Explaining the Effect of Congruency on Processing Collocations: A Modified Replication of Wolter and Yamashita (2015)

Hyeonah Kang
Second Language Acquisition and Teaching, University of Arizona, USA.

Abstract

Using a lexical decision task, Wolter and Yamashita (2015) showed that collocations that exist only in L1 but not in L2 were not processed faster than collocations that only exist in L2 but not in L1 or a random combination of two words. This result seems to support the age/order of acquisition effects (Carroll & White, 1973) over Jiang's (2000) model of lexical acquisition for the processing advantage of congruent collocations. To replicate and extend Wolter and Yamashita (2015), the present study utilized both on-line (lexical decision task) and off-line (meaning recall test) measures. In line with Wolter and Yamashita (2015), results from the lexical decision task showed that there was no L1 activation effect by L2 (L1 Korean) learners, regardless of their L2 proficiency levels. The meaning recall test result, on the other hand, showed that L2 learners showed more tendency to rely on their L1, as suggested by Jiang's (2000) model. The discrepancy between on-line and off-line performances can imply that the findings of Wolter and Yamashita (2015) should be reexamined using different methods and techniques.

Keywords: Collocations, congruency, processing, age/order of acquisition effect (Carroll & White, 1973), Jiang's (2000) model of lexical acquisition

Introduction

Collocations, a part of multiword units, are often defined as "frequently recurring two- or three-word syntagmatic units which can include both lexical and grammatical words" (Henriksen, 2013). Examples include a combination of an adjective and a noun, such as "heavy traffic", or a verb plus a noun, such as "set objectives".

The important role of collocations in L2 learning has been well-documented. According to Hill (2000), collocations account for almost two thirds of English discourse, and the number of collocations exceeds the number of individual lexical items. In addition to a considerable number of collocations in a language, collocational knowledge has largely been reported to provide processing advantages (Conklin & Schmitt 2008; Wray, 2002; Wray & Perkins, 2000). Specifically, a collocation is stored as a chunk, treated as a single entity not as two individual lexical units (e.g., a word combination such as "deep sleep" is likely to be stored as a whole unit, rather than stored as "deep" and "sleep" respectively), thus a collocational competence can diminish a learner's cognitive load and processing time (Konopcka & Bock, 2009; Nattinger & DeCarrico, 1992; Sprenger, Levelt, & Kempen, 2006).

Collocations have enjoyed ample attention in L2 acquisition literature (Boers & Webb, 2018; Choi, 2017; Durrant & Schmitt, 2009; Sonbul & Siyanova-chanturia, 2021; Tsai, 2020; Webb & Kagimoto, 2011; Yamashita & Jiang, 2010). One major research strand in L2 collocation research is the influence of congruency on the processing and production of

collocations (e.g., Nesselhauf, 2003; Peters, 2016; Yamashita & Jiang, 2010). Specifically, a collocation can be divided into congruent or incongruent collocation, based on whether a word-for-word L1 translation equivalent exists or not. An example of a congruent collocation for L1-Korean speakers is "deep sleep" since this expression has a literal translation equivalent in Korean as "깊은 잠", which sounds natural and nearly identical in meaning to the English expression "deep sleep". On the other hand, for an English collocation "strong coffee", a word-for-word Korean translation is "강한 커피" (powerful coffee), resulting in an expression that would be perceived as unnatural in the Korean language.

Empirical evidence suggests that congruent collocations are processed faster than incongruent ones (e.g., Wolter & Gyllstad, 2011; Yamashita & Jiang, 2010). The question that arises here is why this is the case. In an attempt to explain the processing advantage of congruent collocations, this manuscript reports the findings of a replication study of Wolter and Yamashita (2015) conducted as a part of a coursework.

Background

Many researchers have begun to wonder whether congruency matters in processing L2 collocations–that is, would congruent collocations be processed significantly faster than incongruent ones? To answer this question, they have used a lexical decision task to measure the reaction time to the congruent collocations versus incongruent ones. Previous studies have overall converged in support of the advantage of processing congruent collocations over incongruent ones. To illustrate, Wolter and Gyllstad (2011) investigated the L1 transfer effect on the processing of L2 collocation by 35 L1-English speakers and 30 advanced L2 English learners. Verb plus noun collocations were presented in one of the three conditions: (a) congruent, (b) incongruent, and (c) control (a random combination of two words that neither exist in L1 nor L2). Participants were asked to respond to the lexical decision task assessing whether the target noun, preceded by the prime word (i.e., verb), is a real English word or not. Results showed that for L1 speakers, both congruent and incongruent collocations were almost equally processed faster than the control counterparts, suggesting that congruency had little effect on the processing of collocations for L1 speakers. In contrast, for L2 learners, only congruent collocations were processed faster than control pairs, while this pattern was not found for incongruent collocations. Similar results were reported in Yamashita and Jiang (2010) who further divided L2 learners based on their English proficiency levels (beginner, advanced). Results showed that unlike advanced L2 English learners showing comparable performances in processing congruent and incongruent collocations, L2 beginner English learners took significantly longer time to process incongruent collocations than congruent ones.

