Korean Journal of English Language and Linguistics, Vol 23, April 2023, pp. 269-284
DOI: 10.15738/kjell.23..202304.269

KOREAN JOURNAL OF ENGLISH LANGUAGE AND LINGUISTICS

ISSN: 1598-1398 / e-ISSN 2586-7474

http://journal.kasell.or.kr

The Effect of Frequency and L2 Proficiency on the Processing of the Multiword Units Extracted from the Corpus of Contemporary American English*

Shinwoong Lee (Hanyang University)

Received: February 16, 2023
Revised: April 12, 2023
Accepted: April 18, 2023

Shinwoong Lee
Professor, Dept. of English Language and Literature, Hanyang University
Tel: 02-2220-0745
Email:shinwoonglee@hanyang.ac.kr

* This paper was supported by the Ministry of Education of the Republic of Korea and the National Research Foundation of Korea (NRF-2020S1A5A2A01044244).

ABSTRACT

Lee, Shinwoong. 2023. The effect of frequency and L2 proficiency on the processing of multiword units extracted from the Corpus of Contemporary American English. *Korean Journal of English Language and Linguistics* 23, 269-284.

The primary purpose of the current study was to explore the psycholinguistic reality of the multiword units extracted from the Corpus of Contemporary American English (COCA) by investigating how the frequency and L2 proficiency influenced the EFL learners' processing of the multiword units. Forty Korean EFL university students and 18 native speakers of English in Korea participated in the current study. The participants were asked to decide whether the multiword units given were appropriate or not, and their reaction time and error rate were measured. It was found that not only the native speakers of English but also both the advanced and intermediate Korean EFL learners processed the high-frequency multiword units significantly faster than the low-frequency multiword units, and the errors rates on the high-frequency multiword units were significantly lower than those of the low-frequency multiword units even if there existed a significant difference in the reaction time and the error rate between the groups. The findings suggested that some of the high-frequency multiword units extracted from the COCA may exist as independent lexical entities in the mental lexicon of Korean EFL learners.

KEYWORDS

multiword units, psycholinguistic reality, EFL learners, frequency, L2 proficiency, reaction time, error rate, COCA

1. Introduction

There has been converging evidence that multiword units are ubiquitous in most human languages, and they make up a relatively large proportion of both spoken and written languages in general (Biber 2009, Erman and Warren 2000, Nattinger and Decarrico 1992). It has been argued that the efficiency of language processing is mainly due to the lexical patterns or chunks stored in the long-term memory of language users because one can save cognitive resources by processing them holistically as an independent unit instead of processing every single element that comprises them (Schmitt and Carter 2004, Wray 2002). According to the pattern-based model of language acquisition, what human beings are born with are not innate principles or parameters but the proclivity to extract patterns from language input they receive and to store them in their long-term memory (Ellis 2002). It has been further contended that once those language patterns are stored as an independent unit, people can have the native-like selection of words and use appropriate grammatical patterns automatically while bypassing the necessity of proper grammatical and lexical choices (Pawley and Syder 1983). In this way, native speakers are typically equipped with "a platform for fluent and accurate output" (Conklin and Schmitt 2012) through the process of lexical chunking or pattering, which explains well why and how they can comprehend and produce language so fluently and accurately.

As it becomes increasingly apparent that lexical patterning or chunking plays a crucial role in the efficiency of L1 comprehension and production, L2 researchers have begun to explore whether there exists a psycholinguistic reality in processing L2 multiword units (MWUs; hereafter), including formulaic sequences and collocations, and some of the studies (e.g., Conklin and Schmitt 2008, Jiang and Nekrasova 2007, Lee 2021) found that L2 learners may be able to process L2 MWUs in a similar way as native speakers do, supporting the existence of psycholinguistic reality of L2 MWUs, and this was especially true with high-frequency MWUs (e.g., Kim and Kim 2012, Lee 2021). On the other hand, it was also found that the processing of L2 MWUs was strongly influenced by L1 knowledge. For example, the L2 collocations that were predictable from L1 were processed faster than the unpredictable ones (Lee 2016, Yamashita and Jiang 2010). In a similar vein, it was reported that L1 often became the source of the errors in L2 collocation production (e.g., Laufer and Waldman 2011, Nesselhauf 2003). Based on research findings, it has been suggested that L2 collocation knowledge should be emphasized in L2 instruction, and especially the ones unpredictable from L1 should be explicitly taught in L2 classrooms (Lee 2016).

As such, several studies have centered on L2 collocation learning and processing as the knowledge of MWUs has been found essential to L2 fluency. However, a relatively small number of studies were conducted on the psycholinguistic reality of L2 MWUs in EFL contexts (e.g., Han 2019, Lee 2021, Yamashita and Jiang 2010). Besides, only a few studies attempted to investigate whether the MWUs extracted from native corpora can be applied to EFL learners. In this vein, the primary purpose of the current study was to explore the psycholinguistic reality of the MWUs extracted from the Corpus of Contemporary American English (COCA; hereafter) with Korean EFL learners.

2. Literature Review

2.1 Multiword Units

A multiword unit, as Nation and Webb (2011) stated, is probably a kind of blanket term to cover a number of different names of lexical phrases, including collocations, formulaic sequences, lexical bundles, and idioms, and one of the properties that all of these terms share in common is probably the fact that they consist of at least two

words or more. Among these terms, formulaic sequences and collocations are most often utilized in L1 and L2 experimental research (e.g., Conklin and Schmitt 2008, Jiang and Nekrasova 2007, Lee 2016, Yamashita and Jiang 2010, Wolter and Yamashita 2018).

