. (2009). English L1 and L2 speakers' knowledge of lexical bundles. *Language Learning, 59*(3), 647-686.

Shin, Y. K., Choi, H., Kim, D., Ko, S.-J., Yoo, H., Yoo, H., … & Yoo, I. W. (2019). Syntactic complexity of recurrent multiword sequences in the writings of published authors and L1 and L2 English apprentice writers. *The Journal of Asia TEFL*, *16*(2), 516-530.

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.

Warga, M. (2005). "Je serais tres merciable": Formulaic vs. creatively produced speech in learner's request-closing. *Canadian Journal of Applied Linguistics, 8*(1), 67-93.

Wray, A. (2000). Formulaic sequences in second language teaching: Principle and practice. *Applied Linguistics, 21*(4), 463-489.

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.

Zipagan, M., & Lee, K. (2018). Korean English learners' use of lexical bundles in speaking. *The Journal of Asia TEFL, 15*(2), 276-291.