Korean Journal of English Language and Linguistics
Vol. 19, No. 3, Fall 2019, 325-346
10.15738/kjell.19.3.201909.325

Frequency, MI, and Congruency in Collocation Processing by Korean EFL Learners: Evidence from Reading Aloud*

Sumi Han
(Hallym University)

Han, Sumi. 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. This study examined intermediate-level Korean EFL learners' sensitivity to collocations versus noncollocations, and frequency, mutual information (MI), and congruency of verb-noun collocations from a psycholinguistic perspective. A read-aloud task was used to investigate the (controlled) oral production of 225 stimuli, 180 collocations and 45 noncollocations, which were classified into 15 stimuli sets with regard to frequency, MI, and congruency. Results of linear mixed-effects modeling showed a processing cost for collocations over noncollocations, which provides counterevidence for Wray's (2002) holistic hypothesis. Significant effects of frequency and congruency of collocations were also found, indicating that more frequent or congruent collocations were orally processed faster than less frequent or incongruent collocations. These findings were further discussed in terms of the usage-based model and phraseology-based tradition as well as methodological and educational implications for future research in the field of L2 collocation processing.

Keywords: collocation, reading aloud, processing, Korean, EFL, intermediate, frequency, mutual information (MI), congruency, holistic hypothesis

1. Introduction

The role of multiword or formulaic sequences, such as collocations, idioms, phrasal verbs, and proverbs, in second language (L2) learning and teaching has received significant attention from researchers and practitioners over the last decades (Schmitt 2004, Webb and Kagimoto 2009). Pawley and Syder (1983) argue that L2 speakers can achieve nativelike fluency and accuracy by storing and using multiword sequences; Instead, a lack of an appropriate amount of such knowledge often results in slow and

* This work was supported by the Ministry of Education of the Republic of Korea and the National Research Foundation of Korea (NRF-2016S1A5B5A07918204). The literature review and the method sections were partly based on the author's unpublished Ph. D. dissertation.

inaccurate utterances by L2 speakers. As described as islands of reliability (Dechert 1983), the use of multiword sequences is beneficial for L2 speakers to maintain the flow of real-time speaking or writing and to compensate for their nonnativeness of L2 use.

More research has examined the representation and processing of multiword sequences. The key issue of this line of research is whether multiword sequences are stored and retrieved as a whole in language processing (Yamashita and Jiang 2010). Wray's (2002) holistic hypothesis behind this issue explains that multiword sequences are stored and retrieved holistically from memory rather than analytically (i.e., bottom-up processes of each word) and this processing efficiency leads to fast and effortless language use. To date, the processing advantage of multiword sequences for first language (L1) speakers and proficient L2 speakers have often been found, but the results differed across L2 proficiency groups (e.g., Jiang and Nekrasova 2007).

The literature also shows that a variety of methodology from target types and features of multiword sequences to experimental tasks, which led to inconclusive results pertaining to formulaic sequence processing. In L2 processing research, collocations as a subtype of multiword sequences, such as verb-noun collocations (e.g., *make a decision*) and adjective-noun collocations (*a heavy smoker*), have been most widely examined (e.g., Wolter and Gyllstad 2011, 2013). It has been generally found that advanced L2 speakers' receptive collocation processing was influenced by phrasal frequency and congruency (word-to-word match of L1-L2). Little has been done with productive collocation processing, other levels of L2 proficiency, or multiple collocational features. Thus, an in-depth understanding of the collocation processing of L2 speakers has been limited (Siyanova-Chanduria and Martinez 2015).

To narrow the gaps in the literature, this study aimed to explore the productive collocation processing of L2 Korean English as a Foreign Language (EFL) learners using a read-aloud task. Along with any processing advantages of collocations over noncollocations, the effects of frequency, mutual information (MI; association strength between words), and congruency were examined. Evidence from this study will be used to offer theoretical, methodological, and educational implications for future research in L2 vocabulary learning and acquisition.

2. Literature Review

2.1 What are Collocations?

Language is highly recurrent and multiword sequences are ubiquitous in language use (Nattinger and DeCarrico 1992). Erman and Warren (2000) found that 58.6% of spoken English discourse and 52.3% of written discourse are made up of recurrent multiword sequences. Using such ready-made, recurrent chunks, language users are more likely to keep the flow in discourse. Ironically, a considerable interest in multiword sequences seems to lead to a lack of a consensus on the terminology of multiword sequences. As Wray (2002) lists, over fifty terms for the multiword combination phenomenon have been used in the literature of linguistics and applied linguistics: both multiword sequences and formulaic sequences have been most widely used as umbrella terms for subtypes such as prefabricated patterns (e.g., *Ladies and Gentlemen*), lexical bundles or ngrams (e.g., *to the door*), idioms (e.g., *hit the books*), or collocations (e.g., *heavy rain*). As the one of the widely examined types of multiword sequences, the term collocations was first coined by Firth (1957), who argued that a word can be judged by the company it keeps.

To identify or classify multiword sequences, distributional information underlying them has been heavily used regardless of their subtypes. In this frequency-based tradition, statistical properties such as frequency and association strength measures are computed and used (Church and Hanks 1990, McEnery, Xiao and Tono 2008). In particular, association strength measures, including mutual information (MI), t-score, and log-likelihood, are also based on frequency counts, showing the degree of association strength between words in the sequence. Among them, MI is the most widely used with a minimum frequency threshold of 3 (Church and Hanks 1990). MI is computed with the frequency count of the whole multiword sequence divided by the product of expected frequency counts of each word.