Then why are congruent collocations processed more quickly than incongruent ones? Previous studies to date have suggested two theoretical explanations. The first theory concerns the age of acquisition (AoA) effect (Carroll & White, 1973) suggesting that congruent collocations are usually acquired earlier due to pre-existing L1 corresponding patterns, thus facilitating the processing of congruent collocations over incongruent ones (Wolter & Yamashita, 2015). Another explanation is based on Jiang's (2000) model of adult L2 vocabulary acquisition. According to this model, the processing of L2 collocation will automatically lead to its lemma-based information, composed of L1 semantics and L1 syntax. This L1 information, in turn, primes its L1 collocates, which leads to more rapid recognition of the congruent collocation

in the L2. Based on this assumption, it would be apparent that congruency plays a role in processing L2 collocations.

Wolter and Yamashita (2015) took further steps to explore the relative plausibility between AoA effects (Carroll & White, 1973) and Jiang's (2000) model in taking account of the advantage of processing collocations over incongruent ones. By recruiting 50 L1 Japanese speakers, they hypothesized that if Jiang's (2000) model had been plausible, J-only items (collocations that only exist in Japanese but not in English, such as "far eye") would have been processed faster than E-only (collocations that only exist in English but not in Japanese, in other words, incongruent collocations, such as "busy road") or baseline items (i.e., a random combination of two words, such as "low parents", which does not exist in either Japanese or English). Contrary to their hypothesis, however, results showed that there was no significant difference in reaction time between J-only and E-only items, thus lending support to AoA effects (Carroll & White, 1973) over Jiang's (2000) model.

While the results of Wolter and Yamashita (2015) are promising, research exploring theoretical accounts to explain the processing benefits of congruent collocations over incongruent ones is still in its infancy and is characterized by a few methodological limitations. First, consider the paucity of experimental work investigating this issue: Wolter and Yamashita's (2015) study is the first and the only one that looked at comparing the relative effectiveness of the two accounts, AoA effect (Carroll & White, 1973) and Jiang's (2000) model. Notwithstanding paucity of evidence on the topic, explicit knowledge of target collocations was not examined in Wolter and Yamashita (2015). Although no significant difference was discernible in processing J-only items and E-only items, it is plausible that L1 Japanese participants would have provided more accurate translation equivalents to the J-only items relative to either E-only items or baseline items. If this is true, then this could be supporting evidence to Jiang's (2000) model. Lastly, it would be important to replicate these findings with different populations of participants, such as L1 Korean speakers, further divided into their L2 proficiency levels.

The current study aims to extend the findings of Wolter and Yamashita's (2015) study, as far as the acquisition of explicit knowledge is concerned with Korean speakers of English. Research questions that guide the present study are the following:

1. Are there any differences in processing (i.e., reaction time) based on different types of collocations (K-only, E-only, and baseline) and whether this effect is mediated by learners' proficiency levels (intermediate vs. advanced)?
2. Are there any differences in explicit knowledge based on different types of collocations (K-only, E-only, and baseline) and whether this effect is mediated by learners' proficiency levels (intermediate vs. advanced)?

Concerning the first research question, I assumed that L1 Korean speakers of English would process K-only word combinations (expressions that only exist in Korean but not in English) faster than the other types of word combinations, E-only (expressions that only exist in English but not in Korean) and baseline (a random or anomalous combination of two words that do not exist in either Korean or English), and this result would be mainly driven by L2-intermediate learners rather than by L2-advanced learners. This hypothesis was based on Choi (2016) who demonstrated that L2 English learners (L1 Korean) showed more tendency to use their L1 (Korean) over L2 (English) in the task of translating the meaning of new English vocabulary, thus supporting Jiang's (2000) L1 mediation effect. Compared to L2-intermediate learners, L2-advanced learners would show different results: K-only items would not be

processed faster than either E-only or baseline items. This result might derive from their near-native English proficiency or frequent repeated exposures to the English input. Lastly, L1-English speakers would process E-only items significantly faster than the other two items, K-only word or baseline items, given that E-only items are the only word combinations that exist in English.

With regard to the second question, hypotheses are made separately to E-only versus K-only items. To begin with *E-only* items, L1-English speakers would show the most accurate responses, followed by advanced and intermediate L2 learners, respectively. On the contrary, in terms of *K-only* items, L2 learners would provide definitions that conform to the expressions that exist in Korean, and this result will be more pronounced for L2-intermediate learners than L2-advanced learners. L1-English speakers, however, are likely to produce definitions that sound awkward or unnatural in Korean language.

Method

Participants and Research Design

Fifteen participants were recruited, including one group of L1 English speakers (n = 5) and two groups of L2 English learners (n = 5, each). L1 English speakers consisted of graduate students at the University of Arizona. L2 English learners, on the other hand, were recruited from a university in Korea whose native language was Korean. They were further divided into two groups, based on their English proficiency levels (intermediate or advanced), determined by the Common European Framework of Reference (CEFR). Specifically, L2 English learners who have attained over 95 TOEFL scores (C2: the highest CEFR level) were deemed as the advanced learners, whereas those who have attained under 72 TOEFL scores (B2: intermediate CEFR level) were regarded as intermediate learners (Papageorgiou et al., 2015). Thus, following this standard, advanced learners consisted of those who have reported over 100 TOEFL scores, whereas intermediate learners were composed of those who have reported under 70 TOEFL scores. Participants were informed that they could withdraw from the experiment at any time and completed a survey for demographic information.

