JoLaCE

*Journal of Language and Cultural Education*, 2024, *12*(3)
ISSN 1339-4584

sciendo

DOI: 10.2478/jolace-2024-0016

# The effects of residency experience in English-speaking countries on English lexical bundle usage among Korean university students

**Julian Chung & Kyoung Rang Lee***
Sejong University, South Korea
kranglee@sejong.ac.kr

**Abstract**

This study aimed to analyse the usage of English lexical bundles among university students who transitioned to Korean universities after extended periods in English-speaking countries. It was assumed that the duration of their stay would lead to distinct English language usage patterns. Ten participants were recruited: those who resided abroad for more than 5 years (six learners) and those who lived abroad for less than 5 years (four learners). Interviews were conducted, yielding an average of 30 minutes for each interview. The results revealed significant differences in word frequency and trigram usage in terms of students' duration of stay. In other words, the residency experience in English-speaking countries influenced their language use patterns even though their current proficiency remained the same. Students with more extended stays showed greater trigram diversity and native-like patterns with VP and NP-PP fragments. In comparison, students with shorter stays displayed a prevalence of skewed VP fragment use and had narrower trigram usage. The factors contributing to their differing language use patterns should be investigated further when maintaining similar proficiency levels. Despite some limitations, such as excluding the impact of individual motivation, the findings highlight the importance of individualised language learning approaches, even among learners with similar proficiency levels.

**Key words**: English proficiency development, lexical bundle, residential experience, English-speaking countries

## 1 Introduction

Frequently used language expressions provide considerable insight into the knowledge and ability of the users of a particular language, especially language learners (Chung & Lee, 2020). The common expressions of chunked words are fundamental building blocks in improving the understanding and fluency of English learners in conversation (Hyland, 2008). These language expressions refer to *lexical bundles* (*formulaic language, fixed expressions, prefabricated patterns, language chunks, ready-made utterances, n-grams,* and *multi-word expressions*) that often appear together when analysing spoken or written language data. Research on lexical bundles involves contrasting various types of speakers, including native English speakers and English learners (Chen & Baker, 2010), and comparing learners based on proficiency levels (Zipagan & Lee, 2018). Lexical bundles with three words, *trigrams*, have been considered proper for exploring English learners' spoken language usage since quadrigrams are produced less and bigrams have too many combinations without meaningful information (Hyland, 2012).

Recently, there have been attempts to compare the trigram use of Korean Americans, who are native English speakers but can also speak Korean as a heritage language, with Korean English learners (Chung & Lee, 2020). The differences in vocabulary usage between native English speakers and Korean Americans, despite both being native English speakers, offer new insights and possibly a bridge in the understanding of English learners in our country. Especially for English learners, comparisons based on experiences as opposed to sole knowledge are necessary, as the number of international students returning to Korea and enrolling in Korean universities due to COVID-19 has increased. Rather than solely distinguishing them based on proficiency, it is essential to investigate whether the experiences of acquiring English in an English-speaking country have influenced their current language usage.

Similarly, concerning language acquisition, those who immigrated to the United States relatively at young ages, before 12, were considered 1.5 generations to differentiate them from those born in the United States (Southern California Public Radio, 2012). While the two groups looked similar since they are native English speakers with little proficiency in their heritage language, Korean, their actual language uses were different (Chung & Lee, 2020; Zipagan & Lee, 2018). The results suggest that even when maintaining similar proficiency levels, the duration of their stay might have caused different language uses, which has to be dealt with using the more scrutinised lens.

Furthermore, research suggests that the criteria for evaluating English usage differ for those who have resided in a specific country or institution for more than 5 years compared to newcomers who have assimilated into the culture of that country or institution (Lee, 2016). The results suggested that living and learning English in an English-speaking country for more than 5 years would mark a significant difference.

Staying in a particular country for more than 5 years seems to have affected the professors' rating behaviours and learners' language use. In other words, it is necessary to explore possible differences between English language learners in terms of their experiences of living in an English-speaking country for more than 5 years compared to those with a shorter stay (4 years or less). The first group is referred to as L students to emphasise their longer duration of stay, while the latter group is referred to as S students to highlight their shorter stay. This research is expected to be valuable for future English education and research to explore possible variables for understanding English language learners in addition to differences in their proficiency.

