




RESEARCH ARTICLE

Development of automaticity in processing L2 collocations: The roles of L1 collocational knowledge and practice condition

Hyojin Jeong^{1*}  and Robert DeKeyser²

¹University of Maryland, College Park, MD, USA and Sungkyunkwan University, Seoul, Korea; ²University of Maryland, College Park, MD, USA

*Corresponding author. E-mail: bjtoto@skku.edu

(Received 21 December 2021; Revised 23 November 2022; Accepted 07 December 2022)

Abstract

This study examined the development of automaticity in processing L2 collocations, and the roles of L1 collocational knowledge and practice conditions in the development process. Korean learners of English were assigned to one of two practice conditions (practice in identical or varied contexts). The learning gains for word combinations with and without equivalent counterparts in the L1 (L1-only and L2-only collocations) were assessed using response times (RTs) and coefficients of variation (CV) from a phrasal decision task. The results demonstrated that the learners in both groups showed significantly improved collocation processing for both types of items in terms of speed (RT) and automaticity (CV) over time. The RT and CV analyses indicated that both groups' improvements in collocation processing in the later stages of learning were associated with automatization. Interestingly, L1 collocational knowledge played a facilitative role in processing speed only in the early stages of learning. No reliable evidence for the differential effects of the two types of practice conditions on developing automaticity in collocation processing was found.

Introduction

Automatic lexical processing is essential for successful communication and fluent language use (Hulstijn, 2001; Schmitt, 2008). For example, a number of previous studies demonstrated a significant association between automatic word recognition skill and fluent reading performance (Favreau & Segalowitz, 1983; Koda, 1996; Schoonen et al., 2003; Segalowitz & Hulstijn, 2005; Van Gelderen et al., 2004). Moreover, it is widely believed in L1 and L2 reading studies that the efficiency of the lower-order components, such as lexical access, is a prerequisite for the higher-order ones, such as text comprehension (Ehri, 1991; Perfetti, 2007; Segalowitz, 2010). Developing fluency in low-level processes frees up a greater amount of processing resources, allowing them to be used for higher level processing (Grabe, 2009; Harrington & Jiang, 2013). Thus, this automaticity enhances the fluidity and efficiency of

underlying cognitive processing and improves the quality of performance as a whole (Segalowitz, 2000, 2010).

To date, a large number of empirical studies have investigated the development of automaticity in L2 lexical processing under incidental or intentional learning condition using the coefficients of variation (CV) measure (e.g., Elgort, 2011; Hui, 2020; Pellicer-Sánchez, 2015; Segalowitz & Segalowitz, 1993). The results overall showed that repetitive encounters or various types of explicit exercises led to increasing speed and automaticity with which learned words are processed. Although informative, the focus of these studies has been limited to single-word processing; no study has investigated the development of automatic lexical access to newly learned multiword units. Given the ubiquity of multiword units and their crucial roles in language use (e.g., Erman & Warren, 2000; Kremmel et al., 2015; Martínez & Murphy, 2011; Wray & Perkins, 2000), however, automatic processing of multiword units can be equally or even more crucial for accurate and fluent language use than single-word processing. Specifically, the importance of having automatic access to multiword units in production of fluent speech has been highlighted in previous studies (e.g., Boers & Lindstromberg, 2012; Wood, 2009).

Among different types of multiword units, such as collocations, idioms, and lexical bundles, the current study focuses on collocations—grammatically well-formed combinations of words that frequently co-occur (Kjellmer, 1994). Collocations usually do not have a high frequency of occurrence (Webb et al., 2013). In terms of compositionality, they may vary between combinations that are fully decomposable (e.g., *pay a salary*) and patterns that involve figurative meaning (e.g., *pay attention*). Owing to such complex and distinct features, the learning process of collocations can differ from that of single words. Some studies have pointed out that factors such as idiomaticity and complexity have negative effects on the learning of lexical items (Laufer, 1997; Peters, 2014).

Despite a great number of studies on learning and processing collocations, limited research has investigated the automatic processing of collocations (e.g., Durrant & Schmitt, 2010; Sonbul & Schmitt, 2013). The focus of these studies was also limited to learning word combinations at the surface level. Moreover, none of these studies examined the role of L1 congruency in the learning process, which has been reported to have a significant influence on learning and processing collocations (e.g., Bulut & Çelik-Yazici, 2004; Nesselhauf, 2003, 2005; Wolter & Gyllstad, 2011, 2013; Wolter & Yamashita, 2018; Yamashita & Jiang, 2010). Congruency is determined based on the word-for-word overlap between L1 and L2 form and meaning (Bahns, 1993). Thus, congruent collocations have a direct translational equivalent that also sounds natural in the L1 (considering, however, the general syntactic rules of the two languages). Conversely, incongruent collocations do not have an equivalent L1 construction and are expressed using a different word or words in the L1.

Moreover, despite the importance of the practice condition for facilitating learning and the call for fluency development activities in language instruction (Nation, 2007; Segalowitz, 2010), studies examining different types of treatment that facilitate the development of automaticity in L2 lexical processing are very scarce (with the exception of Durrant & Schmitt, 2010; Pellicer-Sánchez, 2015; Sonbul & Schmitt, 2013). Although previous studies offered useful suggestions, their scope was limited (i.e., developing knowledge of the composition of a collocation) or did not focus on processes of automatization of multiword processing (i.e., Pellicer-Sánchez, 2015). Therefore, the current study attempts to address research gaps in previous studies on automatic lexical processing by examining the development of automaticity in collocation processing,

and the roles of L1 collocational knowledge and different types of practice conditions in the development process.