The present study adopted a mixed-factorial design in which independent variables consisted of Group (L1-English, L2-English advanced, L2-English intermediate) as a between-subjects variable, and Item type (K-only, E-only, baseline) as a within-subjects variable. For dependent variables, mean reaction time to each item type and the number of correct responses in the meaning recall test were included.

Procedure

Through an online platform study, Findingfive, each participant completed the experiment (FindingFive Team, 2019). First, a survey was conducted to obtain information about the participants' background. Then the experiment began, consisting of the lexical decision task and a meaning recall test (see Figure 1 below).

Specifically, the lexical decision task began with a total of 50 English expressions (45 target items plus 5 practice items) appearing one at a time on the computer screen for 3,000ms. During this task, participants were asked to press either “YES” or “NO” button, according to their judgments on the acceptability of each appearing English expression (see Figure 2 below). Five practice items were given prior to the 45 test items. Each participant's reaction times, and error rates were recorded. Upon the completion of the lexical decision task, the participants were

given a thirty-second short break and then they were asked to respond to the meaning recall test either by translating or paraphrasing the meaning of each occurring English expression on the screen. The entire experiment lasted approximately 15-20 minutes.

Figure 1

Schematic Description of the Experimental Procedure

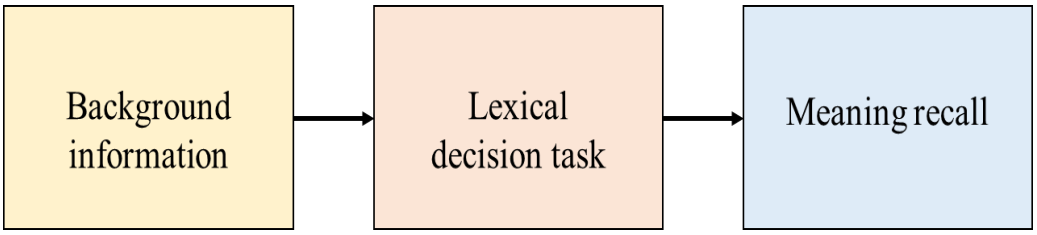
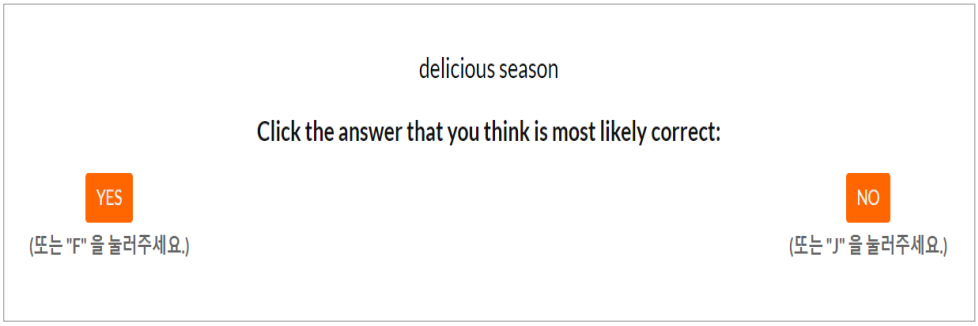


Figure 2

Sample Image of the LDT (Lexical Decision Task)



Materials

Target items

As can be seen in the Table 1 below, the target items were 45 total, which were divided into three groups: (a) K-only (n = 15), (b) E-only (n = 15), and (c) Baseline (n = 15). K-only items referred to the expressions that only exist in Korean but not in English (e.g., a Korean collocation "맛있는 계절" is translated as "delicious season", which sounds unnatural in English). In contrast, E-only items were defined as the collocations that only exist in English but not in Korean (e.g., an English collocation "strong coffee" is translated as "강한 커피" (powerful coffee), which sounds awkward in Korean). Alternatively, the English expression "thick coffee" sounds more natural when it is translated in Korean "진한 커피" to convey the original meaning of "strong coffee". Lastly, the baseline was composed of two-word combinations that were random or anomalous in either Korean or English language (e.g., rich yard). The purpose of the baseline was that it will be used for baseline reaction times to gauge the relative reaction times to either the K-only or E-only items. Note that the present study did not address congruent collocations (i.e., L2 collocations that have the translation equivalents in learner's L1, such as "deep sleep" (깊은 잠) for L1 Korean speakers of English); rather, the primary purpose of this study was to investigate whether K-only items would be processed faster than the other two

types of the two-word combinations, E-only items and baseline, as an evidence of L1 activation effect.

The procedure of selecting each target item consisted of several steps. First, in terms of K-only items, six J-only collocations were chosen from Wolter and Yamashita (2015), based on their presence in the Korean language. Then, the remaining nine legitimate combinations in only Korean were chosen and then these were translated into English. These K-only items were verified by two Korean secondary school teachers of English and Korean that those items only exist in Korean but not in English.