## 2 Literature review

### 2.1 Concept and definitions of lexical bundles

One of the most commonly used definitions in lexical bundles research is provided by Biber and Conrad (1999), who define lexical bundles as expressions where three or more words are likely to co-occur. While some scholars do not explicitly mention the number of co-occurring words, there is a tendency in research to study lexical bundles where approximately three words occur together in spoken language and four words in written language. Lexical bundles include fixed expressions (e.g., *beat around the bush*), where specific expressions appear in a fixed form, but in most cases, they consist of a wide range of combinations (e.g., *I need to know, but I think*). It is suggested that combinations that occur more than 60 times per million words among this variety of lexical bundles should be considered significant. This calculation is based on extrapolating occurrences from a sample of 50,400 words to one million words (3 / 50,400 * 1,000,000 = 60; Conrad & Biber, 2007, p. 268).

When analysing lexical bundles in conversation, researchers often distinguish them based on their functional and structural characteristics (Biber et al., 2004; Conrad & Biber, 2004). First, to analyse the functional features of lexical bundles, they are divided into stance expressions, discourse organisers, and referential expressions. Stance expressions refer to phrases such as *I don't know, I was going to,* and *it is possible* to convey the speaker's attitude and assessment of certainty, obligations, and directives. Discourse organisers include expressions like *What do you mean* and *Do you know what,* which do not directly reveal the content but functionally assist the speaker in linking ideas during communication. Referential expressions refer to tangible and intangible entities directly and consist of phrases like *in the form of* and *shown in the figure*.

Additionally, to compare the structural characteristics, lexical bundles are categorised into verb phrase (VP) fragments, dependent clause fragments, and noun phrase (NP) and prepositional phrase (PP) fragments. English learners tend to use language based on structural/grammatical units rather than functionally distinguishing them, so it is more effective to compare speech patterns based on syntactic types rather than functional features (Zipagan & Lee, 2018). When comparing functional features, lexical bundles are further divided based on the grammatical elements of the central word (e.g., noun, verb, preposition). Although lexical bundles are not strictly divided by grammar, they can be differentiated by specific grammatical structures (Conrad & Biber, 2004). For example, lexical bundles like *you want me to* are expressions centred around verbs, while *in the case of* is, an expression centred around nouns and prepositions. These three main categories can be subdivided into seven, five, and five subcategories, respectively (see Table 1), providing a valuable framework for analysing learners' lexical bundles.

Table 1. Structural Types of Lexical Bundles (Biber, Conrad, & Cortes, 2004)

| Types | Examples |
|---|---|
| 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 them, have a lot of, take a look at* |
| 1e. VP with a passive verb | *It 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 fragments | *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* |

In Chung and Lee's (2020) study comparing the use of trigrams in the speech of English learners in Korea and Korean-Americans speaking Korean as a heritage language in the United States, it was found that English learners used the fewest number of trigrams, while native Korean-Americans utilised the most. Over half of the trigrams spoken by most participants were VP fragments (75.76% for English learners and 68.85% for Korean-Americans). Interestingly, English learners rarely used dependent clause fragments, making using wh- and if-clause fragments challenging. This was in contrast to Korean-Americans, who used VP fragments significantly more, which differs from research results on other English native speakers (Kim, 2013). However, it aligns with previous research (Zipagan & Lee, 2018) that Korean English learners tend to use VP fragments more frequently, regardless of proficiency level.

In summary, regardless of whether they were native English speakers or learners, groups that share Korean as their common language tend to prefer VP fragments over NP fragments. Therefore, it would be worth investigating whether English learners who had early experiences living and learning in an English-speaking country also prefer VP fragments similar to those previously observed.

**2.2 Previous research on lexical bundles**

Existing research has shown that native English speakers and non-native speakers (learners) use lexical bundles differently when speaking or writing in English (Chen & Baker, 2010; De Cock, 2000; Warga, 2005). Much of the research has focused on written data rather than spoken language, with Chen and Baker (2010) being a prominent example. They compared academic writing by native English speakers with essays written by English learners. Native English speakers' published writings contained various VP and NP fragments, while English learners primarily used VP fragments, similar to Korean English learners. Interestingly, even when comparing native English speakers who were students, they exhibited patterns more similar to English learners than expert authors, suggesting that native speakers also require education and learning to produce proficient writing.