Literature review

CV as an index for automaticity

It is generally agreed that automatic processing is highly rapid and largely error-free. However, fast processing, despite its prominent role, is not sufficient to index automaticity (e.g., Segalowitz & Segalowitz, 1993; Segalowitz et al., 1998). Ehri (1991) argued that the ultimate goal in word recognition development is associated with restructuring a word-recognition mechanism or with increasing cognitive efficiency in word recognition. Segalowitz and Segalowitz (1993) proposed a similar point—simply being able to recognize words quickly is not sufficient to claim automatic lexical processing (e.g., Segalowitz & Segalowitz, 1993; Segalowitz et al., 1998). They argued that automatization is associated with increased cognitive efficiency involving changes in how underlying processes are organized or in the internal organization of a given process. Thus, when learners become more proficient, lexical processing, for instance, becomes less variable, as controlled and more effortful mechanisms drop out and are replaced with more efficient processes (Segalowitz & Segalowitz, 1993; Segalowitz et al., 1998).

To assess whether improved lexical processing is a result of a qualitative change in underlying processes, Segalowitz and Segalowitz (1993) proposed an innovative approach using relative variability of performance (as measured by CV). CV is a standard deviation (SD) of RT divided by mean RT; CV reflects processing stability with individuals' processing speed taken into account. According to this conceptualization, when one improves performance by simply speeding up its component processes, mean RT and SD decrease proportionally, resulting in no change in CV. However, when one's performance improves because of increased efficiency in a processing mechanism, SD decreases more drastically than mean RT and, thus, CV is reduced. Given a set of mean RT and CV pairs for a group of participants, the correlation between mean RT and CV should be positive. This approach has great practical potential, allowing a researcher to assess whether the improved performance is attributable to generalized speedup or qualitative changes in component processes. This proposal of relative variability of performance as a valuable indicator of automaticity has been supported in a series of empirical studies carried out by Segalowitz and his associates (Phillips et al., 2004; Segalowitz, 2000; Segalowitz & Segalowitz, 1993; Segalowitz et al., 1998) as well as other researchers (Akamatsu, 2008; Elgort, 2011; Harrington, 2006; Hui, 2020; Lim & Godfroid, 2015; Suzuki & Sunada, 2018). For example, Hui (2020) examined the early and later stages of intentional and incidental word learning and revealed a roughly inverted U-shaped development in CV in the intentional word learning experiment. In the initial stage of learning, CV increased as a result of establishment of new representations in the lexicon, followed by a decrease in CV, indicating the automatization of such knowledge. Although a similar trajectory in CV was not observed in the incidental word learning experiment, significantly more stable processing of high-frequency control words (with reduced CV) was found in comparison to that of low-frequency target words due to participants' prior word knowledge.

Hulstijn et al. (2009), however, made some cautioning remarks regarding using CV as an index for automatization based on some cases of failure to replicate CV effects and

potential confounds between improvement of knowledge and improvement of processing efficiency. Specifically, Hulstijn et al. (2009) analyzed previously collected RT data of secondary-school learners and failed to find evidence for automatization as defined by Segalowitz and Segalowitz (1993) in four out of 14 CV analyses. However, in a follow-up study using the same tasks and stimuli as Hulstijn et al.'s (2009) experiment 1 but with slight modification, Lim and Godfroid (2015) reported CV effects in support of Segalowitz and Segalowitz (1993). As was pointed out by Lim and Godfroid (2015), the lack of CV effects reported by Hulstijn et al. (2009) could be due to the fact that the secondary-school learners in their studies may have not yet progressed to the automatization stage due to the lack of sufficient exposure and practice. In addition, the methodological concern of potential confounding of improvement of knowledge and improvement of processing efficiency in CV measurement has often been addressed by limiting the use of CV to those learners with sufficient levels of proficiency as in Lim and Godfroid (2015). This concern has also been addressed in instruction studies by including a study phase to ensure solid declarative knowledge of the target structure or vocabulary before having learners engage in procedural learning and testing their performance, in the current study as well as others (e.g., Li & DeKeyser, 2019; Suzuki, 2017). At the same time, ensuring solid declarative knowledge is essential for improving processing efficiency. According to skill acquisition theory (Anderson, 1992; DeKeyser, 2015), sufficiently solid declarative knowledge is necessary for successful proceduralization and automatization. That is, declarative knowledge must be active enough for procedural knowledge to be developed (Kim et al., 2013). Kim et al. (2013) also suggested that procedural rules can only be generated when declarative knowledge can be retrieved quickly enough. In the present study, RT and CV measures were used to assess automaticity in collocation processing. The following sections will discuss potential factors affecting collocation learning in detail.

L2 collocation learning and L1 collocational knowledge

To date, collocations have been increasingly researched and recognized as an important part of vocabulary knowledge, playing a crucial role in enabling the comprehension and expression of messages (e.g., Boers & Lindstromberg, 2012). Despite their significant role, however, a number of studies have shown limited mastery of collocations even by highly advanced L2 speakers living in an immersion context (e.g., Abrahamsson & Hyltenstam, 2009; Granena & Long, 2013; Spadaro, 2013; Wray, 2002).