Next, concerning E-only items, four collocations were chosen from Bang and Choi (2018), and the remaining nine collocations were chosen from the researcher. According to the Corpus of Contemporary American English (COCA), the entire E-only collocations were above 3 MI (Mutual Information) scores[1], which is often taken to indicate a significant collocation threshold in previous studies (Ackermann & Chen, 2013; Hunston, 2002). Lastly, the baseline items were composed of 15 two-word combinations that do not exist in either Korean or English and these were verified by the two same evaluators, along with an English native speaker.

Table 1

Target Items

K-only	E-only	Baseline
delicious season	strong coffee	rich yard
light pocket	hard evidence	narrow rainbow
narrow thought	immediate family	heavy library
shallow injury	false teeth	small night
warm gift	busy road	soft knee
weak rain	hard evidence	slim soul
straight heart	full speed	bright mystery
dark society	narrow escape	happy pen
flowery road	heavy rain	buy phrase
rosy life	small talk	drop colors
blue dream	soft drink	short coffee
red lies	slim chance	major sleep
dizzy times	bright idea	low child
buy anger	buy insurance	total nose
pour love	drop hints	isolated teeth

[1] MI (Mutual Information) score, ranging from zero to 17, is a standard measure used to show the strength of co-occurrence of the words. For example, the two-word combination "extenuating circumstances" will have a very high MI score, because whenever we read or hear the word "extenuating", it is most likely to co-occur with the word "circumstances" (Siyanova & Schmitt, 2008).

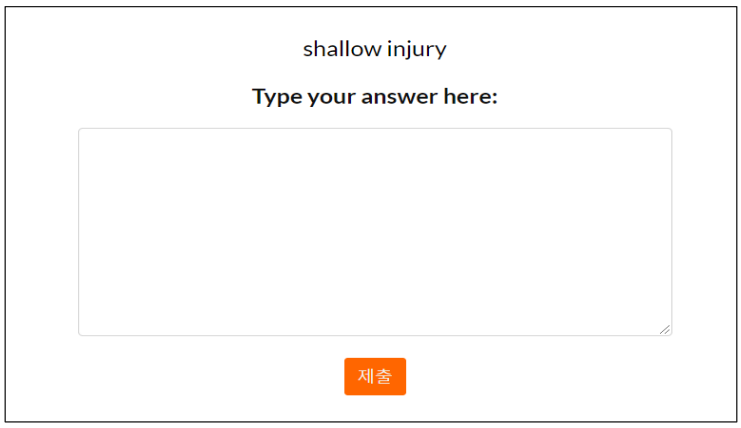
Meaning recall test

Meaning recall test was used to measure explicit knowledge of given English expressions (see Figure 3 below). Specifically, when an expression appeared on the computer screen one at a time, participants were asked to provide a translation equivalent or a short definition using either their L1 or L2. If they felt uncertain about how to define the meaning of the given item, they were instructed to type "I don't know". This test was also provided to native English speakers to make each group's experimental condition to be identical. All participants were given 20 seconds to respond to each item.

Unlike the E-only items, however, note that a word-for-word English translation of each K-only item is not semantically acceptable in English. Considering this therefore, the number of incorrect or "I don't know" responses should have been the highest for K-only collocations, regardless of participants' L1. However, despite the nonsensical form in the English, if L2 learners showed any attempts to provide the meaning of K-only collocations and this tendency was more salient than native English speakers, then I suspect that this will be evidence of the L1 mediation effect (e.g., Choi, 2017; Jiang, 2000).

Figure 3

A Sample Image of Meaning Recall Test



Scoring

Meaning recall test was scored using a lenient scoring method for each item type, E-only, K-only, and baseline. In terms of *E-only* items, responses with either exact or roughly equivalent meanings were awarded one point (e.g., for "strong coffee", "진한 커피" (= thick coffee) or "카페인이 많은 커피" (= coffee with lots of caffeine) were scored as correct); otherwise, zero points were given. Regarding *K-only* items, any attempts to define the meaning of each item were given one point; otherwise, zero points were given. In terms of *K-only* items, however, further scoring was conducted, considering the presence of the literal translations of K-only items in Korean. Therefore, the data of K-only items were scored by giving one point for a response containing the correct meaning in Korean and zero for the remaining responses. In contrast to E-only or K-only items, a different scoring criterion was used for *baseline* items, considering its non-existence in either Korean or English; therefore, responses that showed learners' uncertainty (e.g., "I don't know") were given one point, otherwise zero points.

Results

Analysis of reaction time results

Based on the data pre-screening process conducted in previous research (Wolter & Yamashita, 2015; Yamashita & Jiang, 2010), all erroneous responses (i.e., "No" responses to E-only items and "Yes" responses to either K-only or baseline items) were eliminated. Next, correct responses that took less than 350 ms were also excluded. Table 2 shows the trimmed mean reaction time (ms) for each item type (K-only, E-only, and baseline) across three groups (L1-English, L2-English advanced, and L2-English intermediate).