Similar results were found in spoken language, where native English speakers used more lexical bundles than non-native speakers. Specific expressions such as *something like that* were frequently used by non-native speakers (De Cock, 2004). Native speakers often used expressions like *sort of*, which were rarely used by non-native speakers (Erman & Lewis, 2015). These findings suggest that comparing lexical bundle usage is appropriate for studying various language users, regardless of their native or non-native status.

Another interesting comparison is based on vocabulary frequency, where non-native speakers used low-frequency vocabulary more frequently than native speakers (Erman & Lewis, 2015). This suggests that when designing vocabulary textbooks for English learners, it may be beneficial to exclude low-frequency words that native speakers do not commonly use and instead focus on presenting a variety of high-frequency words.

Research has shown that Korean English learners often mix spoken and written lexical bundles, even though there are differences between the two registers (Yoon & Choi, 2015). This trend is similar to what has been observed in Japanese English learners (Ishikawa, 2015). In contrast to previous results, there was also a finding that as proficiency increased, learners used lexical bundles less frequently (Zipagan & Lee, 2018). Based on a comparison of Japanese English learners by proficiency level, it was suggested that it might be more appropriate for learners with lower English proficiency to initially focus on high-frequency expressions rather than teaching the differences between spoken and written language (Ishikawa, 2015).

While some attempts have been made to categorise English native speakers further into groups like student native speakers or Korean-Americans who can speak their heritage language (i.e., Korean), English learners' categorisation has been primarily based on proficiency levels. Also, few studies have explored the influence of residency duration on the language usage patterns of English learners who transitioned to Korean universities. Therefore, it is necessary to investigate how the experience of living and learning English in an English-speaking country relates to their current uses of words and trigrams. Thus, this study aims to compare the word and trigram usages of Korean university students who have had experiences living in an English-speaking country for a long time with those who lived for a relatively short time.

To achieve this, the following research questions were formulated:

1) Do Korean university students who have had experiences living in an English-speaking country for a long time use words and trigrams differently from those who had shorter stays? If so, how do they differ?

2) Are there differences in the structural types of trigrams used by the two groups? If so, what are these differences?

## 3 Research methods

### 3.1 Participants

This study was conducted at a university in Seoul, South Korea, and participants were recruited by posting notices in the facilities where students majoring in English literature and linguistics commonly attended classes. The recruitment aimed to target Korean university students with experience living in English-speaking countries, and those locations were assumed to be the ideal areas for recruitment. The information on the posts emphasised that the study would involve interviews focusing on their experiences of living in English-speaking countries during their early years and later attending a Korean university. Students who voluntarily agreed to participate and had experiences matching the criteria were included in the study.

A total of 10 students participated, with six students having long-term residency experiences (Long /L): 3 females and three males) and four students having short-term residency experiences (Short /S): all females). The participant demographics are presented in Table 2. Even in cases where the countries of residence were in Asia, for our study, we considered them to be English-speaking countries. This is because these students all attended English-speaking international schools, and their inner circles of daily life comprised native English speakers. Their immediate community shared similar environments with those in the United States, Canada, or Australia. Among the students with long-term experiences in English-speaking countries, all except L1 students found English more comfortable than Korean. All the students with short-term experiences found Korean more comfortable than English. Despite the significantly extended residency period, L1 students considered Korean more fluent than English, while L4 students had a relatively shorter residency period and had higher English proficiency. All participants' English proficiency levels were advanced in all four skills. Students with short-term experiences tended to answer in Korean even when questioned in English, so they were encouraged to respond in English.