When it comes to collocations, the phraseology-based tradition has also been used to further define or identify collocations from corpora (Cowie 1998). In this view, collocations often refer to two- or three-word strings, syntactically and semantically constrained at a moderate degree, such as *kick the habit* and *heavy rain* (Gyllstad 2007, Webb and Kagimoto 2009). Thus, collocations are distinct from free combinations, which are least fixed structurally but most literal semantically (e.g., *kick the ball*) and from idioms, which are most fixed structurally but least literal semantically (e.g., *kick the bucket*). In the view of the phraseology-based tradition, lexical bundles or ngrams such as *they said that* and *according to*, respectively, are more like free combinations, structurally free and semantically literal.

The influence of L1 is another important topic in L2 collocation research. Specifically, congruency (word-to-word match of L1-L2) has been investigated to grasp a cross-linguistic influence on L2 collocation learning: A collocation is congruent if its concept can be expressed in an L1 collocation based on word-to-word translation; if not, it is incongruent (Nesselhauf 2005). Collocation congruency is likely to determine the amount of learning burden. For most Korean EFL learners, the congruent collocation *do homework* is easier to learn and process than the incongruent collocation *cause damage.* This cross-linguistic aspect of collocations needs more investigation to reveal L2 collocational knowledge development as L2 proficiency increases.

2.2 L2 Collocation Processing and Its Theoretical Backgrounds

Theoretical explanations about L2 collocation processing have been proposed to understand what happens when L2 collocational knowledge is learned, acquired, and then used and to further elaborate the nature of L2 collocation use.

First, the explosive interest can be explained by holistic processing of L1 from Wray's (2002) dual processing model. The dual processing model proposed that two types of language processing, holistic and analytic, differently involve four phases of development from infancy to adulthood. L1 speakers tend to develop holistic and analytic knowledge from their birth to childhood and in their late teenage, they achieve the balance between holistic and analytic processing. However, L2 speakers, especially who are adults, are likely to resort to analytic language processing. Suppose the collocation *strong coffee* is produced in communication. English native speakers easily produce it holistically by retrieving it whole because it is stored as whole in their lexicon. In contrast, many non-native speakers with weak collocational knowledge may form it analytically, based on grammatical rules and lexical knowledge of single words. This analytic processing of the collocation often results in a non-nativelike erroneous collocation such as *powerful coffee.* In the view of the holistic hypothesis, the key to successful L2 learning is to expand L2 collocational knowledge stored as a unit long term memory in the mental lexicon of the speaker and to reduce the cognitive burden (Nesselhauf 2005, Schmitt 2004).

The second model to be used to explain L2 collocation processing is the usage-based approach (Ellis 2002). The usage-based models mainly investigate how acquiring frequent constructions is exemplar-based, supposing that a large amount of exposure to L2 targeted input can facilitate L2 acquisition. Regarding language acquisition as statistical learning, the usage-based approach sees frequency influences the efficiency of collocation retrieval (Ellis 2002, Ellis and Ogdens 2017). That is, when L2 learners encounter certain collocations in their input more frequently, and thus they would process them easier than less frequent ones.

Overall, Wray's holistic hypothesis and the usage-based approach generally describe how a large amount of a language is formulaic, arguing that L2 collocations can improve processing efficiency. Using such ready-made chunks reduces cognitive load and allows to buy time to pay attention to other processing difficulties, and thus improves fluency. Native-like accuracy can also be obtained by using L2 collocations, which, if stored correctly, can serve as *islands of safety* or *reliability* (Dechert, 1983). L2 learners may speak confidently when stored L2 collocations are accessible quickly. In this sense, the use of L2 collocations are theoretically described and motivated.

2.3 Previous Studies on L2 Collocation Processing

The interest in on-line processing of L2 collocations has been steadily growing to date, but it is still limited (Siyanova-Chanduria and Martinez 2015). The major question is whether multiword or formulaic sequences are stored in the mental lexicon and retrieved as a whole in language use as seen in Wray's (2002) holistic hypothesis.

To test the holistic hypothesis, processing advantages of multiword sequences (formulaic) over free combinations (non-formulaic) have been widely explored and proven in L1 and L2 research for lexical bundles or ngrams (e.g., Jiang and Nekrasova 2007, Tremblay, Derwing, Libber and Westbury 2011). The processing advantages of collocations over noncollocations (or free combinations) have also been examined by Wolter and his colleagues with conflicting findings: A processing advantage was reported (e.g., Wolter and Yamashita 2014) while a processing cost was found for collocations (e.g., Gyllstad and Wolter 2016). Gyllstad and Wolter (2016) further suggest that this cost tends to derive from the semantically semi-transparent nature of collocations as seen in the phraseology-based tradition, rather than the congruency of collocations.