The primary purpose of this study was to investigate whether there is a difference in reaction time to three different items, K-only, E-only, and baseline, and whether this is mediated by learners' proficiency levels. A 3 (Group: L2-English intermediate, L2-English advanced, L1-English) x 3 (Item type: K-only, E-only, baseline) mixed ANOVA was run. Results showed that there was no significant main effect of Group ($F_{1,12} = 1.547$, $p = .252$, partial eta squared = .205), while a significant main effect of Item type was found ($F_{2,24} = 22.328$, $p < .001$, partial eta squared = .650). However, there was no significant interaction effect of Group and Item type ($F_{4,24} = 1.482$, $p = .239$, partial eta squared = .198).

As a follow-up to finding the significant main effect of Item type, a one-way ANOVA was performed for each group. First, for the L1-English group, there was a significant difference in reaction time between the Item types ($F_{2,8} = 6.080$, $p = .025$, partial eta squared = .603). A Post-hoc analysis revealed that E-only items were processed significantly faster than K-only items ($p = .022$), whereas the remaining comparisons all failed to reach statistical significance. Similar results were found for the L2-English advanced group where the reaction time between each item type differed significantly ($F_{2,8} = 5.193$, $p = .036$, partial eta squared = .565). Post hoc analysis revealed that E-only items were processed significantly faster than K-only items ($p = .013$), while the other comparisons (each mean differences between the K-only and baseline and between the E-only and baseline) failed to reach statistical significance. Lastly, for the L2-English intermediate group, a significant difference in processing time between the item type was found ($F_{2,8} = 17.608$, $p = .001$, partial eta squared = .815). This was followed by a post hoc analysis, showing that the E-only items were processed significantly faster than K-only items ($p = .002$), and baseline items were significantly processed faster than K-only items ($p = .014$).

Table 2

Summary of Results

	Reaction time (ms)			Meaning recall test		
	K-only	E-only	Baseline	K-only	E-only	Baseline
L1-English	1733.33 (437.19)	1339.67 (541.95)	1626.98 (511.82)	5.8 (2.68)	9.6 (3.65)	3.8 (2.49)
L2-English advanced	2158.59 (280.69)	1661.62 (160.13)	2026.11 (496.45)	8.2 (4.09)	9.0 (4.80)	2.4 (1.52)
L2-English intermediate	2340.20 (519.01)	1712.41 (540.37)	1840.83 (274.48)	13.60 (1.34)	8.4 (4.51)	7.20 (4.66)

Analysis of meaning recall test results

This study also aimed to measure learners' ability to provide explicit knowledge of each item type and whether this is mediated by learners' proficiency levels. As mentioned earlier, each item was scored with different scoring criterion: (a) E-only: One point was awarded for the responses containing the correct meaning of each item, otherwise zero points; (b) K-only: One point was awarded for the responses that show any attempts to provide the meaning of the given item or provide the meaning that exists in Korean, otherwise zero points; and (c) Baseline: One point was given for the responses that show the uncertainty of the existence of each item, otherwise zero points. A 3 (Group: L2-English intermediate, L2-English advanced, L1-English) x 3 (Item type: K-only, E-only, baseline) mixed ANOVA was performed. Results revealed a significant main effect of Group ($F_{2,12} = 4.290$, $p = .039$, partial eta squared = .417) and Item type ($F_{2,24} = 8.384$, $p = .002$, partial eta squared = .411). However, there was no significant interaction effect between Group and Item type ($F_{4,24} = 2.586$, $p = .063$, partial eta squared = .301).

Due to the significant main effect of Group, a one-way ANOVA was performed for each Item type. First, in terms of E-only items, no significant difference was found between the groups ($F_{2,12} = .095$, $p = .910$, partial eta squared = .016). By contrast, for the K-only ($F_{2,12} = 9.315$, $p = .004$, partial eta squared = .608) and baseline ($F_{2,12} = 3.901$, $p = .050$, partial eta squared = .394), the three groups differed significantly; thus, post hoc analyses were run for each K-only and baseline to identify where significant differences occurred. For the K-only items (with the scoring criterion that awards responses with any attempts to provide the meaning of the given item as correct), results showed that learners in the L2-English intermediate group produced significantly more attempts than those in the L2-English advanced group ($p = .013$). In addition, learners in the L2-English intermediate group produced responses significantly more than those in the L1-English group ($p = .001$). By contrast, no significant differences occurred between learners in the L2-English advanced and L1-English group. In a similar vein, in terms of the baseline items, the three groups differed significantly ($F_{2,12} = 3.901$, $p = .050$, partial eta squared = .394). Post hoc analyses revealed that learners in the L2-English intermediate group

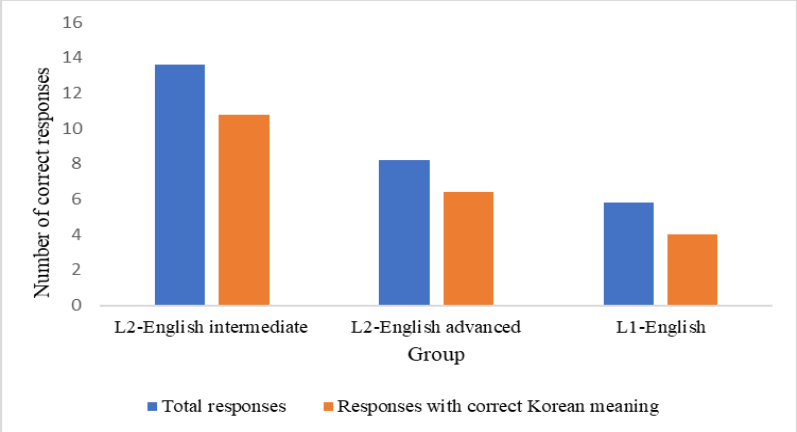
yielded significantly more correct responses than those in the L1-English group (p = .020), whereas the other comparisons failed to reach statistical significance.