Table 2. Profile of participants

| | Gender | Age | Country they stayed | Duration (Ages) | Self-rated proficiency |
|---|---|---|---|---|---|
| L1 | female | 25 | Malaysia | 7 years (10-18) | Korean > English |
| L2 | female | 22 | Singapore | 9 years (6-14) | English > Korean |
| L3 | male | 27 | Canada | 6 years (10-15) | English > Korean |
| L4 | male | 23 | America | 5 years (7-11) | English > Korean |
| L5 | male | 27 | India | 17 years (1-17) | English > Korean |
| L6 | female | 23 | Philippines | 14 years (5-18) | English > Korean |
| S1 | female | 23 | Philippines | 4 years (14-17) | Korean > English |
| S2 | female | 23 | Australia | 3 years (16-18) | Korean > English |
| S3 | female | 23 | America | 1 year (20-21) | Korean > English |
| S4 | female | 23 | America | 1 year (20-21) | Korean > English |

### 3.2 Instruments

As seen in Table 3, participants engaged in interviews lasting approximately 30 minutes. After explaining the study and obtaining their informed consent through signed forms, the conversations were recorded using mobile phones with the participants' consent. The interviews were conducted in a classroom setting where participants felt comfortable, starting with small talk about their age and current courses. Semi-structured interviews were conducted to allow for flexibility, with prepared questions to ensure each participant covered similar content.

The pre-prepared questions included "*When and where did you first learn Korean/English?*" "*Did you speak Korean only with your parents?*" "*How proficient are you in Korean and English?*" "*What do you consider is your mother tongue?*" "*Would you want your child to learn Korean?*" and "*Who do you speak English with these days?*" These questions served as starting points, and the conversations naturally expanded from these topics. The questions primarily focused on language acquisition and usage situations, as well as the use of language in Korean universities and language maintenance after transitioning to Korean universities.

Table 3. Duration of interview

| | Long-term students | | | | | | Short-term students | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| average | 37:05 | | | | | | 32:23 | | | |
| Participant | L1 | L2 | L3 | L4 | L5 | L6 | S1 | S2 | S3 | S4 |
| Duration | 47:25 | 30:16 | 28:22 | 38:08 | 42:43 | 35:33 | 34:18 | 35:16 | 28:22 | 31:35 |

**3.3 Data collection procedure**

In total, 10 students were interviewed over one week. The interviews began with small talk, including questions about their names, ages, majors, and the courses they were currently taking. This initial conversation was designed to create a relaxed atmosphere.

After initiating the conversation with questions about the participants' country of residence and the reasons for their stay there, the interview topics shifted to their language learning experiences. Participants were asked when and where they began learning English and Korean, which language they found more comfortable, and which language they primarily used in Korea.

While the interviews were expected to last approximately 30 minutes, the interview with the L1 student took longer due to the extensive content related to their major courses, anecdotes about their parents, and other introductory small talk. The interviews with the remaining participants were similar in length.

**3.4 Data analysis procedure**

The interviews were transcribed by a graduate student majoring in English education who had prior experience transcribing interviews. As a bilingual speaker fluent in Korean and English, this student had no difficulty transcribing the interview content. It took approximately 3 hours to transcribe each of the 30-minute interviews, totalling 30 hours for all 10 interviews. The entire transcription process took about a month to complete.

The transcribed files were saved with coded names for each participant. Participants who had lived in English-speaking countries for an extended period were coded as *L* for *Long*, followed by a sequential number corresponding to the order of the interviews. Participants who had lived in English-speaking countries for a shorter duration were coded as *S* for *Short* and numbered sequentially. Since all the participants who had lived for a shorter duration were female, gender coding was not applied.

An English education expert with a PhD and over 20 years of university teaching experience listened to the interview files and reviewed and verified the transcriptions for accuracy.

The interview data was analysed using the free corpus analysis tool *AntConc*. Due to the limited number of participants, trigrams were collected and analysed using *N-Gram*. Following the definition by Conrad and Biber (2007), a minimum frequency of 3 was set for extracting trigrams for comparison.

The vocabulary usage was assessed using type and token frequency (type-token ratio; TTR) data provided by *AntConc*. The extracted vocabulary was categorised into three structural types (VP fragments, dependent clause fragments, NP and PP fragments) with 7, 5, and 5 subcategories, respectively, as outlined in Table 1. Since the utterances varied across groups, normalised frequency data was extracted using *AntConc* and compared. Ratios were calculated based on the total number of utterances (words and trigrams) for comparison purposes.