Wray (2002) defined a formulaic sequence as "a sequence, continuous or discontinuous, of words or other elements, which is, or appears to be, prefabricated: that is, stored and retrieved whole from memory at the time of use, rather than being subject to generation or analysis by the language grammar" (p. 9). She further argued that formulaic sequences do not always conform to standard constituents. In other words, they can be grammatically incomplete (e.g., *on the basis of*). According to her, what counts important in defining formulaic sequences is whether they are stored as a whole and thus can be retrieved as an independent unit, and have little to do with being analyzed and generated by language rules.

It has also been argued that formulaic sequences are essential in relation to their pragmatic value (Schmitt 2010). Language users often draw on formulaic expressions daily to meet recurrent communication needs. For example, people use "Could you do...?" to request and "I feel sorry to hear..." to express sympathy. As such, many of the routinized expressions used daily consist of formulaic sequences, and they are often essential for successful communicative functions in our daily lives.

Meanwhile, collocation is often defined as a combination of words in a language that occurs more frequently than expected if words are combined randomly (Nesselhauf 2003). The definition of collocation and formulaic sequence are similar to some extent in that the combination of the words are more frequent than random probability. However, the experimental definitions of collocation in many L2 collocation studies (e.g., Durrant and Doherty 2010, Lee 2016, Wolter and Gyllstad 2011, Wolter and Yamshita 2018) are somewhat different from that of formulaic sequences in that collocations typically have grammatical completeness (e.g., *have a decision*, *heavy rain*), whereas formulaic sequences do not necessarily have one (e.g., *in the middle of*, *get out of*). In addition, most studies on formulaic sequences utilized three- or four-word sequences (e.g., Conklin and Schmitt 2008, Jiang and Nekrasova 2007), whereas the studies on collocations typically drew on two-word sequences in the experiments (e.g., Durrant and Doherty 2010, Lee 2021, Yamashita and Jiang 2010, Wolter and Gyllstad 2013, Wolter and Yamashita 2018).

It appears that the terms of MWUs, such as collocations, formulaic sequences, and lexical bundles, are not to be defined easily. It is almost impossible to differentiate each other clearly since they share parts of their definitions in common, and there does not seem to be a clear boundary among the concepts. In this vein, Nation and Webb (2011) stated, "It is highly unlikely that there will be a standardization of the terminology used for multiword units" (p. 179). They further contended that it might be more important to describe the experimental definition of MWUs and how they are collected and make sure that the rules for data collection and analyses are consistently applied.

2.2 Studies on the Processing of Multiword Units

As a usage-based language acquisition or learning model receives much attention from L1 and L2 researchers, a number of studies (Conklin and Schmitt 2008, Ellis Simpson-Vlach and Maynard 2008, Jiang and Nekrasova 2007, Kim and Kim 2012) have delved into the processing of L1 or L2 MWUs and attempted to find out whether they truly have a psycholinguistic reality and thus expedite language comprehension and production. This line of thought has enjoyed popularity because it can explain well where language fluency originates from and how people become fluent in languages, and thus sparked several studies on the processing of MWUs. Overall, the findings of the studies (e.g., Conklin and Schmitt 2008, Jiang and Nekrasova 2007, Kim and Kim 2012, Lee 2021, Wolter and Yamashita 2018) seemed to support the idea that MWUs are stored in our brain as independent lexical units, and

they are processed holistically.

Conklin and Schmitt (2008) examined the effect of formulaic sequences by measuring the reading time of the text with idioms or formulaic sequences. They found that there existed processing advantages for the formulaic sequences over the nonformulaic ones, and both native and nonnative speakers of English appeared to have a processing advantage for the formulaic sequences. Also, Kim and Kim (2012) investigated the effect of frequency on the processing of L2 phrasal verbs drawing on a self-paced reading task with native and nonnative speakers of English in ESL settings. The results revealed a frequency effect on the processing of two-word MWUs in both groups. However, while the native speakers of English could process most of the MWUs as a chunk, the nonnative speakers of English could only process high-frequency MWUs holistically. The researchers attributed the results to the fact that even if most participants were advanced L2 learners, they spent a relatively short time in the U.S. Thus, the input they received was not adequate to make those MWUs processed holistically. The findings also implied that the formulaicity of MWUs may be related to how frequently they are used.

Along the same line, Jiang and Nekrasova (2007) investigated the processing of MWUs by measuring the reaction time and the error rate on the formulaic and nonformulaic sequences with the native and the nonnative speakers of English in ESL settings. They found that the native speakers of English processed both formulaic and nonformulaic sequences significantly faster with a lower error rate than the nonnative speakers of English. Still, it was also revealed that both groups processed the formulaic sequences faster with fewer errors than the nonformulaic sequences. The results supported the holistic nature of the processing of formulaic sequences. They implied that the psycholinguistic reality of formulaic sequences might exist in the mental lexicon of both native and nonnative speakers of English.

On the other hand, some of the studies (e.g., Schmitt Grandage and Adolphs 2004, Schmitt and Underwood 2004, Underwood Schmitt and Galpin 2004) on the processing of MWUs did not seem to support the holistic nature of MWUs, especially with nonnative speakers of English. Schmitt, Gradage and Adolphs (2004) examined the corpus-derived MWUs to see whether they could be processed as a chunk drawing on an oral dictation task. The findings suggested that the nonnative speakers of English did not appear to process MWUs holistically. Along the same line, Underwood, Schmitt and Galpin (2004) investigated how formulaic sequences are processed by the native and the nonnative speakers of English drawing on an eye-tracking method. It was revealed that the nonnative speakers of English used fewer but not shorter fixation points at the last word of MWUs, implying that the data partially supported the automatic processing of the MWUs. As such, some of the findings of the studies on MWUs were somewhat inconclusive, especially with nonnative speakers of English. Jiang and Nekrasova (2007) argued that the conflicting results of those studies on the MWUs were due to the materials used in the experiments. They contended that some of the materials used in the experiments were not formulaic sequences and were possibly unknown to the nonnative participants. They further maintained that utilizing an oral dictation task or measuring reading times in a self-paced reading task might not be sensitive enough to capture a formulaic effect on the processing. Rather online grammaticality judgment would be more sensitive to detect it.