Next, due to the significant main effect of Item type, a one-way ANOVA was performed for each participant group. First, in terms of the L1-English group, a significant difference between item types was found ($F_{2,8}$ = 32.062, p < .001, partial eta squared = .889). Post hoc analyses showed that E-only items were scored significantly better than K-only (p = .004) and baseline items (p = .005), respectively. However, the mean differences between K-only and baseline items were not significant (p = .230). For the L2-English advanced group, a significant difference between item types was found ($F_{2,8}$ = 5.969, p = .026, partial eta squared = .599). This was followed by post hoc analyses and revealed that E-only items were scored significantly higher than the baseline (p = .021). In addition, K-only items were scored significantly higher than the baseline (p = .025). However, there was no significant difference between the K-only and E-only items. Lastly, regarding the L2-English intermediate group, the mean differences between each item type reached the level of statistical significance ($F_{2,8}$ = 4.969, p = .040, partial eta squared = .554). Post hoc analyses revealed that the mean differences between the E-only and K-only items (p = .068) and between the E-only and baseline were marginally significant (p = .079).

As mentioned above, an additional scoring criterion was applied for K-only items, given the presence of the literal translation of K-only items in Korean. Therefore, responses that contain only the correct Korean meanings were further analyzed by performing a one-way ANOVA (see Figure 4 below). As can be seen in Figure 4 below, results showed that the three groups differed significantly ($F_{2,12}$ = 5.663, p = .019, partial eta squared = .486). Post hoc analyses revealed that learners in the L2-English intermediate group performed significantly better than those in the L1-English group (p = .018), whereas the other comparisons were all not significant.

Figure 4

Meaning Recall Test Results for the K-only Items



Discussion

Two theories can account for the faster processing of congruent collocations over incongruent ones. One theory is called AoA effects (Carroll & White, 1973), suggesting that

earlier-learned words can be recognized faster than later-learned words (Izura & Ellis, 2002). According to this, a word-for-word translation of a congruent collocation has been already built in learners' L1, which can accelerate its processing in L2. Another theory is based on Jiang's (2000) model of lexical acquisition, which posits that the activation of L1 translation is inevitable in the process of learning L2 words. Thus, the processing advantage of congruent collocations can be attributed to learners' automatic activation of L1 translation. Wolter and Yamashita (2015) found that there was no L1 activation on processing collocations that only exist in learners' L1 (Japanese) but not in L2 (English), thus favoring AoA effects (Carroll & White, 1973) rather than Jiang's (2000) model. However, how generalizable are these findings? This study reports the findings of a coursework paper that sets out to replicate the findings of Wolter and Yamashita's (2015).

The first research question asked whether there are any differences in the performance on the lexical decision task based on different types of collocations (K-only, E-only, and baseline), as a function of learners' different proficiency levels. To this question, I hypothesized that unlike native English speakers, L2 English learners would process K-only items more rapidly than the other two item types (E-only and baseline) and this effect would be driven by L2 English intermediate learners rather than advanced learners. Contrary to my hypothesis, however, results showed that K-only items took the longest time to process, irrespective of learners' proficiency levels. Specifically, for L2 English intermediate learners, K-only items were processed significantly longer than E-only (p = .002) and baseline (p = .014) items, respectively. Similarly, L2 advanced learners spent significantly more time processing K-only items than E-only ones (p = .013). This result, therefore, can substantiate the findings of Wolter and Yamashita's (2015) study that no L1 activation was found, thus supporting AoA effect (Carroll & White, 1973) over Jiang's (2000) model to account for the processing advantage of congruent collocations over incongruent ones.

The second research question asked whether there are any differences in the ability to provide the meaning of each occurring expression based on different types of collocations (K-only, E-only, and baseline), and whether this effect is mediated by learners' proficiency levels. For E-only items, native English speakers were expected to show the most superior performance, followed by L2 advanced and L2 intermediate learners. Interestingly, however, the three groups did not significantly differ in providing the meaning of E-only items. This result can suggest that L2 learners can show comparable performance to L1 speakers, as far as the explicit knowledge of the form-meaning link is concerned. In contrast, with regard to K-only items, L2-English speakers, regardless of their English proficiency levels, provided significantly more responses that are semantically relevant to the expressions that only exist in Korean. This result, even if only partially, supports Jiang's (2000) model over AoA effect (Carroll & White, 1973) by reflecting L2 learners' tendency to rely on the L1 lexical system, despite the absence of the corresponding form of each K-only item in English.