In particular, processing effects of corpus-derived distributional information and congruency of collocations have been examined in the small but fast growing body of research. Two features of statistical properties underlying collocations, phrasal frequency and MI, were examined and significant effects were found. Gyllstad and Wolter (2016) reported a significant effect of phrasal frequency in advanced L1 Swedish learners' receptive processing collocations by using a semantic judgment task. Yi (2018) also found that 32 advanced L1 Chinese learners demonstrated sensitivity to frequency as well MI (or contingency) of English adjective-noun collocations by using a phrasal acceptability judgment task. Sonbul (2015) also found a processing advantage of phrasal frequency of adjective-noun collocations for nonnative speakers of English for the both data from a rating task and an eye-movement task (first pass reading time only).

The congruency effect to see if L1 knowledge is activated in L2 collocation processing has also been investigated, but conflicting findings were reported. Wolter and Gyllstad (2011, 2013) found that advanced L1 Swedish learners were highly sensitive to congruency of English collocations, recognizing congruent collocations faster than incongruent collocations. However, Yamashita and Jiang (2010) found mixed findings about the congruency effect for two proficiency groups of Japanese English learners: The low-proficiency group was more likely to recognize congruent collocations faster than incongruent collocations, whereas the advanced group showed no statistical difference in processing between congruent and incongruent collocations. It was concluded that the advanced group as ESL learners seemed to recognize L2 collocations independently from their L1.

To our best knowledge, two studies have been conducted for advanced Korean English learners' collocation processing, Kim and Kim (2012) and Han (2015). Kim and Kim (2012) used a self-paced reading task to examine receptive collocation processing by 14 Korean ESL learners. With phrasal verbs such as *turn out* and *point out* as collocations, they found the advanced learners read faster high-frequency collocations than low-frequency ones. Han (2015) conducted a comprehensive study on L2 collocation processing by 50 advanced Korean EFL learners. It sought to test Wray's (2002) holistic hypothesis and the effects of MI and congruency in L2 collocation productive processing. To this end, a read-aloud task was developed and employed with a total of 225 target stimuli, 180 collocations and 45 noncollocations, categorized into 15 sets in terms of frequency counts, MI scores, and congruency.[1]

[1] The present study adopted and slightly modified the materials and research designs developed by Han (2015). A summary of them is presented in Method in this paper.

The results were consistent with Gyllstad and Wolter (2016), in that a processing cost for collocations. No statistically significant effects of the two categorical predictors, MI and congruency, were found with the frequency counts controlled for. A preliminary explanation about the findings was that the advanced group like English native speakers seems to have a large amount of collocational knowledge, for which association strength or congruency of collocations would not affect their online processing.

Overall, the literature review reveals that most of the previous research was concerned with advanced L2 learners' receptive collocational processing using acceptability judgment tasks. Little research was done with other experimental tasks or types of stimuli, or differing intermediate or low proficiency groups. This study was motivated to see if previous research findings can be generalized with other types of tasks and proficiency levels. In this regard, this study aimed to examine how intermediate-level Korean EFL learners process collocations versus noncollocations and whether and how much frequency, MI, and congruency of collocations influence their productive processing. To this end, a read-aloud experiment was conducted (see Method). Considering the main goal of this study, the following two research questions were thus formulated:

1. Are intermediate-level Korean EFL learners likely to process collocations better than noncollocations?
2. Do frequency, mutual information, and congruency of collocations influence the Korean EFL learners' performance on a read-aloud task?

The first question is to examine whether Korean EFL learners process collocations better than noncollocations to obtain evidence for Wray's (2002) holistic hypothesis. The second question is to examine the effects of frequency, MI, and congruency on L2 productive collocation processing. Results and findings will be used to discuss theoretical, methodological, and educational implications for L2 collocation learning and use.

3. Method

3.1 Participants

A total of 25 Korean EFL learners attending a private university located in the northeastern part of Korea were invited, without difficulty in reading aloud and study-abroad experience in any English-speaking country. They received 10,000 won in return and four participants were excluded for analysis due to technical errors. Table 1 lists a summary of the 21 participants' responses to a demographic questionnaire from gender to self-assessment scores of four skills of English (i.e., reading, listening, speaking, and writing; 1 for very low up to 5 very high). Among them, 18 were females and 17 were English majors. The TOEIC scores reported by the participants ranged from 680 to 850 with the average of 823.05. Their average self-assessment reading and listening ability scores were at an intermediate level (about 3/5) whereas their speaking ability was low (2.1/5). Based on the TOEIC scores and the self-assessment scores, the participants were generally considered at an intermediate level of English language proficiency.

Table 1. Detailed Profiles and Self-reported Assessment of Participants (k = 21)

Characteristics		Self-assessment Scores (average)	
Gender	Male 3 Female 18	Reading	3.04/5
Major	English 17 Non-English 4	Listening	3.0/5
Age	Mean 24.05 Range 22-30	Speaking	2.1/5
TOEIC scores	Mean 823.05 Range 680-850	Writing	2.7/5

3.2 Materials

This study used the experiment materials and designs developed by Han (2015) as they were particularly designed for examining Korean EFL learners' processing of collocations and noncollocations. In the following, the target set of stimuli and read-aloud task are explained in detail (see Han 2015 for detailed information).