In the meantime, some of the researchers attempted to examine how the frequency and the strength of the association between the words influence the processing of MWUs. For example, Durrant and Doherty (2010) investigated how the frequency of collocations and the strength of association measured by mutual information (MI) are associated with the processing efficiency of MWUs with native speakers of English. They found that even if the strength of association played a role in the efficiency of collocation processing to some extent, the frequency of collocations appeared to be a significant contributor to the psychological reality of the collocations. Similarly, Ellis, Simpson-Vlach and Maynard (2008) examined how the frequency of formulaic sequences and the strength of associations between the words would influence the recognition of formulaic sequences utilizing a multiple regression analysis. They found that the strength of association between the words measured by MI scores was

more influential on the processing of formulaic sequences for the native speaker of English. In contrast, the frequency of formulaic sequences was more influential on the processing for the nonnative speakers of English.

More relevant to the current study, Wolter and Yamashita (2018) investigated the effect of frequency, L1 congruency, and L2 proficiency on the L2 collocation processing with Japanese EFL learners. They found that the Japanese EFL learners processed the congruent L2 collocations faster than the incongruent ones in general, and there was a significant effect of collocation and word frequency on the processing of L2 collocations in all of the groups. Similarly, Lee (2021) examined the effect of L2 proficiency, collocation frequency, and L1 predictability on the processing of L2 collocations in EFL contexts by measuring the reaction time and the error rate. It was found that collocation frequency was the strongest factor that influenced L2 collocation processing, followed by L2 proficiency and L1 congruency. The findings suggested that L1 influence on L2 collocation processing wanes as L2 proficiency enhances.

Even if a relatively large number of studies were conducted on the processing of L2 MWUs, many of them had only one group of L2 learners who were largely graduate or undergraduate students in English-speaking countries and who were proficient in English in general. Indeed, only a few studies (e.g., Wolter and Yamashita 2018) attempted to investigate the processing of MWUs with EFL learners with different English proficiency levels. Besides, as aforementioned, not many studies (e.g., Han 2019, Kim and Kim 2012) have attempted to investigate the processing of MWUs extracted from native corpora with EFL learners. Thus, it seems necessary to further examine the processing of L2 MWUs with diverse EFL learners drawing on various L2 MWUs. The followings are the research questions of the current study.

1) How does L2 proficiency affect the processing of the L2 multiword units extracted from the COCA?
2) How does frequency affect the processing of the L2 multiword units extracted from the COCA?
3) Is there an interaction effect between L2 proficiency and frequency on the processing of the L2 multiword units extracted from the COCA?

3. Research Method

3.1 Participants and Settings

There were 40 Korean EFL learners and 18 native speakers of English who participated in the current study. All of the Korean EFL learners were undergraduate or graduate students at a university located in Seoul, Korea. They were from 10 different departments at the university, and those who lived in English-speaking countries for more than a year were excluded from the experiment. Only the participants who went from the elementary school to the university in Korea were included to control the degree of exposure to English input. The native speakers were the instructors at the university who taught English presentation skills and academic writing or the instructors at a private language institute in Seoul. The Korean EFL learners were assigned to two groups according to their TOEIC (Test of English for International Communication) scores. Following the guidelines of ETS (Educational Testing Service), 20 EFL learners with scores above 910 were grouped as advanced EFL learners who can communicate effectively in any communicative situation. Meanwhile, 20 EFL learners who had scores between 750 and 850 were classified as intermediate EFL learners who are expected to be able to meet social demands and work requirements in general according to the ETS guidelines. The average TOEIC score of the advanced EFL learners was 955, and that of the intermediate EFL learners was 815. The participants' TOEIC scores were accepted only when they took the test within two years from the day of the experiment. The background information about the

participants is shown in Table 1 below.

Table 1. Background Information about the Participants (Standard Deviation in Parentheses)

Group	Number	Gender (male/female)	Age	TOEIC	Onset of learning English (age)	Duration of learning English (years)
Native	18	8/10	38 (5.5)	N/A	N/A	N/A
Advanced	20	9/11	24 (2.8)	955 (21.5)	9.2 (2.3)	13.8 (2.1)
Intermediate	20	10/10	23 (2.5)	815 (28.2)	10.3 (1.9)	12.5 (2.3)

3.2 Material Development and Experimental Design

In the current study, multiword units (MWUs) were defined as "three or four consecutive words that may have or may not have grammatical completeness but that have meaningful wholeness" (e.g., *at the same time*, *take advantage of*). In selecting the MWUs for the experiment, the 3-gram and 4-gram data from the COCA were utilized, and also some of MWUs were adopted from the materials used in the previous studies (i.e., Cortes 2004, Jiang and Nekrasova 2007). Still, the frequency of each MWU was obtained using the "search" function built in the COCA. The criterion of high-frequency MWUs was set at higher than 10,000 hits approximately in the COCA. Once the 30 high-frequency MWUs were selected, the low-frequency MWUs were constructed by replacing one of the words in the high-frequency MWUs (e.g., *in other words* vs. *in other games*). The mean frequency of the high-frequency MWUs was 16,229, and that of the low-frequency MWUs was 365. An independent *t*-test was conducted to see whether there existed a significant difference between them with the log-transformed frequencies, and a significant difference was found between the high- and low-frequency MWUs (t=19.52, $p<.05$).