**4 Results and discussion**

**4.1 Frequency of words and trigrams**

As shown in Table 4, the usage of trigrams was very low compared to the total number of spoken words (L: 2.99%, S: 2.25%). This result aligns with previous research, such as the study by Chung and Lee (2020), which found that participants used vocabulary expressions at an average rate of 2.66%, with English learners having the lowest rate (1.39%) and second-generation Korean Americans, who are native speakers of English, having the highest rate (3.53%).

The results suggest a positive relationship between English proficiency and trigram usage, contradicting previous studies suggesting beginners use prefabricated phrases (Chen & Baker, 2010; Yoon & Choi, 2015).

This study offers a more precise snapshot of the speaker's actual trigram usage in its more natural state. This is attributed to participants engaging in open-ended, free-flowing conversations. In contrast, previous studies (Staples et al., 2013; Zipagan & Lee, 2018) gathered data through oral and written tests, and thus performative awareness could contaminate the collection of authentic language. The results aligned with participants' self-rated proficiency ratings rather than proficiency levels. The four L students who rated their English more proficient than Korean were found to have used more extensive trigrams and generally spoke more. In other words, those who had lived in English-speaking countries for an extended period exhibited a more native-like English usage pattern (De Cock, 2004). This might highlight the importance of learners' self-confidence in effectively using diverse words and trigrams.

The results are consistent with previous studies indicating that native speakers tend to use more vocabulary more diversely than non-native speakers (Chen & Baker, 2010; De Cock, 2004; Kim, 2013). This study revealed a positive correlation between language production and length of residence in an English-speaking country. While the durations of their interviews were not significantly different (Table 3, L: 37:05, S: 32:23), there was a significant disparity in average word count in their speech, with the L students producing twice as many words as S students (2,861 vs. 1,465). Almost three times as many trigrams (86 vs. 33). These results did not support Skehan (1998)

proposing a negative relationship between an improvement in one area and a decrease in another. The participants in this study showed a positive relationship between word frequency and their English proficiency, as did the other studies (Polat & Kim, 2014), and there was also a positive relationship between trigram frequency and their English proficiency.

A partial explanation of this discrepancy may be found directly in the participants' interview answers, revealing a clear difference in the extent and desire for language maintenance between the L and S students. Upon returning to Korea after living in an English-speaking country, L students consistently showed high motivation and effort to use English in their daily lives with interlocutors. All L students actively and purposefully sought English-speaking situations and friends to integrate language practice into their schedules. S students also expressed a high desire to maintain English; however, they frequently regretted the inability to use English properly in their daily lives in Korea. Most expressed a negative societal pressure regarding their English performance, stating that the expectation of English ability (having lived in an English-speaking country) was burdensome. Some tried to avoid English-speaking contexts and communicated less with friends in English. It is necessary to investigate further how the S group maintained their advanced English proficiency despite this negative pressure and whether this negative pressure might have reduced the use of words and trigrams.

Lexical variety, however, showed a contrasting relationship with trigrams than between trigrams and word frequency. The type-token ratios of the two groups in this study used as an index of lexical diversity indicated a negative correlation with trigrams; the more trigrams were used, the less diverse the vocabulary produced. L students (TTR: 28.24) used less variety of words than S students (TTR: 38.37), unlike native speakers (or higher proficiency speakers) using more diverse words than non-native speakers. Other studies (Vercellotti et al., 2021) have also reported a negative correlation between trigrams and lexical variety. It may be worthwhile to explore in later research why or how this phenomenon occurs.

Future research could involve a broader range of groups in comparing trigram usage patterns. Investigating usage between men and women or amongst groups of learners who have never had experience living abroad would be an informative contribution to the existing literature. In turn, comparing the findings across these studies may create additional avenues of research. For instance, when comparing to the study by Chung and Lee (2020), where English learners (with little to no experience living in English-speaking countries) had a low rate of 1.39% for trigrams usage, the S students in this study had a notably higher rate of trigrams usage. Further research could open up more analyses of the specific structural patterns of trigram usage.