3.2.1 Target stimuli selection

The target stimuli were selected to obtain a balanced set of stimuli of the read-aloud task for collocability (i.e., collocation or noncollocation) and phrasal frequency, MI, and congruency of collocations. First, a total of 1224 verb-noun phrases based on previous collocation research were extracted (e.g., Nesselhauf 2005, Webb and Kagimoto 2009, Yamashita and Jian 2010) and Mark Davies' list of verb-noun phrases (available at https://www.ngrams.info/coca/v_the_n.txt). After two Korean English teachers' judgments of phrases on the degree of phrasal difficulty to Korean EFL learners, a total of 1037 verb-noun phrases remained. Second, the 1037 phrases were categorized in terms of their collocability (C for collocations; NC for noncollocations), frequency, MI, and congruency. The spoken subcorpus of the Corpus of Contemporary American English (COCA, available at https://corpus.byu.edu/coca; Davies 2008) was first consulted to obtain two pieces of frequency-based information for each of the 1037 phrases, raw frequency and MI (the verb as a collocate for the noun with a span size of 4 to the left-size and 0 to the right-size).[2] The frequency counts and MI scores were positively skewed, and thus log-transformed values were used instead. The 1037 phrases which were further classified into three frequency bands, low, mid, and high frequency (LF, MF and HF, respectively). Then, the MI scores of the phrases equal to or above 3 were considered collocations; otherwise, noncollocations (Church and Hanks 1990). With the COCA consultation, the 1037 phrases were split into 578 collocations and 459 noncollocations across the three frequency bands and for the collocations, either high MI (HM) or low MI (LM) groups. Congruency judgment was also conducted by three Korean native speakers with 6 years or more of English teaching experiences to classify the 578 collocations into congruent (CR) or incongruent (ICR) phrases (.85 of the inter-coder reliability or percentage of exact agreement). For instance, an English phrase *do the research* is congruent with the Korean counterpart 연구를 하다 (*the research do*). Another English phrase *kick a habit* is incongruent as the verb *kick* is not correctly translated based on the word-to-word translation. Through discussion, any disagreements were resolved. As a result, a final list of 225 target stimuli was obtained with 180 collocations and 45 noncollocations.

Table 2 presents the 15 sets of target stimuli by collocability, frequency, MI, and congruency with average frequency and MI scores along with the example stimuli. The 15 sets include 12 collocation sets and 3 noncollocation sets with 15 target stimuli each. The collocations within each frequency band were divided into HM (about 7 MI) and LM groups (about 4 of MI), and the collocations in each MI group was further divided into CR and ICR groups. The noncollocations were split into the three frequency bands only with relatively low average frequency from 0.70 to 2.23

[2] At the time of the materials development, the size of COCA was 480 million words and that of the spoken subcorpus was 90 million words.

compared to those of the collocations from 1.69 to 5.89. Their median MI scores were near zero, ranging from −0.40 to 0.36 and every noncollocation was congruent. To avoid any effect of priming, words did not appear more than two times in each set (Trofimovich and McDonough 2011).

Table 2. Sets of Target Stimuli by Collocability, Frequency, MI, and Congruency

Coll.	Freq.	MI	Cong.	Average frequency (log−transformed)	MI (Median)	Example Stimulus
C	HF	HM	CR	5.29	7.16	cancel the order
			ICR	5.89	6.65	wear a watch
		LM	CR	5.82	3.81	melt ice
			ICR	5.34	4.24	propose a toast
	MF	HM	CR	3.46	7.05	show weakness
			ICR	3.69	7.00	draw blood
		LM	CR	3.33	4.08	relieve stress
			ICR	3.63	4.15	kick a habit
	LF	HM	CR	1.98	6.89	hear music
			ICR	1.69	6.52	take notes
		LM	CR	1.83	4.13	sign a contract
			ICR	1.90	4.11	pay attention
NC	HF			2.23	0.36	hold a camera
	MF			1.53	−0.08	need water
	LF			0.70	−0.40	do the research

Note. Coll. = collocability: C for collocation, NC for noncollocation; Freq.: HF for high frequency, MF for medium frequency, LF for low frequency; MI: HM for high mutual information, LM for low mutual information; Cong. = congruency: CR for congruent, ICR for incongruent.

3.2.2 The read−aloud task

The read−aloud task was employed to examine Korean EFL learners' productive processing of English verb−noun collocations and noncollocations. To implement the task, a computer software program called DMDX version 5 (Forster and Forster 2014) was used as it has been widely used in psycholinguistic research to measure reaction or reading time (Wolter and Gyllstad 2013, Yamashita and Jiang 2010). DMDX automatically presents stimuli one at a time on screen and save the read−aloud times for each participant. A DMDX script was written and employed as all the target stimuli were presented on screen in the same order for the participants with 15 practice items and 5 high−frequency stimuli at the beginning of the task. The rest of the stimuli were presented by taking one item from each of the 15 sets.

Figure 1 displays how reading aloud a stimulus *wear a watch* was performed. Once the practice session was over, the experiment session began. Once the participant pressed the Enter key, each stimulus was presented with a fixation point (+) for about 500.1 milliseconds or half of a second and a stimulus in order. Reading aloud *wear a watch* consists of two phases, recognition and articulation. As soon as the participant recognized the stimulus on screen, he or she was asked to articulate it as naturally and accurately as possible and then to press the Enter key to move onto the next item. The time taken for both recognizing and articulating the stimulus was used as the read−aloud time in milliseconds.