Also, when pairing the high-frequency MWUs and their counterparts (i.e., the low-frequency MWUs), it was attempted to minimize the difference in the frequencies of the individual words as much as possible. For example, the frequency of "forward" in the phrase "look forward to" was 78,035, and that of "similar" in the phrase "look similar to" was 78,722 in the COCA. The mean frequency of the different words in the high-frequency MWUs was 252,321 and that of the low-frequency MWUs was 246,693 respectively. An independent *t*-test was conducted with the log-transformed frequencies to see whether a significant difference existed between them, and no significant difference was found (p=.97). It was also attempted to match the number of the letters of the phrases as closely as possible. The average number of the letters of high-frequency MWUs was 11.5, and that of low-frequency MWUs was 11.7. Additionally, an independent *t*-test was conducted to examine the difference in the length of the two types of phrases statistically, and no significant difference was found between them (p=.92).

Regarding the experiment, the two stimulus lists were created. List 1 included the first half of the high-frequency MWUs and the second half of the low-frequency MWUs, and List 2 had the second half of the high-frequency MWUs and the first half of the low-frequency MWUs. In so doing, it was ensured that if a high-frequency MWU (e.g., in the middle of) was in List 1, then its low-frequency counterpart (e.g., in the space of) was assigned to List 2 to control a repetition effect, while the same number of the high-frequency MWUs and the low-frequency MWUs were included in each list (See Appendix A). Also, an equal number of participants in each proficiency level (i.e., native, advanced, and intermediate) were randomly assigned to one of the two order sets (List 1 and 2 or List 2 and 1). Half of the participants in each level were presented with List 1 first and then List 2, and the other half were presented with List 2 and then List 1 to minimize the order effect. Additionally, 50 ungrammatical phrases (e.g., *when at of*) were created, and 25 of them were used as the fillers in each stimulus set.

3.3 Experiment Procedures and Data Analysis

The experiments were carried out in the researcher's office room at a quiet time. Once the participants entered the office room, they were explained how the experiment would proceed and what they needed to do during the experiment. Then they had a practice session with 30 items in the same format as the experiment. This was intended to have the participants understand what they needed to do in the experiment clearly. After the practice session, the experiment was conducted with the test items, which were randomly presented at the center of the computer screen. There was a fixation point between each stimulus for 400 milliseconds, and the participants were asked to judge whether the phrase presented was appropriate or not and hit the assigned key (i.e., yes or no) as quickly as possible.

Once they were finished with the experiment, they were asked if there existed any cases in which they hit the key mistakenly during the experiment. For example, some of the participants reported that they hit 'yes' even though they believed the phrase given was ungrammatical. This was partially due to the instruction that they should hit the key as quickly as possible, and such cases were excluded from the data analysis. Psychopy, a tool for psycholinguistic experiments and analysis, was used for the experiments, and the reaction time and the error rate for each item and person were calculated using the Excel program. In addition, the reaction times which were higher than 2.5 standard deviations away from the mean score of each participant were assumed to be outliers and thus excluded from the analysis.

As for the statistical analysis, a 3 x 2 repeated-measures analysis of variance (ANOVA) was conducted and the effect of phrase frequency and participants' English proficiency on the reaction time and the error rate was examined. In the analyses, the participants' English proficiency was entered as a between-subject variable, and the frequency of the MWUs as a within-subject variable. Also, one-way ANOVAs and paired *t*-tests were conducted for finer analyses.

4. Results

4.1 Reaction Time

As displayed in Table 2 below, the means and the standard deviations of reaction times of the three groups (native speakers, advanced EFL learners, and intermediate EFL learners) were calculated in the two types of MWUs (i.e., low-frequency and high-frequency MWUs). It was found that the native speakers processed both high-frequency and low-frequency MWUs faster than the advanced and intermediate EFL learners, and the advanced EFL learners processed them faster than the intermediate EFL learners. It was also revealed that the high-frequency MWUs were processed faster than the low-frequency MWUs in all groups (See Figure 1).

Table 2. Reaction Time (sec)

	Native Speakers		Advanced EFL Learners		Intermediate EFL Learners	
	Mean	*SD*	*Mean*	*SD*	*Mean*	*SD*
High-frequency MWUs	.86	.16	1.02	.13	1.23	.29
Low-frequency MWUs	.99	.25	1.24	.31	1.45	.32

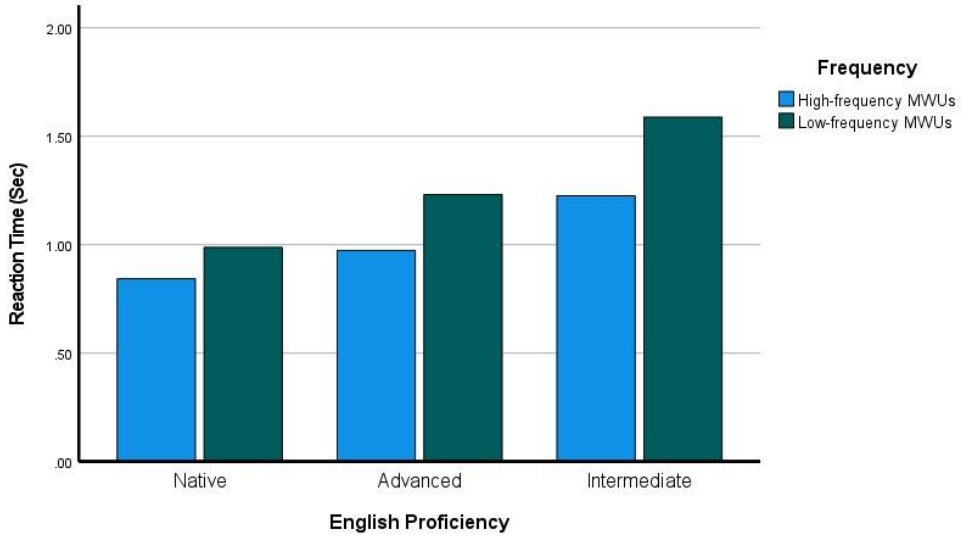

Figure 1. English Proficiency and Phrase Frequency on Reaction Time

In order to examine the effect of English proficiency and phrase frequency on the reaction time statistically, a 3 x 2 repeated-measures ANOVA was conducted. The participants' English proficiency was entered as a between-subject variable, and the frequency of MWUs as a within-subject variable. The results of the analysis are shown in Table 3 below.