Table 4: Frequency and normalised number of trigrams

| | Long-term students | Short-term students | Total |
|---|---|---|---|
| No. of words spoken (Token)<br>Type (TTR) | 17,165<br>4,848 (28.24) | 5,861<br>2,249 (38.37%) | 23,026 |
| No. of trigrams used<br>(% of the total number of words)<br>No. of average words per student<br>No. of average trigrams per student | 514<br>(2.99%)<br>2,861<br>86 | 132<br>(2.25%)<br>1,465<br>33 | 646<br>(2.81%)<br>2,303<br>65 |

**4.2 Structural patterns of lexical bundle usage**

One of the most distinctive findings from previous research comparing vocabulary syntactic patterns was that native English speakers use NP fragments more diversely and frequently than non-native speakers. However, this study found that regardless of their residency duration in English-speaking countries, both groups primarily used VP fragments (Table 5, L: 71.4%, S: 84.5%). This differs from previous studies' findings but is similar to Chung and Lee's (2020), which found that bilingual Korean-Americans, despite being native English speakers, predominantly used VP fragments. However, as mentioned earlier, participants who had lived in English-speaking countries for an extended period exhibited characteristics more similar to native English speakers regarding structural patterns of trigrams.

As shown in Table 5, L students used a wider variety of trigrams. While primarily used VP fragments (71.4%), they used five of the seven subtypes relatively evenly. S students used VP fragments at a higher rate (84.5%) yet primarily used one subtype (1a: 70.5%) involving first-person and second-person subjects. This suggests that living in an English-speaking country for a short period has a less significant impact on promoting a more diverse use of trigrams than long-term residency. Although the dominance of VP fragments may appear similar, they were less varied and more skewed in subtypes for S students.

In terms of dependent clause fragments, the differences were more pronounced. Although the overall usage of dependent clause fragments appeared similar for both groups (L: 8.8%, S: 8.5%), L students used four of the five subtypes evenly, while S students mainly used one subtype (2b: *when I speak, when I was*). This differs from the

study by Chung and Lee (2020), where English learners with little to no experience living in English-speaking countries did not use 2b expressions. Considering that the overall usage of dependent clause expressions was much lower than that of native Korean speakers, it suggests that English learners may struggle with dependent clause fragments. Therefore, there is a need to expose English learners to various dependent clause fragments and provide opportunities for practising them to improve their speaking skills.

The most significant difference was observed in NP and PP fragments. S students rarely used NP and PP fragments (7.0%), producing only two subtypes (3b: a *little bit about*, 3c: *something bad for*), while L students used them extensively (19.8%) and across all five subtypes. This suggests that longer residency in an English-speaking country helped L students use NP and PP fragments more frequently and diversely. Since dependent clause fragments had a similar distribution but lower overall usage, it is crucial to introduce more frequent NP-PP fragments in English classes. English teachers should first teach high-frequency NP, and PP fragments to students without experience living in English-speaking countries and students with short-term residency experiences and then observe the proportion of each expression in their English speech. Further research can investigate whether exposure to these expressions during residency and explicit instruction in school contribute to increased usage proportions.