A noteworthy finding of the present study lies in the discrepancy between the lexical decision task and the meaning recall test result. To reiterate, the lexical decision task result is compatible with AoA effect (Carroll & White, 1973), whereas the meaning recall test aligns with predictions emerging from Jiang's (2000) model. This incongruity between on-line (i.e., lexical decision task) and off-line (i.e., meaning recall test) measures can be explained in several ways. First, the hypothesis of Wolter and Yamashita's (2015) study warrants further examination on whether there was no L1 effect on processing J-only items (expressions exist in Japanese but not in English), compared to the processing of E-only items (expressions exist in English but not in

Japanese), and baseline (a random combination of two words that do not exist either in Japanese or English). To recap, there was no processing advantage of *J-only* items in Wolter and Yamashita's (2015) study and *K-only* items in the present study; however, these results can, in fact, provide supporting evidence to Jiang's (2000) model. In particular, the lexical decision task in each study asked the participants to judge if the two-word combinations made sense or not. However, note that either *J-only* items in Wolter and Yamashita's (2015) study or *K-only* items in this study were non-existent English expressions; thus, the lexical decision task *per se* might have activated learners' L1 lexical system and then garden-pathed them to linger on those expressions whether they exist in English, resulting in the increased reaction time. In brief, the method of processing different types of L2 collocations varying in the plausibility in either learners' L1 or L2 may not be a useful method of examining the L1 mediation effect suggested by Jiang's (2000) model. Therefore, a further investigation is needed to compare AoA effect (Carroll & White, 1973) versus Jiang's (2000) model to find an optimal account.

An alternate explanation to the differences between the lexical decision task and meaning recall test results may be attributable to the different level of task difficulty. Specifically, the lexical decision task required the participants to choose either "Yes" or "No", while the meaning recall test required participants to recall and provide the meaning of the given English expressions. Thus, in terms of the lexical decision task, L2 learners can make a rapid judgment on the legitimacy of the given English expression, regardless of its item type (i.e., K-only, E-only, and baseline). Put differently, the lexical decision task might have enabled even nonnative English speakers to bypass their L1 activation which would not inhibit the decision process required in tasks demanding the plausibility of each item (Milberg & Blumstein, 1981). By contrast, concerning the meaning recall test, L2 English learners, especially intermediate learners, would have shown more tendency to rely on their L1 to extract the meanings of K-only items that are congruent with expressions that exist in Korean, without pondering upon their legitimacy in English.

Limitations

Despite the contribution of the present study, several limitations need to be addressed. First, since this study was conducted as part of a coursework, an IRB approval was not obtained. Therefore, it is acknowledged that the findings of the present study are not generalizable. Furthermore, the sample size is small (five participants in each group: native English speakers, English intermediate learners, and English advanced learners), rendering ultimate conclusions only tentative. Next, the number of letters in three item types (i.e., K-only, E-only, and baseline) was not controlled for. Considering the effect of word length on reaction time (Martens & de Jong, 2006), future research should address this issue. Lastly, the items in the present study only included a combination of an adjective and a noun. Given the result of the differential effect of a collocate-node relationship (i.e., adjective plus noun vs. verb plus noun) on learning gains (Nesselhauf, 2003), it would be important to replicate these findings with different types of collocations before drawing any strong conclusions. Such limitations notwithstanding, this study is expected to have significant implications for future research, given that this study has established guidelines for designing sound research studies using both *on-line* and *off-line* measures.

Conclusion

The present study aims to replicate and extend Wolter and Yamashita's (2015) study by employing both on-line (lexical decision task) and off-line measures (meaning recall test). The lexical decision task result showed similar results with Wolter and Yamashita (2015) that there was no L1 activation effect by L2 (L1 Korean) learners, regardless of their L2 proficiency levels, adding credence to the age/order of acquisition (AoA) effects (Carroll & White, 1973) over Jiang's (2000) model. The meaning recall test result, on the other hand, showed that L2 learners provided significantly more responses for expressions that are only acceptable in their L1 (Korean), which seems to support Jiang's (2000) model. The discrepancy between on-line and off-line measures suggests that the findings of Wolter and Yamashita (2015) should be reexamined with large-scale investigations that can provide a more generalizable and replicable picture of the mechanisms at play in processing L2 collocations.

References

Bang, J., & Choi, S. (2018). The effects of pictures on short-and long-term retention of congruent and incongruent English collocations. *Studies in English Education, 23,* 529-553. https://doi.org/10.22275/see.23.2.05

Boers, F., & Webb, S. (2018). Teaching and learning collocation in adult second and foreign language learning. *Language Teaching*, *51*(1), 77-89. https://doi.org/10.1017/s0261444817000301

Carroll, J. B., & White, M. N. (1973). Word frequency and age of acquisition as determiners of picture-naming latency. *Quarterly Journal of Experimental Psychology*, *25*, 85–95.