Table 3. Results of a 3 x 2 Repeated-measures ANOVA (Reaction time)

	F	*sig.*	η^2
Between-Subject Variable			
English proficiency	19.01	.000*	.41
Within-Subject Variable			
Phrase Frequency	28.93	.000*	.35
Phrase Frequency x English proficiency	.504	.607	

*p<.05

As presented in Table 3, it was found that there existed a significant effect of the participants' English proficiency (F=19.01, p<.05) and the frequency of MWUs on reaction time (F=28.93, p<.05), but no significant interaction effect between them was found (p>.05) suggesting that the effect of frequency on reaction time was not conditional upon the level of English proficiency. The effect size of English proficiency (η^2=.41) was higher than that of phrase frequency (η^2=.35), implying that the influence of English proficiency on reaction time was stronger than that of the frequency of MWUs. For a finer analysis of a between-subject variable, a one-way ANOVA between the groups was conducted at each level of frequency (i.e., high and low). It was found that there existed a significant difference in reaction time between the groups in both high-frequency (F=16.64, p<.05) and low-frequency (F=11.80, p<.05) levels as shown in table 4 below.

Table 4. Results of a One-way ANOVA Between the Groups in Each Level of Frequency (Reaction Time)

	F	*sig.*	η^2
High-frequency MWUs	16.64	.00*	.38
Low-frequency MWUs	11.80	.00*	.30

*p<.05

As a post-hoc procedure, Tukey's HSD tests were also conducted between the groups at each level. As to the reaction time to the high-frequency MWUs, there existed a significant difference between the native speakers and the advanced EFL learners (p<.05), the native speakers and the intermediate EFL learners (p<.05), and the advanced EFL learners and the intermediate EFL learners (p<.05). Meanwhile, concerning the reaction time to the low-frequency MWUs, a significant difference was found between the native speakers and the advanced EFL learners (p<.05) and the native speakers and the intermediate EFL learners (p<.05), but no significant difference was found between the advanced EFL learners and the intermediate EFL learners (p>.05).

In addition, a series of paired sample *t*-tests were conducted to examine whether a significant difference in reaction time existed between the high-frequency MWUs and the low-frequency MWUs in each group. It was found that a significant difference in reaction time existed in all of the groups (See Table 5 below), and the results indicated that not only the native speakers of English but also both groups of EFL learners processed the high-frequency MWUs significantly faster than the low-frequency MWUs.

Table 5. Results of Paired Sample *t*-tests Between the High-frequency MWUs and the Low-frequency MWUs in Each Group (Reaction Time)

Phrase Frequency (high vs. low)	*t*	*sig.*
Native Speakers	2.74	.01*
Advanced EFL learners	3.37	.00*
Intermediate EFL learners	3.30	.00*

*p<.05

To recapitulate the findings of the experiments on the reaction time to the high-frequency MWUs and the low-frequency MWUs, it was revealed that the native speakers of English processed both the high-frequency and the low-frequency MWUs significantly faster than both groups of the EFL learners, and the advanced EFL learners processed the high-frequency MWUs significantly faster than the intermediate EFL learners. Meanwhile, there was no significant difference in reaction time to the low-frequency MWUs between the advanced and the intermediate EFL learners. It was also found that not only the native speakers of English but also both the advanced and the intermediate EFL learners processed the high-frequency MWUs significantly faster than the low-frequency MWUs. The results of the experiment implied that both groups of EFL learners might be able to process the high-frequency MWUs holistically as a chunk, even if their processing efficiency was not as high as that of the native speakers of English.

4.2 Error Rate

As presented in Table 6, the means and the standard deviations of error rates of the three groups (native speakers, advanced EFL learners, and intermediate EFL learners) were calculated for the high-frequency MWUs and the

low-frequency MWUs. It was revealed that the error rate of the native speakers was the lowest, followed by the advanced and the intermediate EFL learners in both the high-frequency and the low-frequency MWUs. It was also found that the high-frequency MWUs were processed with lower error rates than the low-frequency MWUs in all groups, which was similar to the result of reaction time (See Table 6 and Figure 2 below).