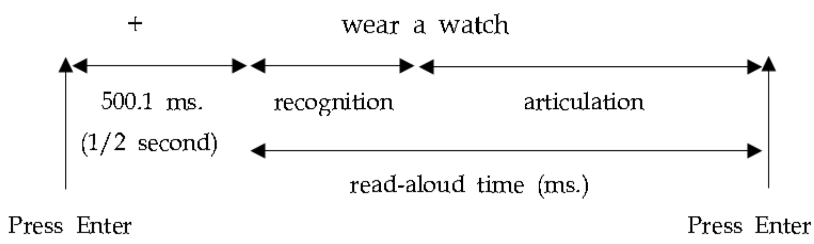

Figure 1. Procedure of Reading Aloud *wear a watch*

3.3 Procedure

The experiment was separately administered with each participant in a quite place for about 40−50 minutes. Once a consent form was signed, the purpose of the study was briefly explained along with the experiment procedure. Then, they performed the read−aloud task within 20 minutes on average. To reduce the influence of any fatigue, a one−minute break was given in the middle of the task. Then, they performed a translation task of the target stimuli (from English to Korean), of which scores were used to exclude the read−aloud times of any unknown phrases from analysis (see Appendix for sample questions of the translation task). The experiment ended with the demographic questionnaire, of which responses were summarized in Table 1.

3.4 Analyses

Descriptive statistics of the read−aloud times of the target stimuli were computed along with the assumption check, and linear mixed−effects models (LMMs) were used with crossed random effects for subjects and items using R (version 3.6.1; R core team 2019). LMMs were specifically adopted to build and identify the best-fit model to the read-aloud time dataset because of their superiority to other general linear models such as repeated-measures ANOVAs for analyzing grouped data (Baayen 2008, Baayen, Davison and Bates 2008) and their flexibility and robustness for unbalanced designs and missing data (Gelman and Hill 2007). A total of six distributional assumptions for LMMs from linearity to normality were also examined and all were met (see Crawley (2012) and Gelman and Hill (2007) for the assumptions of LMMs).

Variables included one dependent variable of the read-aloud times (continuous) and four independent variables, collocability (collocation vs. noncollocation) and congruency (congruent vs. incongruent) (categorical) and frequency (phrasal frequency) and MI (continuous).[3] Due to the non-normal distribution, the read-aloud times and frequency were log-transformed (natural log). To avoid collinearity issues, all the continuous predictors (i.e. frequency, MI) were centered at their means.

For each research question, LMMs were built, compared, and examined using the R package *lme4* (version 1.1-21; Bates, Maechler, Bolker and Walker 2015). At first, a maximal model was built with all the possible parameters of random slopes and intercepts. For instance, the maximal model formula for the first research question on the processing effect of collocability is as follows: (logRT ~ collocability + (1+collocability|subject) + (1+collocability|item), where the by-subject random intercept (denoted as 1) was included to account for the different speech rates across the participants. Then, a series of reduced or nested models were examined to find better-fit models with the function *anova()* based on a maximum likelihood technique. For model comparison, two criteria, Akaike's Information Criterion (AIC; Akaike 1974) and the Likelihood Ratio Test (LRT), were consulted.

4. Results

Before any statistical analysis, the data from the read-aloud task were cleaned by removing the read-aloud times of any unknown phrases based on the translation scores (M = 189.10 out of 225, SD = 8.38) and outliers exceeding about 2.5

[3] Two variables, the sum of letter counts in phrase and the sum of syllables in phrase, were additionally examined in model building as they may influence the read-aloud times of phrases, but no influence was found.

absolute standard deviations from the mean (Baayen 2008). This resulted in 16.27% of data loss. Four participants were eliminated due to errors during the experiment.

4.1 Descriptive Statistics of Read-aloud Times

The descriptive statistics of the read-aloud times (RTs) in milliseconds were computed for the 15 target stimuli sets, as shown in Table 3. As explained in Method, the sets were categorized by collocability (collocation vs. noncollocation), frequency (HF, MF, LF), MI (HM, LM), and congruency (congruent vs. incongruent). The average of the RTs for the collocations was 2102.47 milliseconds (*SD* = 648.80) and for the noncollocations, the average was 1854.51 milliseconds (*SD* = 500.74). In general, the participants read-loud faster highly frequent collocations or noncollocations than less frequent counterparts, and noncollocations than collocations.

Table 3. Descriptive Statistics of Read-aloud Times (RTs in milliseconds) by Stimuli Sets (*k* = 21)

Coll.	Freq.	MI	Cong.	M	SD
C	HF	HM	CR	1788.74	523.30
			ICR	2377.41	773.63
		LM	CR	1903.19	632.64
			ICR	2147.14	626.45
	MF	HM	CR	2058.99	662.19
			ICR	2219.29	565.04
		LM	CR	2074.08	620.28
			ICR	2214.45	621.65
	LF	HM	CR	2056.18	635.54
			ICR	2461.30	711.60
		LM	CR	2025.75	541.46
			ICR	2149.24	604.35
NC	HF			1804.68	505.01
	MF			1834.49	486.15
	LF			1938.91	504.45

Note. M = mean, SD = standard deviation, Coll. = collocability: C for collocation, NC for noncollocation; Freq.: HF for high frequency, MF for medium frequency, LF for low frequency; MI: HM for high mutual information, LM for low mutual information; Cong. = congruency: CR for congruent, ICR for incongruent.