Choi, S. (2017). Processing and learning of enhanced English collocations: An eye movement study. *Language Teaching Research*, *21*(3), 403-426. https://doi.org/10.1080/14640747308400325

Conklin, K., & Schmitt, N. (2008). Formulaic sequences: Are they processed more quickly than nonformulaic language by native and nonnative speakers?. *Applied linguistics*, *29*(1), 72-89. https://doi.org/10.1093/applin/amm022

Durrant, P., & Schmitt, N. (2009). To what extent do native and non-native writers make use of collocations. *International Review of Applied Linguistics in Language Teaching (IRAL)*, *47*(2), 157-177. https://doi.org/10.1515/iral.2009.007

Erman, B., & Warren, B. (2000). The idiom principle and the open choice principle. *Text-Interdisciplinary Journal for the Study of Discourse*, *20*(1), 29-62. https://doi.org/10.1515/text.1.2000.20.1.29

FindingFive Team (2019). FindingFive: A web platform for creating, running, and managing your studies in one place. FindingFive Corporation (nonprofit), NJ, USA. https://www.findingfive.com.

Henriksen, B. (2013). Research on L2 learners' collocational competence and development – a progress report. In C. Bardel, C. Linquist & B. Laufer (Eds.), *L2 vocabulary acquisition, knowledge and use: New perspectives on assessment and corpus analysis* (pp. 29–56). Eurosla.

Hill, J. (2000). Revising priorities: From grammatical failure to collocational success. In: M. Lewis (Ed.), *Teaching collocation: Further developments in the lexical approach* (pp. 47–69). Hove: Language Teaching Publications.

Izura, C., & Ellis, A. W. (2002). Age of acquisition effects in word recognition and production in first and second languages. *Psicológica*, *23*(2).

Jiang, N. (2000). Lexical representation and development in a second language. *Applied Linguistics*, *21*, 47–77. https://doi.org/10.1093/applin/21.1.47

Martens, V. E., & de Jong, P. F. (2006). The effect of word length on lexical decision in dyslexic and normal reading children. *Brain and language*, *98*(2), 140-149. https://doi.org/10.1016/j.bandl.2006.04.003

Milberg, W., & Blumstein, S. E. (1981). Lexical decision and aphasia: Evidence for semantic processing. *Brain and language*, *14*(2), 371-385. https://doi.org/10.1016/0093-934x(81)90086-9

Nattinger, J. R., & DeCarrico, J. S. (1992). *Lexical phrases and language teaching*. 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. https://doi.org/10.1093/applin/24.2.223

Papageorgiou, S., Tannenbaum, R. J., Bridgeman, B., & Cho, Y. (2015). The association between TOEFL iBT® test scores and the Common European Framework of Reference (CEFR) levels. *Educational Testing Service*, 1-28.

Peters, E. (2016). The learning burden of collocations: The role of interlexical and intralexical factors. *Language Teaching Research*, *20*(1), 113-138. https://doi.org/10.1177/1362168814568131

Schmitt, N., & Carter, R. (2004). Formulaic sequences in action: An introduction. In N. Schmitt (Ed.), *Formulaic sequences: Acquisition, processing, and use*. Philadelphia, PA: John Benjamins. https://doi.org/10.1075/lllt.9.02sch

Siyanova, A., & Schmitt, N. (2008). L2 learner production and processing of collocation: A multi-study perspective. *Canadian Modern Language Review*, *64*(3), 429-458. https://doi.org/10.3138/cmlr.64.3.429

Sonbul, S., & Siyanova-Chanturia, A. (2021). Research on the on-line processing of collocation: Replication of Wolter and Gyllstad (2011) and Millar (2011). *Language Teaching*, *54*(2), 236-244. https://doi.org/10.1017/s0261444819000132

Sprenger, S. A., Levelt, W. J., & Kempen, G. (2006). Lexical access during the production of idiomatic phrases. *Journal of memory and language*, *54*(2), 161-184. https://doi.org/10.1016/j.jml.2005.11.001

Tsai, M. H. (2020). The effects of explicit instruction on L2 learners' acquisition of verb–noun collocations. *Language Teaching Research*, *24*(2), 138-162. https://doi.org/10.1177/1362168818795188

Webb, S., & Kagimoto, E. (2011). Learning collocations: Do the number of collocates, position of the node word, and synonymy affect learning? *Applied Linguistics*, *32*(3), 259-276. https://doi.org/10.1093/applin/amq051

Wolter, B., & Gyllstad, H. (2013). Frequency of input and L2 collocational processing: A comparison of congruent and incongruent collocations. *Studies in Second Language Acquisition*, *35*(3), 451-482. https://doi.org/10.1017/s0272263113000107

Wolter, B., & Yamashita, J. (2015). Processing collocations in a second language: A case of first language activation? *Applied Psycholinguistics*, *36*(5), 1193-1221. https://doi.org/10.1017/s0142716414000113

Wray, A. (2002). *Formulaic language and the lexicon*. Cambridge University Press. https://doi.org/10.1017/cbo9780511519772

Wray, A., & Perkins, M. R. (2000). The functions of formulaic language: An integrated model. *Language & Communication*, *20*(1), 1-28. https://doi.org/10.1016/s0271-5309(99)00015-4

Yamashita, J., & Jiang, N. (2010). L1 influence on the acquisition of L2 collocations: Japanese ESL users and EFL learners acquiring English collocations. *TESOL Quarterly*, *44*, 647–668. https://doi.org/10.5054/tq.2010.235998