One crucial factor that influences learning and processing L2 collocations is congruency with L1. Studies have shown that L2 learners' performance on incongruent collocations often lagged behind that of congruent collocations, and facilitation effects have been consistently found on recognition speed for congruent collocations over incongruent collocations (e.g., Sonbul & El-Dakhs, 2020; Wolter & Gyllstad, 2011, 2013; Yamashita & Jiang, 2010). For example, Sonbul and El-Dakhs (2020) examined the effects of congruency and L2 proficiency, and the interaction between the two factors on timed and untimed collocation recognition. A multiple-choice test and a timed acceptability judgment task were given to L2 EFL learners. The results showed that congruency and estimated proficiency had effects on both untimed and timed recognition. Moreover, there was a clear interaction effect on timed recognition only, showing a gradual decrease in the congruency effect as proficiency increased. This was interpreted as the learners showing more nativelike collocation processing with increasing proficiency.

Previous studies have often drawn on the account proposed by Yamashita and Jiang (2010) to explain how L1 knowledge may be involved in learning and processing L2 collocations. Based on Jiang's model of L2 lexical development (see Jiang, 2000, 2004), it is suggested that learning an L2 collocation follows three steps. The first step consists of recognizing a word string as a legitimate collocation in L2, which generally takes place with understanding its meaning. The second step is the integration of the collocation in long-term memory through repeated exposure and practice. In the first two steps, an L2 collocation is primarily connected to its L1 counterpart or both its L1 counterpart and corresponding concept (but more strongly to the former than the latter). Thus, the use of L2 collocations is mediated by their L1 counterparts; activation of L1 counterparts in L2 use, however, may decrease with increasing experience in L2. The third step involves establishing a direct link between the new collocation and the concept that is strong enough to allow automatic access to it without activating L1 equivalents. Under this hypothesis, Yamashita and Jiang (2010) further suggested that with the knowledge of L1 congruent collocations, the acquisition of L2 congruent collocations can be accelerated compared to that of incongruent collocations, especially in the initial stage of learning. However, they also predicted that due to direct reliance on L1 equivalents, making links with concepts may be more demanding and require more practice and exposure for congruent collocations once they are stored in memory. On the contrary, because no L1 counterpart exists for an incongruent collocation, incongruent collocations may be at an advantage in this process to allow automatic processing.

To the best of our knowledge, however, the actual learning trajectory of collocations with and without equivalent counterparts in L1 have never been examined using the CV measure. Such data are important for elucidating the role of L1 knowledge in different stages of learning collocations. Moreover, the data from both RT and CV measures will provide useful insights into how collocations with and without equivalent counterparts in L1 are processed and are represented in adult L2 learners. Note, however, that the current study examined the development process of L1-only collocations (word combinations that are acceptable in participants' L1 but not in the L2) and L2-only (what have been called incongruent in other studies) collocations that are acceptable in the L2 but not the L1. Although congruent collocations (that exist in both L1 and L2) were not directly examined, such data can still allow us to determine the influence of L1 collocational knowledge in learning word combinations in the L2.

Practice condition

Automatization is considered a slow process that requires repeated practice with the same items (e.g., Gatbonton & Segalowitz, 2005). At the same time, practice condition is recognized to be an important factor for creating effective practice that best promotes learning (Suzuki et al., 2019). Considering that the potential for learning most collocations incidentally may be relatively small due to the lack of a sufficient number of exposures (Webb et al., 2013), identifying an effective practice condition that facilitates collocation learning should be of use in both ESL and EFL contexts.

When creating repetitive exercises with meaningful contexts for fostering automaticity in collocation processing, at least two different types of practice conditions can be considered: practice in identical or varied contexts. The effects of repetitive tasks with the same content on fluency development have been empirically shown by a number of studies (e.g., Boers, 2014; De Jong & Perfetti, 2011; Thai & Boers, 2016). The authors of

the three related studies suggested that a repetitive activity with or without time pressure supports the development of fluency by freeing up attentional resources that can be used for improving fluency.

In a study on the incidental learning of collocations, Durrant and Schmitt (2010) examined the effects of two different types of repetition conditions of input. The participants were randomly assigned to one of three different treatment conditions in which they were presented with a series of decontextualized sentences containing the target collocations (single exposure vs. verbatim repetition vs. varied repetition). In the single exposure condition, participants were exposed to each target collocation in a sentence, only once. In the verbatim repetition condition, participants were presented with the same target sentences containing the target collocations twice, while in the varied repetition condition, two different sentences were presented for each target collocation. The results showed that both repetition groups outperformed the single exposure group, with the verbatim repetition group performing significantly better than the varied repetition group. Based on their findings, the authors suggested that when teaching collocations, exercises including such verbatim repetition be given more attention, which enable learners to build up fluency with particular strings of language. They argued that when learners are first exposed to a certain language input, a number of cognitive demands (e.g., recognizing the words, decoding the syntax, and generating the appropriate semantic context) that learners have to deal with may prevent learning. However, when exposed to the same language input a second time, those issues can be at least partially resolved; thus, learners can focus more on consolidating and building fluency with the language. Sternberg et al. (1983) also noted that learning vocabulary from multiple-sentence contexts can, in certain cases, cause information overload that may hinder the effective use of sentences.

However, based on the desirable difficulty framework, it has been argued that more demanding conditions of learning that require more effort during practice will benefit learning and retention over time (e.g., Bjork & Kroll, 2015; Brown et al., 2014; Schmidt & Bjork, 1992; Suzuki et al., 2019). According to this view, practice in varied context (i.e., more demanding learning condition) may eventually lead to greater learning than that in identical context. Recent evidence showed that different types of practice conditions that induced the appropriate level of difficulty facilitate learning relative to the less challenging practice conditions (e.g., Kim & Godfroid, 2019; Li & DeKeyser, 2019; Pulido & Dussias, 2020; Strong & Boers, 2019; Suzuki & Sunada, 2020). For example, Pulido and Dussias (2020) examined whether a practice condition that induced L1 interference resulted in greater learning gains for incongruent collocations compared to a noninterference practice condition. In support of the desirable difficulty framework, the results showed that the L1-interference group outperformed the unrelated group for the incongruent collocations on a productive recall test.