Table 6. Error Rate

	Native Speakers		Advanced EFL Learners		Intermediate EFL Learners	
	Mean	*SD*	*Mean*	*SD*	*Mean*	*SD*
High-frequency MWUs	.020	.026	.035	.029	.065	.056
Low-frequency MWUs	.042	.031	.073	.041	.114	.060

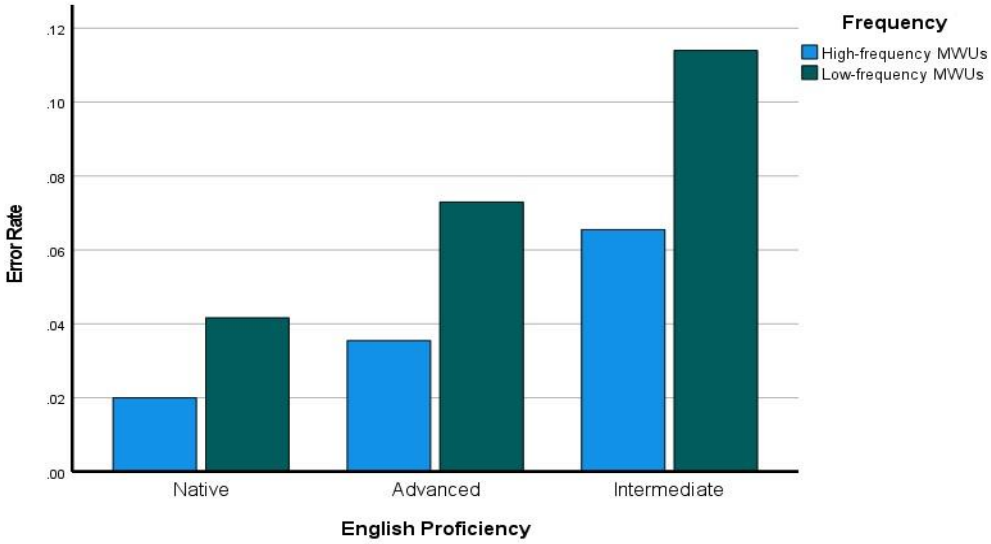

Figure 2. English Proficiency and Phrase Frequency on Error Rate

In order to examine the effect of English proficiency and phrase frequency on the error rate statistically, a 3 x 2 repeated ANOVA was performed. As in the case of reaction time, the participants' English proficiency was entered as a between-subject variable and the frequency of the MWUs as a within-subject variable. The results of the analysis are displayed in Table 7 below.

Table 7. Results of a 3 x 2 Repeated-measures ANOVA (Error Rate)

	F	*sig.*	η^2
Between-Subject Variable			
English proficiency	15.61	.00*	.36
Within-Subject Variable			
Phrase Frequency	23.71	.00*	.35
Phrase Frequency x English proficiency	1.01	.34	

*$p<.05$

As shown in Table 7, it was found that there existed a significant effect of English proficiency (F=15.61, $p<.05$) and phrase frequency (F=23.71, $p<.05$) on the error rate. However, similar to the findings in reaction time, no significant interaction effect between English proficiency and phrase frequency was found ($p>.05$), indicating that the effect of frequency on the error rate was not contingent upon the level of English proficiency. Meanwhile, the effect size of English proficiency (η^2=.36) was similar to that of frequency (η^2=.35), implying that the influence of English proficiency on the error rate was similar to that of the frequency of MWUs.

In addition, a one-way ANOVA between the groups was conducted at each frequency level, and it was found that there existed a significant difference in the error rate between the groups in both the high-frequency (F=6.54, $p<.05$) and the low-frequency (F=11.64, $p<.05$) MWUs as shown in table 8 below.

Table 8. Results of a One-way ANOVA Between the Groups in Each Level of Frequency (Error Rate)

	F	*sig.*	η^2
High-frequency MWUs	6.54	.00*	.20
Low-frequency MWUs	11.64	.00*	.30

*$p<.05$

Tukey's HSD tests, as a post-hoc procedure, were also conducted between the groups at each level of frequency. As to the error rate on the high-frequency MWUs, there was no significant difference between the native speakers(NS) and the advanced EFL learners ($p>.05$), but existed a significant difference between the native speakers and the intermediate EFL learners ($p<.05$) and the advanced EFL learners and the intermediate EFL learners ($p<.05$) as well. Meanwhile, regarding the error rates on the low-frequency MWUs, a significant difference was not found between the native speakers and the advanced EFL learners ($p>.05$), but it was found between the native speakers and the intermediate EFL learners ($p<.05$) and the advanced EFL learners and the intermediate EFL learners ($p<.05$) which was similar to the result of the high-frequency MWUs.

In addition, a series of paired sample t-tests were conducted to examine whether a significant difference in the error rate existed between the high-frequency MWUs and the low-frequency MWUs in each group, and a significant difference in the error rate was found in all of the groups as shown in Table 9 below.

Table 9. Results of Paired Sample *t*-tests Between the High-Frequency and the Low-Frequency MWUs in Each Group (Error rate)

(formulaic vs. nonformulaic)	*t*	*sig.*
Native Speakers	2.51	.02*
Advanced EFL learners	3.55	.00*
Intermediate EFL learners	2.88	.01*

*$p<.05$

To summarize the findings of the error rate on the high-frequency and the low-frequency MWUs, there was a significant effect of English proficiency and phrase frequency on the error rate, but no significant interaction effect was found between English proficiency and phrase frequency. It was also revealed that the error rate of the native speakers was the lowest on both the high-frequency and the low-frequency MWUs, followed by the advanced and the intermediate EFL learners. However, the difference in the error rate between the native speakers and the advanced EFL learners was not statistically significant, while it was significant between the advanced and the intermediate EFL learners. Also, as to the difference in the error rate within the group, a significant difference was found in all of the groups, showing that they all processed the high-frequency MWUs with significantly fewer errors than the low-frequency MWUs. The findings were compatible with those of reaction time in that there existed an advantage for the processing of the high-frequency MWUs in comparison with the low-frequency MWUs in all of the groups.