4.2 Linear Mixed-effects Model of Read-aloud Times for Collocability

Table 4 lists the results of the best-fit linear-mixed-effects model for the logRTs by collocability. As the Intercept-only model, the best-fit model's formula was (logRT ~ collocability + (1|subject) + (1|item). As seen in the estimates of fixed effects, the estimate of Intercept was 7.62, showing that the average read-aloud time of the whole target stimuli for the 21 participants was 2038.56 milliseconds [exp(7.62)]. One main effect of collocability was found (Estimate = −0.13, *SE* = 0.03, *t* = −4.97), suggesting that the average read-aloud time of the collocations was longer than that of the noncollocations by 13% [1 − (exp(−0.13))], that is, about 248.51 milliseconds [exp(7.62) − exp(7.62 − 0.13)]. These two fixed-effects parameters had absolute *t* values above 2 and no zeros in their 95% bootstrapped confidence intervals (7.53 ~ 7.69 for Intercept and −0.18 ~ −0.07 for collocability), which indicate that both had statistically significant effects on the read-aloud performance.

Table 4. Linear Mixed-effects Model Results for Read-aloud Times (logRTs) by Collocability

Parameters	Fixed Effects			Random Effects			
	Estimate	*SE*	*t*	By subject		By item	
				Variance	*SD*	Variance	*SD*
(Intercept)	7.62	0.04	192.20	0.03	0.17	0.02	0.15
Collocability	−0.13	0.03	−4.97				

Note. There are 3961 observations by the participants for reading aloud the 180 collocations and 45 noncollocations. Model formula: logRT ~ collocability + (1| subject) + (1|item). The collocations were the reference for collocability.

4.3 Linear Mixed-effects Model of Read-aloud Times for Collocations by Frequency, MI, and Congruency

Table 5 presents the results of the best-fit linear-mixed-effects model for the logRTs of the collocations in logged milliseconds by frequency, MI, and congruency after model comparisons. As the Intercept-only model, the best-fit model's formula was logRT ~ frequency + congruency + (1|subject) + (1|item).

The estimates of fixed effects show that the estimate of Intercept was 7.55 (logged in millisecond), showing the average read-aloud time of the collocations for the 21 participants was 1900.74 milliseconds [exp(7.55)]. Two main effects of frequency and congruency were found. As for the effect of frequency (Estimate = −0.02, *SE* = 0.07, *t* = −2.37), one unit of increase in frequency (logged and centered) of the collocation led to a decrease in the read-aloud time by 2% [(1 − exp(−0.02)]. As for the effect of congruency (Estimate = 0.13, *SE* = 0.02, *t* = 5.84), the average read-aloud time of the congruent collocations was shorter than that of the incongruent collocations about 263.88 milliseconds [exp(7.55 + 0.13) − exp(7.55)]. All the three fixed-effects parameters had absolute *t* values above 2 and no zeros in their 95% bootstrapped confidence intervals (7.47 ~ 7.64 for Intercept, −0.03 ~ −0.01 for collocability, and 0.09 ~ 0.17 for congruency), which indicates that all had statistically significant effects on the read-aloud performance.

Table 5. Linear Mixed-effects Model Results for Read-aloud Times (logRTs) by Frequency, MI, and Congruency for Collocations

Parameters	Fixed Effects			Random Effects			
	Estimate	*SE*	*t*	By subject		By item	
				Variance	*SD*	Variance	*SD*
(Intercept)	7.55	0.04	183.67	0.03	0.18	0.02	0.14
Frequency	−0.02	0.07	−2.38				
Congruency	0.13	0.02	5.84				

Note. There are 3045 observations by the participants for reading aloud the 180 collocations. Model formula: logRT ~ frequency + congruency + (1|subject) + (1|item). The congruent collocations were the reference for congruency.

5. Discussion and Conclusion

5.1 Processing Collocations and Noncollocations

The first research question was addressed by the results for processing of the collocations versus noncollocations. The intermediate-level Korean EFL learners did not elicit significantly faster read-aloud times for the collocations over the noncollocations. Importantly, this finding does not support Wray's (2002) holistic hypothesis that multiword or formulaic sequences are stored and retrieved as a whole chunk from the mental lexicon. To date, this hypothesis has been most empirically tested and proven across L1 and L2 studies for lexical bundles or ngrams (Jiang and Nekrasova 2007, Tremblay et al. 2011) and for collocations (Wolter and Yamashita

2014), and substantial evidence has been drawn for the importance of L2 knowledge of multiword sequences in language use. However, this study with the intermediate-level Korean EFL learners found that the productive processing of the collocations was slower than that of the noncollocations about 248.51 milliseconds on average. This is consistent with the finding by Han (2015) with advanced Korean EFL learners, indicating that collocations still pose a processing burden for both the intermediate and advanced proficiency groups.