Considering the relevant theoretical and empirical findings in the previous studies (e.g., Bjork & Kroll, 2015; Brown et al., 2014; Durrant & Schmitt, 2010; Pulido & Dussias, 2020; Schmidt & Bjork, 1992; Suzuki et al., 2019), the role of different types of practice conditions such as practice in identical or varied contexts in learning collocations remains to be explored.

The present study

The present study investigated the development of automaticity in processing collocations and specifically focused on two potential factors that can influence the

development of automatic collocation processing: L1 collocational knowledge and practice conditions (practice in identical or varied contexts). The following are the research questions the current study attempted to address:

1. Does repeated practice with collocations following declarative study lead to the development of automaticity in processing collocations (as reflected by a decrease in CV and statistically positive correlations between CV and RT for the target expressions in the phrasal decision task)?
2. What is the role of L1 collocational knowledge in the development of automaticity in processing collocations?
3. What is the role of different types of practice conditions (identical-sentence context and varied-sentence context) associated with exercises in the development of automaticity in processing collocations?

Methods

Participants

Seventy-eight Korean learners of English who were either graduate or undergraduate students at universities in the United States were recruited for the study. All participants were native speakers of Korean. Six participants did not return after the first session; thus, a total of 72 participants provided data for the study. During the screening process, those learners who scored at the B or C levels on DIALANG, corresponding to independent or proficient users according to the Common European Framework of Reference, were invited to participate in the study (Council of Europe, 2001). Based on the pilot test, such proficiency requirements were determined sufficient and necessary to complete the vocabulary exercises and tasks in the present study.

The participants were randomly assigned to one of the two experimental conditions (identical- or varied-sentence condition). Participants in the two groups were comparable in terms of mean age, $F(3, 64) = .454, p = .716$, and gender distribution (Pearson Chi-Square test, $X^2 = 2.776, p = .096$). They were also comparable in DIALANG scores ($U = 589.5, z = -.80, p = .424$) and length of residence in the United States, $F(3, 64) = .454, p = .716$. Table 1 summarizes the background information of the participants.

Table 1. Description of the participants

	Identical-sentence group (N = 36)	Varied-sentence group (N = 36)
Age	Mean = 24.64 Range = 19–35	Mean = 24.56 Range = 19–35
Gender	Female = 17 Male = 19	Female = 24 Male = 12
DIALANG scores	B1 (n = 2) B2 (n = 28) C1 (n = 5) C2 (n = 1)	B1 (n = 4) B2 (n = 21) C1 (n = 8) C2 (n = 3)
Years of residence in the United States	Mean = 4.65 Range = 0.33–6.5	Mean = 4.47 Range = 0.67–6.67

Materials

Target items

The critical items for learning consisted of 16 L1-only collocations (verb-noun combinations that exist in Korean but not in English) and 16 L2-only (incongruent) verb-noun collocations in English, both of which are not semantically transparent. Thus, the meaning of these critical items cannot be comprehended by merely combining the meaning of individual words. It is important to note that, in the present study, direct word-for-word translations of Korean collocations that do not exist in English (e.g., *roll money*) were used as the target items rather than actual congruent collocations. Pilot tests showed that learners tend to identify the correct meaning of unknown congruent collocations in the L2 based on their knowledge of L1 collocations, especially when constituent words were already known. Thus, it was almost impossible to identify a sufficient number of congruent verb-noun collocations that were potentially new to participants using frequency counts and pilot tests. Using translated forms of L1 expressions to investigate the influence of L1 knowledge on understanding L2 phrases is not uncommon (Carrol & Conklin, 2014, 2017; Wolter & Yamashita, 2015, 2018). More importantly, although artificially constructed items were included, the participants were told that all the target items to be learned were authentic English expressions; they did not know that artificially constructed items were used until the debriefing. Thus, learning those items can be considered comparable to the experience of learning actual congruent collocations.

For the selection and construction of the target lexical materials, a list of potential L2-only verb-noun collocations was first identified using various resources such as collocations dictionaries and textbooks (e.g., O'Dell & McCarthy, 2008; Oxford University Press, 2009). The selected collocations were then reduced based on the following criteria: (a) they could not have direct translational equivalents in Korean, which were attested by the nonexistence of their translations in two Korean collocation dictionaries (Kim, 2007; Kim & Joo, 2016) and the Korean language corpus (a corpus of 36 million words created by the National Institute of Korean Language); (b) their constituent words had to be highly familiar to the participants (the majority of individual words were found in the lists of 3,000 basic words for English education constructed by the National Curriculum Information Center in Korea in 2015); and (c) they needed to co-occur with relatively low frequency to ensure that the participants are unlikely to have any knowledge of them before participating in the experiment (all well below 0.5 times per million occurrences according to the Corpus of Contemporary American English [COCA]). Finally, the candidate collocations were piloted with 12 L2 learners similar to the experiment participants in terms of key language background characteristics such as L2 proficiency, and years of residence, to ensure that they were all essentially unknown. Only the items that were identified as unfamiliar to all the participants remained in the final list (eight items were removed in this process).