5. Discussion

In the current study, it was found that not only the native speakers of English but also both groups of Korean EFL learners could process the high-frequency MWUs significantly faster with fewer errors than the low-frequency MWUs although there existed a significant difference in the reaction time and the error rate between the groups. The results of the current study seem compatible with those of previous studies (e.g., Durrant and Doherty 2010, Kim and Kim 2012, Lee 2021, Wolter and Gyllstad 2013, Wolter and Yamashita 2018) in which the processing advantage of high-frequency MWUs was identified. It was also suggested that some of the high-frequency MWUs extracted from the COCA in the current study could have been perceived as formulaic to both the advanced and the intermediate EFL learners, while most of the low-frequency MWUs were not. As several studies on the processing of the MWUs (e.g., Durrant and Doherty 2010, Ellis Simpson-Vlach and Maynard 2008, Kim and Kim 2012) suggested, the frequency could be one of the most critical factors in determining the degree of formulaicity of MWUs, especially to the EFL learners. Then, what may have happened in the current study would be that the EFL learners had to pay more attention to the low-frequency MWUs which were perceived as non-formulaic, inspecting their syntactic well-formedness and semantic appropriateness (Jiang and Nekrasova 2007), whereas it was not a necessary step for the high-frequency MWUs which were perceived as formulaic.

As such, the findings of the current study suggested that the nonnative speakers of English situated in EFL contexts might have some formulaic sequences stored in their mental lexicon, and appeared to support the results of some previous studies (e.g., Ellis Simpson-Vlach and Maynard 2008, Jiang and Nekrasova 2007) in which nonnative speakers of English were able to process formulaic sequences faster with fewer errors than nonformulaic sequences. However, it should be noted that those studies were mostly conducted in ESL contexts where nonnative

speakers of English were typically graduated students at a university in the U.S. and their English proficiency was relatively high, and they were supposedly exposed to a large amount of English input as they used English on a daily basis. In other words, the linguistic environment of those ESL learners who participated in the previous studies (i.e., Ellis Simpson-Vlach and Maynard 2008, Jiang and Nekrasova 2007) seemed very different from that of the participants in the current study. Then, the question would be what contributes to the holistic representation of MWUs in the minds of EFL learners in Korea, where English input is relatively scarce and the use of English is limited. In fact, not only the advanced EFL learners but also the intermediate EFL learners were found to have some degree of efficiency in processing the high-frequency MWUs in the current study. Considering the amount of time of formal English education at school in Korea, it is hardly attributable to the learning at school. Instead, the efficiency that Korean EFL learners in the current study showed in processing high-frequency multiword units may be related to the emergence of the various media platforms that are readily available on the Internet. For example, such media platforms as Youtube, Netflix, and Apple TV have enabled most EFL learners around the world to access English contents relatively easily. In other words, they can have access to such English contents as News, movies, and dramas through those media platforms at any time and any place as long as they are online. Indeed, the accessibility to English input has significantly increased with the help of online technology, and accordingly, EFL learners seem to have more opportunities to be exposed to a large amount of English input. Furthermore, since they can choose the contents of their own interests through multiple online resources, they can be intrinsically motivated to use them, being exposed to the type of English materials that they prefer for an extended period of time.

Consequently, the use of online multimedia platforms seems to contribute to the enrichment of English learning environments in diverse EFL countries around the world, which is probably why the EFL learners in the current study showed a sensitivity to the high-frequency MWUs. In the meantime, it should also be noted that the intermediate EFL learners, in particular, in the current study showed significantly slower reaction times and higher error rates on the high-frequency MWUs than the other groups. This implies that an explicit introduction or teaching of the high-frequency MWUs in EFL classrooms would be necessary to make them more noticeable to EFL learners (Lee 2021). The very first step would be to unearth frequently used MWUs by referring to such general English corpora as the Corpus of Contemporary American English (COCA) or British National Corpus (BNC) since the frequency information of MWUs from such resources would be very conducive to deciding which MWUs should receive more attention than others. Once the frequently used MWUs are selected, EFL teachers can introduce or even teach them explicitly in the classroom so that EFL learners can pay selective attention to them. Furthermore, it would also be necessary to use those frequently used MWUs in the textbooks so that they become more salient to EFL learners, which should be considered at the beginning stage of the development of the textbooks or teaching materials.

6. Conclusion

The findings of the current study strongly upheld the holistic view of language processing in general, and they implied that not only native speakers of English but also EFL learners might have some English MWUs stored in their mental lexicon holistically as an independent lexical unit, which would facilitate their processing ultimately. As aforementioned, this kind of language processing efficiency that EFL learners have may be attributable to a significant enhancement of the accessibility to a large amount of diverse English input through online technology. Meanwhile, some limitations exist in interpreting the results of the current study. Most importantly, the frequency data of the MWUs extracted from the COCA may not adequately represent the degree of exposure in the minds of

Korean EFL learners because their experiences are largely shaped by the linguistic and cultural environments of Korea. Thus, some of the test items in the current study might be under- or over-represented in the minds of Korean EFL learners (Lee 2021). It should also be noted that the age of the native speakers of English was not controlled tightly, which might have introduced a bias to the findings of the current study. Therefore, the findings of the current study should be interpreted with some caution and not be directly generalized into other EFL contexts.

References

Biber, D. 2009. A corpus-driven approach to formulaic language in English: Multi-word patterns in speech and writing. *International Journal of Corpus Linguistics* 14(3), 275-311.

Conklin, K. and N. Schmitt. 2008. Formulaic sequences: Are they processed more quickly than nonformulaic language by native and nonnative speakers? *Applied Linguistics* 29(1), 72-89.

Conklin, K. and N. Schmitt. 2012. The processing of formulaic language. *Annual Review of Applied Linguistics* 32, 433-445.

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

Durrant, P. and A. Doherty. 2010. Are high-frequency collocations psychologically real? Investigating the thesis of collocation priming. *Corpus Linguistics and Linguistic Theory* 6(2), 125-155.