There are two possible explanations for the processing disadvantage of collocations. First, L2 productive processing of collocations and noncollocations may be distinct from L2 receptive processing. Wolter and Yamashita (2014) examined the receptive processing, whereas this study focused on the productive processing. Second, scrutiny of the target stimuli showed that almost all the noncollocations were semantically transparent or literal, including *join the war*, *give a fact*, or *buy a company*. As seen in the phraseology-based tradition (Cowie 1998), those semantically literal noncollocations are free combinations. In contrast, a number of the collocations were semantically less transparent, such as *fill the gap*, *catch one's eye*, or *shoot the film*. A few collocations were like idioms (e.g., *save face*, *pave the way*) in the view of the phraseology-based tradition. Note that the target stimuli were classified in terms of mutual information in the frequency-based tradition: above 4 ~ 7 for collocations and near zeros for noncollocations. In this regard, the processing cost for collocations in this study may stem from the semantically less transparent nature of collocations as seen in the phraseology-based tradition (Wolter and Gyllstad 2016).

5.2 Effects of Frequency, Mutual Information, and Congruency on Processing Collocations

The second research question was addressed by the results for frequency, mutual information, and congruency of collocations. Among them, frequency and congruency yielded statistically significant effects: The intermediate-level Korean EFL learners elicited significantly faster read-aloud times for more frequent collocations and for congruent collocations, respectively. Mutual information (or the association strength between two words) had no significant influence in the productive collocation processing.

The significant productive processing effect of frequency from this study is consistent with the findings reported in previous studies of receptive processing with advanced learners of English and judgment tasks, including Gyllstad and Wolter (2016) and Yi (2018). The facilitative effect of frequency supports the usage-based model in that frequent input can increase the amount of exposure to the input or are more salient for learners, thereby promoting language acquisition as statistical learning (Ellis 2002).

As for the congruency effect, similar results were reported in Wolter and Gyllstad (2011, 2013) and Yamashita and Jiang (2010). These previous studies examined the receptive processing, but differ in proficiency groups, advanced learner groups for Wolter and Gyllstad (2011, 2013) and low proficient learners for Yamashita and Jiang (2010). In contrast, Han's (2015) research did not find any statistically significant effect of congruency on the advanced Korean EFL learners' productive processing. Despite the use of the same experimental materials and procedure, the discrepancy in the finding between the current study and Han's (2015) study may derive from different proficiency groups (intermediate and advanced, respectively).

5.3 Conclusion and Future Research

A growing body of research on processing L2 collocations has been witnessed. In line with this accumulated interest, this study conducted a read-aloud experiment with intermediate-level Korean EFL learners to see whether a processing advantage for collocations over noncollocations exists and to assess the effects of frequency, mutual information, and congruency of collocations in the (controlled) oral production. It was found that the read-aloud times of collocations were longer than those of noncollocations, which provides counterevidence for Wray's (2002) holistic hypothesis. Another finding from this study is that both phrasal frequency and congruency of collocations significantly influenced the productive processing.

These results and findings can be discussed to offer theoretical, methodological, and educational implications for the field of L2 collocation processing research. Theoretically, the phraseological continuum model (or phraseology-based tradition) needs more attention for explaining what are collocations or why those collocations are processed faster than others (Cowie 1998, Gyllstad and Wolter 2016). These questions have been successfully addressed by the usage-based model (along with the frequency-based tradition) for lexical bundles or ngrams, but not for collocations. The degree of semantic transparency also seems to play a role in collocation processing efficiency, and thus, the phraseological continuum model as well as the usage-based model can explain how different collocations are processed.

Methodologically, target stimuli and experimental tasks used in previous research need to be re-examined in their reliability and validity. As also shown in this study, the target stimuli needed further classification in terms of their semantic transparency. The read-aloud task also needs investigation to see if it really reflects the oral production of collocations. Therefore, replication studies are strongly recommended, with different proficiency groups or modified tasks, to examine if a research finding is robust across studies. Targeting different types of multiword sequences such as idioms or lexical bundles or different collocational features can also lead to a fuller understanding of the processing of multiword sequences.

Educationally, the finding of the processing burden for collocations over noncollocations can be used as strong motivation for teaching collocations rather than noncollocations in L2 learning. It was found that the advanced and the intermediate-level Korean EFL learners processed the collocations slower than the noncollocations. To facilitate L2 learners' learning and acquisition of collocations, careful selection of target collocations for instruction is necessary based on frequency and congruency of collocations along with language proficiency: In general, more frequent and congruent collocations can be taught before less frequent and incongruent collocations.

More and more psycholinguistic research has been conducted to examine the processing of multiword sequences beyond single words. It is hoped that the findings and implications above can lead to more studies in this growing area of research.

References

Akaike, H. 1974. A new look at the statistical model identification. In *Selected Papers of Hirotugu Akaike*, 215-222. Springer, New York, NY.

Baayen, R. H. 2008. *Analyzing Linguistic Data: A Practical Introduction to Statistics Using R*. Cambridge, UK: Cambridge University Press.

Baayen, R. H., D. J. Davidson and D. M. Bates. 2008. Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language* 59(4), 390-412.

Bates, B., M. Maechler, B. Bolker and S. Walker. 2015. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software* 67(1), 1-48. Available at http://CRAN.R-project.org/package=lme4

Church, K. and P. Hanks. 1990. Word association norms, mutual information, and lexicography. *Computational Linguistics* 16(1), 22-29.

Cowie, A. P. 1998. Introduction. In A. P. Cowie, ed., *Phraseology: Theory, Analysis and Applications,* 1-22. Oxford: Clarendon Press.