For constructing L1-only collocations, possible items were first selected using two Korean collocation dictionaries (Kim, 2007; Kim & Joo, 2016) and translated into English as closely as possible by the first author. The selected items were then checked to ensure they were not acceptable English expressions using the COCA. As Korean does not require overt marking of the number (singular vs. plural) or countability of nouns (count vs. mass) (Ionin, 2003; Ionin et al., 2007; Lee Amuzie & Spinner, 2012), nouns in Korean noun-verb collocations typically appear in bare form, without an article or plural marking. English verb-noun collocations, of course, often include articles between two lexical components (e.g., *beat a retreat*, *tip the scales*). Thus, for comparative purposes,

Table 2. Lexical properties of the target items (16 items in each of the two conditions)

	L1-only	L2-only	<i>t</i>
Length	11.75	11.81	.10
Noun frequency	4.66	4.46	-.92
Verb frequency	4.51	4.55	.34

Note: Length refers to the total number of letters of each expression. Log adjusted frequencies are shown for noun and verb frequencies. All comparisons are pairwise.

articles (definite and indefinite) were inserted in some of the L1-only items to match the phrase structure and length between the two lists of items. The artificially constructed items were then piloted with four native speakers of Korean with high English proficiency to ensure their equivalent status. It should be noted that a great deal of effort has been made to ensure that both item types (L1-only and L2-only items) in the current study are equally new and unknown to the participants before the experiment. Thus, in both conditions, learning would involve the process of establishing membership of constituent words in collocations and links between the target word combinations and their meaning.

A summary of the lexical properties of these test items can be found in Table 2. There were no major differences between the lists of the L1-only and L2-only items in terms of length and individual word frequency. The L1-only and L2-only collocations can be found in Appendix A in the Supplementary Materials.

Sentences used for the fill-in-the-blank exercises

Twelve sentences were created for each target expression to construct the exercise materials for the two practice conditions (practice in identical or varied contexts). All the sentences were taken from English or Korean dictionaries and corpora such as the COCA (e.g., Kim, 2007; Oxford University Press, 2009). However, when necessary, the sentences were slightly modified to ensure all the running words were potentially known by all participants. When selecting and constructing a series of different sentences for any given item, it was assured that the given target expression expressed the same meaning for all the sentences. The constructed sentences were divided into 12 sets, each containing 32 sentential contexts (one for each of the 32 target items). None of the target verbs or nouns in the target items was repeated in any sentence. Sample sentences for the target item, *split hairs*, can be found in Appendix B in the Supplementary Materials.

Stimuli for the phrasal decision task

The test stimuli included 16 L1-only and 16 L2-only collocations and 32 noncollocational filler items. Following previous studies (Wolter & Gyllstad, 2011, 2013; Wolter & Yamashita, 2015, 2018; Yamashita & Jiang, 2010), the filler items consisted of atypical combinations of two high-frequency words such as *jump the smell* or *lay an age*. All high-frequency words were selected from the New General Service List (Browne et al., 2013).

Treatment

Declarative collocation study phase

All the tasks and tests involved in the treatment were presented by computer. Before engaging in practice treatment (fill-in-the-blank exercises) on Day 1, participants

engaged in learning declarative knowledge of the 32 target collocations. In the study phase, the participants went through three sessions of presentation of the target items, each followed by a meaning recall test. In each presentation session, each target item to be learned (in English) with its meaning in Korean was presented on the screen for 10 seconds. After every two cycles of all 32 target items, a short recall test with different target items each (16 items) was given to test explicit knowledge of the target items; for the last session (session 3), however, a combined test of all 32 items was administered. The participants were tested until they demonstrated complete explicit knowledge of the target items. The declarative study phase was included to ensure that the participants had the declarative knowledge of the form and meaning of each target item before engaging in the fill-in-the-blank exercises for procedural learning.

Both the second and third sessions (Day 2 and 3) started with a declarative collocation review in which 32 target items learned on Day 1 were tested again in an identical manner to that of the recall test on Day 1 until they reached 100% accuracy, to prevent any attrition of declarative knowledge.

Practice phase for procedural learning

All the participants in either the identical- or varied-sentence condition were provided with a series of fill-in-the-blank exercises, a format that has been used successfully in previous studies to foster automaticity in L2 word processing (Fukkink et al., 2005; Snellings et al., 2004) and found to be a popular format used in students' course materials for learning collocations (Boers et al., 2017). The fill-in-the-blank exercises were administered through an online testing site, ClassMarker (<http://www.classmarker.com>). In this exercise, the participants were presented with sentences, each containing a gap, where the target item was missing. Four sentences each with a gap, appeared on the screen for each exercise item. The participants were asked to choose the appropriate expression that fit each sentential context among four target expressions provided in the list under time pressure. Each target expression occurred only once in the exercise.

The two groups were given the same number of exercises (four rounds of exercises) in the same exercise format (fill-in-the-blank) in each practice session but the exercises differed in terms of items. That is, during each practice session, participants in the identical sentence condition completed the fill-in-the-blank exercises four times with the same sentence context for each target collocation (the same set of items was repeated throughout all the practice sessions). A participant assigned to the varied-sentence condition, however, completed the fill-in-the-blank exercises four times with four different sets of items in each session (12 different sets of items in total in the three practice sessions together).

In all four rounds, participants were asked to respond as accurately and quickly as possible and received feedback on every response regarding the correctness and the correct answer in case of erroneous responses.