Erman, B. and B. Warren. 2000. The idiom principle and the open-choice principle. *Text* 20, 29-62.

Ellis, N. 2002. Frequency effects in language processing: A review with implications for theories of implicit and explicit language acquisition. *Studies in Second Language Acquisition* 24, 143-188.

Ellis, N., R. Simpson-Vlach and C. Maynard. 2008. Formulaic language in native and second-language speakers: Psycholinguistics, corpus linguistics, and TESOL. *TESOL Quarterly* 41(3), 375-396.

Han, S. 2019. Frequency, MI, and congruency in collocation processing by Korean EFL learners: Evidence from reading aloud. *Korean Journal of English Language and Linguistics* 19(3), 325-346.

Jiang, N. and T. Nekrasova. 2007. The processing of formulaic sequences by second language speakers. *The Modern Language Journal* 91(3), 433-445.

Kim, S. and H. Kim. 2012. Frequency effects in multiword unit processing: Evidence from self-paced reading. *TESOL Quarterly* 46(4), 831-841.

Laufer, B. and T. Waldman. 2011. Verb-noun collocations in second language writing: A corpus analysis of learners' English. *Language Learning* 61(2), 647-672.

Lee, S. 2016. L1 influence on the processing of L2 collocation: An experimental study of Korean EFL learners. *Linguistic Research* 33(special edition), 137-163.

Lee, S. 2021. The effect of L2 proficiency, L1 congruency, and collocation frequency on L2 collocation processing: An experimental study of Korean EFL learners. *Korean Journal of English Language and Linguistics* 21, 1060-1084.

Nation, P. and S. Webb. 2011. *Researching and Analyzing Vocabulary*. Boston: Heinle.

Nattinger, J. and J. DeCarrico. 1992. *Lexical Phrases and Language Teaching*. Oxford: Oxford University Press.

Nesselhauf, N. 2003. The use of collocations by advanced learners of English and some implications for teaching. *Applied Linguistics* 24(2), 223-242.

Pawley, A. and F. Syder. 1983. Two puzzles for linguistic theory: Nativelike selection and nativelike fluency. In J. Richards and R. Schmidt, eds., *Language and Communication*, 191-225. London: Longman.

Peirce, J. and M. MacAskill. 2018. Building experiments in psychology. London: Sage.

Schmitt, N. 2010. Researching vocabulary: A vocabulary research manual. New York: Palgrave Macmillan.

Schmitt, N. and R. Carter. 2004. Formulaic sequences in action: An introduction. In N. Schmitt, ed., *Formulaic Sequences: Acquisition, Processing and Use* 1-22. Amsterdam: John Benjamins.

Schmitt, N., S. Grandage. and S. Adolphs. 2004. Are corpus-derived recurrent clusters psycholinguistic ally valid? In N. Schmitt, ed., *Formulaic Sequences: Acquisition, Processing and Use* 127-151. Amsterdam: John Benjamins.

Schmitt, N. and G. Underwood 2004. Exploring the processing of formulaic sequences through a self-paced reading task. In N. Schmitt, ed., *Formulaic Sequences: Acquisition, Processing and Use* 173-189. Amsterdam: John Benjamins.

Sinclair, N. 1991. Corpus, concordance, collocation. Oxford: Oxford University Press.

Underwood, G., N. Schmitt. and A. Galpin. 2004. The eyes have it: An eye-movement study into the processing of formulaic sequences. In N. Schmitt, ed., *Formulaic sequences: Acquisition, Processing and Use* 153-172. Amsterdam: John Benjamins.

Wolter, B. and H. Gyllstad. 2011. Collocational links in the L2 mental lexicon and the influence of L1 interlexical knowledge. *Applied Linguistics* 32(4), 430-449.

Wolter, B. and H. Gyllstad. 2013. Frequency of input and L2 collocational processing. *Studies in Second Language Acquisition* 35, 451-482.

Wolter, B. and J. Yamashita. 2018. Word frequency, collocational frequency, L1 congruency, and proficiency in L2 collocational processing. *Studies in Second Language Acquisition* 40, 395-416.

Wray, A. 2002. *Formulaic Language and the Lexicon*. Cambridge: Cambridge University Press.

Yamashita, J. and N. Jiang. 2010. L1 influence on the acquisition of L2 collocations: Japanese ESL users and EFL learners acquiring English collocations. *TESOL Quarterly* 44(4), 647-668.

Examples in: English
Applicable Languages: English
Applicable Level: tertiary

Appendix: Test Items

		High-Frequency MWUs	Low-Frequency MWUs
	1	make it (easy)	make it (real)
	2	in the (middle) of	in the (space) of
	3	at the (end) of	at the (side) of
	4	in the (case) of	in the (study) of
	5	for the (rest) of	for the (role) of
	6	on the (basis) of	on the (trail) of
	7	in (relation) to	in (respect) to
	8	as (soon) as	as (short) as
	9	as a (result)	as a (system)
	10	the most (important)	the most (different)
	11	at this (point)	at this (price)
	12	on the (phone)	on the (paper)
	13	look (forward) to	look (similar) to
	14	by the (way)	by the (day)
	15	on the other (side)	on the other (line)
	16	around the (world)	around the (school)
	17	in the first (place)	in the first (case)
	18	all of a (sudden)	all of a (piece)
	19	fall in (love)	fall in (value)
	20	in other (words)	in other (games)
	21	get (rid) of	get (hold) of
	22	take (advantage) of	take (photographs) of
	23	for a (long) time	for a (short) time
	24	in (favor) of	in (fear) of
	25	for the first (time)	for the first (year)
Average number of frequency (phrase)		16,229	365
Average number of letters		11.5	11.7
Average number of frequency (the word in the blank)		246,693	252,321