Davies, M. 2008. *The Corpus of Contemporary American English (COCA): 560 Million Words, 1990-present.* Available at https://corpus.byu.edu/coca/.

Dechert, H. W. 1983. How a story is done in a second language. In C. Faerch and G. Kasper, eds., *Strategies in Interlanguage Communication,* 175-195. London, UK: Longman.

Ellis, N. C. 2002. Frequency effects in language processing and acquisition: A review with implications for theories of implicit and explicit language acquisition. *Studies in Second Language Acquisition* 24(2), 143-188.

Ellis, N. C. and D. C. Ogden. 2017. Thinking about multiword constructions: Usage-based approaches to acquisition and processing. *Topics in Cognitive Science* 9(3), 604-620.

Erman, B. and B. Warren. 2001. The idiom principle and the open choice principle. *Text* 20(1), 87-120.

Firth, J. R. 1957. *Idioms and idiomaticity.* In *Papers in Linguistics, 1934-1951.* London, UK: Oxford University Press.

Forster, K. I. and J. C. Forster. 2014. *DMDX Help: DMDX 5.* Available at http://psy1.psych.arizona.edu/~jforster/dmdx

Gelman, A. and J. Hill. 2007. *Data Analysis Using Regression and Multilevel/Hierarchical Models.* Cambridge, UK: Cambridge University Press.

Gyllstad, H. 2007. *Testing English Collocations: Developing Receptive Tests for Use with Advanced Swedish Learners.* Doctoral dissertation, Lund University, Sweden.

Gyllstad, H. and B. Wolter. 2016. Collocational processing in light of the Phraseological Continuum Model: Does semantic transparency matter? *Language Learning* 66(2), 296-323.

Han, S. 2015. *Processing of Formulaic Sequences by Native and Nonnative Speakers of English: Evidence from Reading Aloud.* Doctoral dissertation. Northern Arizona University, AZ, USA.

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

Kim, S.-H. and J.-H. Kim. 2012. Frequency effects in L2 multiword unit processing:

Evidence from self-paced reading. *TESOL Quarterly* 46(4), 831-841.

McEnery, T., R. Xiao and Y. Tono. 2006. *Corpus-based Language Studies: An Advanced Resource Book*. New York, NY: Routledge.

Nattinger, J. R. and J. S. DeCarrico. 1992. *Leixcal Phrases and Language Teaching.* Oxford: Oxford University Press.

Nesselhauf, N. 2005. *Collocations in a Learner Corpus*. Philadelphia, PA: John Benjamins Publishing Company.

Pawley, A. and F. H. Syder. 1983. Two puzzles for linguistic theory: Nativelike selection and nativelike fluency. In J. C. Richards and R. W. Schmidt, eds., *Language and Communication,* 191-226. London, UK: Longman.

R Development Core Team. 2019. *R: A Language and Environment for Statistical Computing (version 3.6.1)*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org/

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

Siyanova-Chanturia, A. and R. Martinez. 2015. The idiom principle revisited. *Applied Linguistics* 36(5), 549-569.

Sonbul, S. 2015. Fatal mistake, awful mistake, or extreme mistake? Frequency effects on off-line/on-line collocational processing. *Bilingualism: Language and Cognition* 18(3), 419-437.

Tremblay, A., B. Derwing, G. Libben and C. Westbury. 2011. Processing advantages of lexical bundles: Evidence from self-paced reading and sentence recall tasks. *Language Learning* 61(2), 569-613.

Trofimovich, P. and K. McDonough. 2011. Using priming methods to study L1 learning and teaching. In P. Trofimovich and K. McDonough, eds., *Applying Priming Methods to L2 Learning, Teaching and Research,* 3-20. Amsterdam: John Benjamins.

Webb, S. and E. Kagimoto. 2009. The effects of vocabulary learning on collocation and meaning. *TESOL Quarterly* 43(1), 55-77.

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

Wolter, B. and H. Gyllstad. 2013. Frequency of input and L2 collocational processing: A comparison of congruent and incongruent collocations. *Studies in Second Language Acquisition* 35(3), 451-482.

Wolter, B. and J. Yamashita. 2014. Processing collocations in a second language: A case of first language activation? *Applied Psycholinguistics* 36(5), 1193−1221.

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

Yi, W. 2018. Statistical sensitivity, cognitive aptitudes, and processing of collocations. *Studies in Second Language Acquisition* 40(4), 831−856.

Examples in: English
Applicable Languages: English
Applicable Level: Tertiary

Han, Sumi
Professor, Hallym University
Department of English Language and Literature/
Digital Arts and Humanities
1 Hallymdaehack-gil, Chuncheon
Kwangwon-do, Korea
Tel: (033) 248-1532
E-mail: sumihan@hallym.ac.kr

Received: August 11, 2019
Revised: September 05, 2019
Accepted: September 09, 2019

Appendix

Samples of Translation Task

Directions: Please look at each of the phrases you saw in the read−aloud task. Say the Korean meaning of each phrase as quickly and accurately as possible.

1	2	3
heal the wounds	take steps	find a website
cause damage	give a reminder	need water
send a message	buy a company	want justice
cross the line	have lunch	require cooperation
sing a song	polish shoes	commit errors
hold a camera	do the research	have a question
give a reminder	have lunch	want the book
read a message	lose the war	make goals
leave a decision	do business	take pleasure