Procedure

Each participant in either the identical- or varied-sentence condition engaged in three practice sessions (one session per day) in a quiet room within a single week. Spacing for each experimental group was almost identical (mostly consecutive); thus, the potential effects of a different distribution of practice is controlled. Table 3 presents the overview

Table 3. Overview of experimental sessions

Day 1	Consent form Declarative study PDT 1 Fill-in-the-blank exercises PDT 2
Day 2	Declarative study Fill-in-the-blank exercises PDT 3
Day 3	Declarative study Fill-in-the-blank exercises PDT 4 Questionnaire

Note: PDT: phrasal decision task.

of experimental sessions. The phrasal decision task was conducted using DMDX (Forster & Forster, 2003).

Upon completing the consent form, the first session (Day 1) started with the declarative collocation study phase. Then, the pretraining phrasal decision task (PDT 1) was given, in which the participants were asked to decide whether or not the word combination presented was acceptable in English by pressing a specified key (Yes/No) on the keyboard as quickly as possible. After instructions and 12 practice items, the experimental items (i.e., 32 collocation and 32 noncollocational control items) were presented in an individually randomized order. Each trial began with a series of asterisks in the center of the screen for 1,000 ms for the purpose of eye fixation. The asterisks were followed immediately by the presentation of an item to which the participant responded. RTs were measured from the onset of an item to a key press. No feedback was provided during the test. The phrasal decision task was followed by a series of fill-in-the-blank exercises (four rounds) with a short break between each round (when necessary). The participants in the identical-sentence condition were assigned in equal numbers to one of the 12 sets of items to control for any effect of a particular item set on their performances while in the varied-sentence group, the same number of participants were randomly assigned to one of the six counterbalanced orders of items sets to control for order effects. The order of the questions in the exercise of each round was randomized across the participants. Upon completing the fill-in-the-blank exercises, the participants completed the post-training phrasal decision task (PDT 2). At the end of the first day's session, the participants were instructed not to study or practice the expressions they learned outside the sessions until the study concluded.

The second session (Day 2) started with the review of declarative knowledge of the target collocations first and then the fill-in-the-blank exercises followed by the post-training phrasal decision task (PDT 3), as illustrated in Table 3.

On Day 3, upon completing the review of declarative knowledge of the target collocations, the participants again carried out fill-in-the-blank exercises followed by the post-training phrasal decision task (PDT 4), and finally the language experience and proficiency questionnaire (Marian et al., 2007).

Data analysis

As recommended and practiced by previous studies (Lim & Godfroid, 2015; Suzuki & Sunada, 2018), only RTs of correct responses were analyzed (resulting in a removal of

2.9% of the RT data); no additional data cleaning procedures were employed. Concurring with Lim and Godfroid (2015), Suzuki and Sunada (2018) specifically suggested that a raw data analysis is preferred for automatization studies using RT and CV measures, especially when accuracy is high (near ceiling) because data cleaning may artificially and incorrectly affect results.

For the statistical analysis, RT and CV scores from the phrasal decision task were analyzed separately using a linear mixed-effects model. R version 2.11.1 (R Core Team, 2018) with the *lme4* (Bates et al., 2015) and *lmerTest* packages (Kuznetsova et al., 2020) was used for the statistical analysis of the RT and CV data. RTs were log transformed before the statistical analysis to reduce skewness in the distribution (Baayen, 2008). For both RT and CV analyses, Time (each testing time, four levels), Item type (L1-only or L2-only items), Condition (identical- or varied-sentence conditions), and their two- and three-way interactions were modeled as fixed effects, and subjects and items as random effects. The fixed effects variables were contrast coded in which time 1 was treated as the reference level for Time, the L1-only items for Item type, and the identical-sentence condition for Condition. Thus, in the mixed-effects models, the intercept represents the grand mean (across all conditions). When fitting the mixed-effects models, random intercepts for participants and/or items were included in all models. After attempts to build a maximal random effects structure (Barr et al., 2013), which led to convergence issues, the random effects structure was simplified (Bates et al., 2015; Cunnings & Finlayson, 2015). Thus, the final model for the RT analysis included by-subject slope for Time in addition to random intercepts for both participants and items, while the final model for the CV analyses included a random intercept for participants and by-subject slope for Time. All models were built as forced-entry models for fixed effects and random intercepts. Effect sizes were computed using the *MuMIn* package in R (Barton, 2016). Both marginal R^2 (the variance explained by the fixed effects only), and conditional R^2 (the variance explained by both the fixed and random effects) are reported for each model. Post-hoc comparisons, whenever needed, were conducted using the R package *emmeans* (Lenth, 2020). To lower the chances of Type I errors from multiple comparisons, a false discovery rate (FDR) procedure was used where appropriate (Benjamin & Hochberg, 1995). In addition to linear mixed-effects modeling, correlational analyses were carried out with RT (log transformed) and CV.

Results

Table 4 presents the descriptive statistics for RTs, SDs, CVs, and accuracy rates for each group across three practice sessions. In addition, the correlations between the RT and CV for each group on each outcome measure is presented together in the table. As illustrated by the descriptive statistics in Table 4, both groups performed near ceiling on all four tests (95%–99% accuracy). Note that the two groups did not differ in terms of accuracy scores for both L1-only and L2-only items at any time point (all $ps > .05$). Thus, participants in both groups engaged in each practice session with a solid declarative knowledge of the target collocations. More importantly, RTs, SDs, and CVs of both types of items generally decreased in both groups across practice sessions at the descriptive level. The results are presented separately for RT and CV.