Table 5. Structural types of trigrams

| ***Structural Types*** | L (Normalised Frequency) | S (Normalised Frequency) | Total (Normalised Frequency) | Example |
|---|---|---|---|---|
| **1. LB with VP fragments** | **367 (21,396) 71.4%** | **109 (18623) 84.5%** | **476 (40,019) 74.0%** | |
| 1a. (connector +) 1st/2nd person pronoun + VP fragments | 254 (14,808) 49.4% | 91 (15,548) 70.5% | 345 (30,356) 53,7% | *I would say I have to* |
| 1b. (connector +) 3rd person pronoun + VP fragments | 46 (2,682) 8.9% | 7 (1,196) 8.2% | 53 (3,878) 8.2% | *it's a and there are* |
| 1c. Discourse marker + VP fragment | 23 (1,341) 4.5% | - | 23 (1,341) 3.6% | *I believe that I think it* |
| 1d. VP (with the non-passive verb) | 41 (2,390) 8.0% | 11 (1,879) 8.5% | 52 (4,270) 8.1% | *be able to able to speak* |
| 1e. VP with a passive verb | - | - | - | - |
| 1f. *yes-no* question fragments | 3 (175) 0.6% | - | 3 (175) 0.5% | *do you mean* |
| 1g. Wh-question fragments | - | - | - | - |
| **2. LB with dependent clause fragments** | **45 (2,623) 8.8%** | **11 (1,879) 8.5%** | **56 (4,503) 8.7%** | |
| 2a. 1st/2nd person pronoun + dependent clause fragment | - | - | - | - |
| 2b. Wh-clause fragments | 13 (758) 2.5% | 11 (1,879) 8.5% | 24 (2,637) 8.7% | *when I was but when I* |
| 2c. *If*-clause fragments | 6 (350) 2.5% | - | 6 (350) 3.7% | *if I wanna so if I* |
| 2d. (verb/adjective+) *to*-clause fragment | 23 (1,341) 4.5% | - | 23 (1,341) 3.6% | *to read and to use English* |
| 2e. *that* clause fragments | 3 (175) 0.6% | - | 3 (175) 0.5% | *that English is* |
| **3. LB with NP and PP fragments** | **102 (5,946) 19.8%** | **9 (1,538) 7.0%** | **111 (7,484) 17.3%** | |
| 3a. (connector +) NP with *of-* phrase fragments | 32 (1,866) 6.2% | - | 32 (1,866) 5.0% | *a lot of kind of like* |
| 3b. NP with other post-modifier fragments | 19 (1,108) 3.7% | 3 (513) 2.3% | 22 (1,620) 3.4% | *English in my my child to* |
| 3c. Other NP expressions | 18 (1,049) 3.5% | 6 (1,025) 4.7% | 24 (2,074) 3.7% | *Korean and I Korean and English* |
| 3d. PP expressions | 30 (1,749) 5.8% | - | 30 (1,749) 4.7% | *in the Philippines of my friends* |
| 3e. Comparative expressions | 3 (175) 0.6% | - | 3 (175) 0.5% | *a bit easier* |
| **Total** | **514 (29,966) 100.0%** | **129 (22,040) 100.0%** | **643 (52,006) 100.0%** | |

When examining the structural vocabulary patterns used by English learners, it was evident that L students exhibited trigram usage patterns more similarly to those of native speakers than S students. In other words, the experience of living in an English-speaking country for more than 5 years seemed to have positively impacted English learners in terms of language development. This impact persisted even after several years of living in Korea, where they had limited opportunities to use English.

While exploring methods for applying native-speaker language patterns to English learners is essential, it is equally crucial to leverage the language patterns of long-term residency English learners who did not significantly differ from native speakers during their Korean-language environment experiences. This could be useful in designing learning materials for beginning English learners.

S students may appear to have lower proficiency in trigrams' quantity and diversity than L students. However, their English proficiency in general was not lower, even though their self-ratings were indeed lower. Although they resided in English-speaking countries for a shorter period, they primarily went there for educational purposes. This would make a significant difference for L students who might not intend to learn English, often accompanying their parents. That intent and attitude concerning residency abroad may play an essential role in language development. Therefore, it is encouraged that circumstances and motivations of language learning be more carefully scrutinised in future studies, as there could be underlying differences in their actual English language learning experiences and overall academic abilities among groups. This study did not objectively compare the two groups using the same English tests or overall academic achievements, which should be considered in future research. Examining how evidence from explicit learning compares with trigram usage would be beneficial.

In addition, research is needed to determine whether S students would begin using previously unused expressions when introduced and taught. This could help understand whether their usage results from implicit exposure during their stay in English-speaking countries or explicit instruction from an English classroom or textbook. Subsequent research could explicitly group learners into long-term, short-term, and inexperienced categories, teach them target dependent clause fragments, NP-PP fragments, etc., and then analyse the interview results to determine the impact of explicit instruction.

In summary, this study's findings highlight the influence of living in an English-speaking country on the structural patterns of English learners' trigrams. It emphasises the importance of considering the duration of residency in English language learning experiences. Future research should focus on comparing different groups of English learners, examining the impact of motivation, and investigating the effectiveness of explicit instruction in using specific expressions. These insights could contribute to more effective English language teaching and curriculum development for diverse learner backgrounds.