Changes in RT

The results of the linear mixed-effects modeling are presented in Appendix C in the Supplementary Materials. The analysis of the RT showed that there were highly

Table 4. Means of RT (in milliseconds), SD, CV, accuracy, correlation between CV and log RT

Test	Identical-sentence condition (N = 36)					Varied-sentence condition (N = 36)				
	Mean RT ^a (SD)	Mean SD (SD)	Mean CV (SD)	Accuracy Rate (SD)	r (CV, RT)	Mean RT ^a (SD)	Mean SD (SD)	Mean CV (SD)	Accuracy Rate (SD)	r (CV, RT)
L1-only										
Test 1	996.17 (243.88)	296.14 (170.96)	0.28 (0.11)	0.98 (0.04)	.563**	916.52 (271.33)	249.01 (119.40)	0.26 (0.09)	0.99 (0.03)	.266
Test 2	851.10 (162.26)	225.70 (118.52)	0.25 (0.10)	0.97 (0.05)	.766**	857.00 (209.52)	261.08 (150.90)	0.29 (0.11)	0.96 (0.06)	.633**
Test 3	805.12 (137.23)	206.58 (101.53)	0.25 (0.09)	0.97 (0.04)	.640**	771.24 (138.72)	201.92 (125.97)	0.25 (0.11)	0.97 (0.04)	.700**
Test 4	771.51 (144.29)	197.27 (115.82)	0.24 (0.10)	0.97 (0.04)	.642**	710.34 (115.23)	152.55 (74.63)	0.21 (0.08)	0.95 (0.06)	.465**
L2-only										
Test 1	1046.33 (253.27)	339.73 (199.92)	0.31 (0.11)	0.98 (0.03)	.570**	964.06 (286.18)	286.54 (128.26)	0.29 (0.10)	0.99 (0.03)	.208
Test 2	945.23 (229.21)	299.24 (197.05)	0.29 (0.13)	0.97 (0.04)	.815**	886.88 (215.95)	273.57 (174.59)	0.29 (0.13)	0.97 (0.05)	.664**
Test 3	817.98 (123.15)	208.51 (113.08)	0.24 (0.10)	0.96 (0.05)	.725**	784.32 (164.17)	204.80 (154.12)	0.24 (0.13)	0.96 (0.04)	.647**
Test 4	768.73 (126.20)	175.40 (93.67)	0.22 (0.10)	0.96 (0.06)	.500**	734.54 (129.06)	174.80 (112.87)	0.23 (0.11)	0.96 (0.05)	.650**

**p < .01.

^aRTs were log transformed for the statistical analysis, but RT values presented in this table are mean results before the transformation.

significant effects of Time, with significantly lower RTs for both item types after each practice session (all $ps < .001$). A marginally significant effect of Item type was also observed ($t = 1.81, p = .08$), with slower RTs for the L2-only items. However, there was no significant effect of Condition ($t = -1.58, p = .12$). Critically, an interaction between Time and Item type was significant (Time 3, $t = -2.79, p = .005$, and Time 4, $t = -2.76, p = .005$, respectively). The rest of the two-way interactions and a three-way interaction of Time, Item type, and Condition were not statistically significant: the interaction between Condition and Time, and the interaction between Condition and Item type and the three-way interaction among Condition, Time, and Item type (all $ps > .05$).

Following the significant interaction of Time and Item type, post-hoc comparisons with FDR adjustments for multiple comparisons revealed significant mean differences over time for both L1-only and L2-only items in both groups (Appendix D in the Supplementary Materials). Specifically, for each item type, both groups' RTs decreased significantly between times, as shown in Figure 1 (all except from Time 1 to Time 2 for the L1-only items in the varied-sentence group, $p = .06$ with FDR correction). For both groups, RTs for the learned items at Times 3 and 4 were always significantly lower than those of preceding tests, indicating faster processing speed of the target items by the participants over time.

Crucially, the pairwise comparisons also revealed that a processing advantage for the L1-only items was only present at Times 1 and 2 in the identical-sentence group and at Time 1 in the varied-sentence group (Appendix E in the Supplementary Materials). That is, RTs to the L1-only items were significantly faster than to the L2-only items at Times 1 and 2 in the identical-sentence group and at Time 1 in the varied-sentence group ($ps < .05$). Thus, the interaction between Time and Item type showed significant processing advantages afforded by L1 collocational knowledge at Time 1, before completing exercises for procedural learning, and Time 2, after the first practice session in comparison to the lack of such advantages at Times 3 and 4. In sum, both groups' speed of performance for both types of learned items became significantly faster over three practice sessions. A significant advantage for the L1-only items was observed in

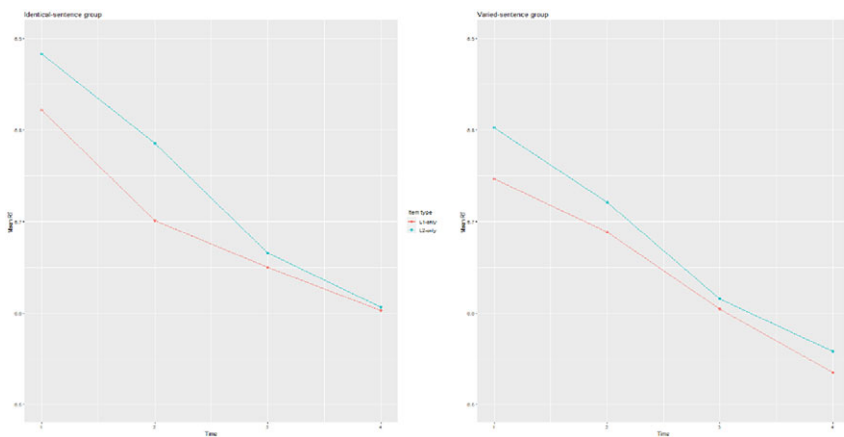


Figure 1. Mean RT (log transformed) for two item types across practice sessions for the identical-sentence group (left) and the varied-sentence group (right).

processing time only at Times 1 and 2 in the identical-sentence group and Time 1 in the varied-sentence group.