**5 Conclusion**

This study compared the usage of English words and trigrams of university students who had notable lengths of residency in English-speaking countries, followed by enrolling in Korean universities and learning English in the same environment as typical English learners. The groups were divided based on the duration of their residency: students who had lived long-term in English-speaking countries and those who had lived for a relatively short period. The study did not explore the impact of individual motivation on what made them reside in English-speaking countries, either their willingness to learn English or accompanying their parents. Still, it was anticipated that the effects of the residency duration would be substantial even with a similar level of English proficiency; therefore, the study focused on these factors.

A total of 10 participants, including six L students (3 males, three females) and four S students (all females), participated in interviews. With consent, the recorded interviews were approximately 30 minutes long. Still, there was a notable difference in word count and trigram usage between the groups, with L students exhibiting significantly greater word count and trigram diversity. Moreover, L students used a wider variety of structural patterns, much like English native speakers, including both VP and NP-PP fragments. In contrast, S students strongly preferred VP fragments, consistent with previous research on English learners in Korea. Interestingly, there was a greater lexical diversity compared to their L students. Despite similarities in English proficiency or reported test scores, these differences suggest the need to examine the specific characteristics of each group to tailor effective education and learning strategies.

While this study provides valuable insights into the impact of residency duration on the structural patterns of trigrams, it has certain limitations. The limitations include the small sample size, lack of male participants, and exclusion of groups such as those without experience living in English-speaking countries. As more data about lexical bundles are unearthed, subsequent research should also involve using bigrams and quadrigrams for more in-depth exploration. However, the study's significance lies in breaking away from the traditional categorisation of English learners based on proficiency and instead exploring the impact of the duration of residence in English-speaking countries as a variable to compare the frequency and structural types of trigrams. This research calls for further

investigation and encourages adaptations in English language education. Even within similar proficiency levels, nuanced differences can exist based on factors such as past residency experiences, warranting a more detailed classification of English learners and personalised English learning environments.

**Acknowledgements**

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

**References**

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*(3), 371-405.

Biber, D., & Barbieri, F. (2007). Lexical bundles in university spoken and written registers. *ESP, 26*, 263-286.

Chen, Y., & Baker, P. (2010). Lexical bundles in L1 and L2 academic writing. *Language Learning & Technology, 14*(2), 30-49.

Chung, J., & Lee, K. (2021). Structural analysis of lexical-bundles in uncontrolled environment of spoken discourse: The same heritage language groups. *The Journal of ASIA TEFL, 17*(2), 479-492.

Conrad, S., & Biber, D. (2004). The frequency and use of lexical bundles in conversation and academic prose. *Lexicographia, 20,* 56-71.

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.

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

Hyland, K. (2008). As can be seen: Lexical bundles and disciplinary variation. *ESP, 27*, 4-21.

Hyland, K. (2012). Bundles in academic discourse. *Annual Review of Applied Linguistics, 32*, 150-169.

Ishikawa, S. (2015). A consideration of the difference between the spoken and written English of native speakers and Japanese learners: *A corpus-based study. Discourse and Interaction, 8*(1), 37-52.

Kim, J. (2013). Lexical bundles in Korean college students' English essays: A corpus-based comparative study. *English Language & Literature Teaching, 19*(3), 157-179.

Krashen, S. D. (1975). The critical period for language acquisition and its possible bases. *Developmental Psycholinguistics and Communication Disorders, 263*(1), 211-224.

Lee, K. (2016). Diversity among NEST raters: How do new and experienced NESTs evaluate Korean English learners' essays? *The Asia-Pacific Education Researcher, 25*(4), 549-558.

Lenneberg, E. H. (1967). The biological foundations of language. *Hospital Practice*, *2*(12), 59-67.

Polat, B., & Kim, Y. (2014). Dynamics of complexity and accuracy: A longitudinal case study of advanced untutored development. *Applied Linguistics, 34*(2), 184-207.

Southern California Public Radio (March 21, 2012). "Gen 1.5: Where an immigrant generation fits in". Southern California Public Radio. Retrieved August 14, 2023.

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.

Vercellotti, M. L., Juffs, A., & Naismith, B. (2021). Multi-word sequences in English language learners' speech: The relationship between trigrams and lexical variety across development. *System, 98,* 102494. https://doi.org/10.1016/j.system.2021.102494

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.

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.