Changes in CV

The analysis on CV showed significant effects of Time; relative to CVs of Time 1, CVs were significantly lower at Time 3 ($t = -3.15, p < .01$), after the second practice session and at Time 4, after the third practice session ($t = -5.38, p = .001$). There was a marginally significant effect of Item type, with overall larger CVs for the L2-only items ($t = 1.68, p < .09$). The effect of Condition was not significant ($t = -0.30, p = .76$). The interaction between Time and Item type was marginally significant at Time 3 ($t = -1.81, p = .07$) and Time 4 ($t = -1.77, p = .08$). The rest of the two-way interactions and the three-way interaction of Time, Item type, and Condition were not statistically significant: the interaction between Condition and Time, the interaction between Condition and Item Type, and the three-way interaction between Condition, Time and Item Type (all $ps > .05$). It was observed in the follow-up tests (Appendix F in the Supplementary Materials) that for the identical-sentence group, although descriptively CVs for the L1-only items decreased between times, this decrease in CVs was marginally significant for the L1-only items only at Time 4 compared to Time 1 ($p = .09$). However, CVs for the L2-only items significantly decreased in the later sessions; the decrease in CVs was significant at Time 3 compared to Times 1 and 2, and significant at Time 4 compared to Times 1 and 2, (all $ps < .05$). For the varied-sentence group, CVs for both items significantly decreased in the later sessions. Specifically, the decreases in CVs for the L1-only items approached significance at Time 3 compared to Time 2 ($p = .06$), but were significant at Time 4 compared to Times 1 and 2 ($ps < .05$ and between Time 3 and Time 4 with approaching significance, $p = .06$). The decreases in CVs for the L2-only items were marginally significant at Time 3 compared to Times 1 and 2 ($p = .06, p = .05$, respectively) but significant at Time 4 compared to Times 1 and 2 (all $ps < .05$). Figure 2 graphically presents the change in CV for both item types in the two groups across practice sessions. The results of the follow-up tests (Appendix G in the Supplementary Materials) also revealed that there

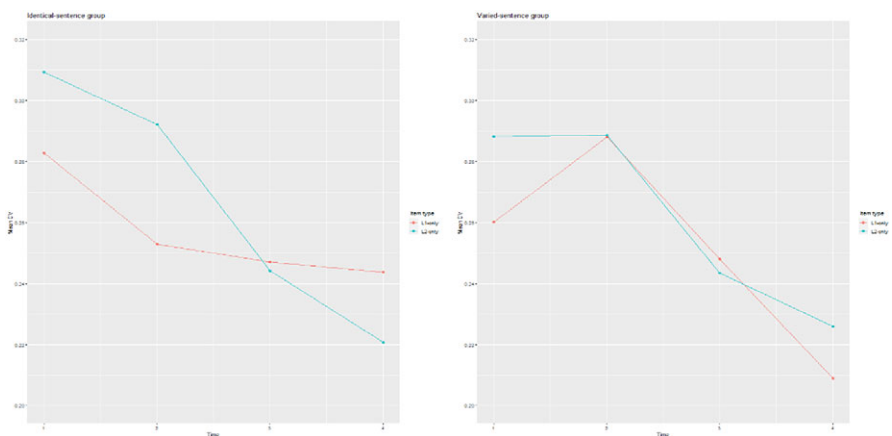


Figure 2. Mean CV for two item types across practice sessions for the identical-sentence group (left) and the varied-sentence group (right).

was a trend of larger CVs for the L2-only items in the earlier session (with significant CV difference only at Time 2, $p = .022$ in the identical-sentence group) while this pattern changed with negligible CV differences in the later sessions (CV differences at Times 3 and 4 were nonsignificant). This pattern of change resulted in a marginally significant interaction between Time and Item type.

To summarize the results from CV data, both groups' CVs for the learned items overall remained constant after the first practice session but declined considerably in the following sessions. Interestingly, CVs for the L2-only items tended to be larger in the earlier session but became similar to those of L1-only items in the later sessions.

Correlational analyses

The correlational analysis of RT and CV showed that overall positive and significant correlations between the two were found for both the L1-only and L2-only items in both groups across the tests (see Table 4). Overall, correlation coefficients between RT and CV for both items in both groups increased in all the subsequent posttests compared to that of pretraining test (Test 1) and remained high and significant.

Specifically, decreases in both RTs and CVs, along with the positive correlations between them, were found for the L1-only items at Time 4 and for the L2-only items at Times 3 and 4 in the identical-sentence group, while they were found for both item types at Times 3 and 4 in the varied-sentence group. This indicates that faster and more efficient processing of the learned items was achieved in both groups as a result of automatization in the second and/or the third session.

Discussion

The present study investigated the development of automaticity in processing collocations in three practice sessions and examined the influences of L1 collocational knowledge and two different types of practice conditions (practice in identical or varied contexts).

Development of automaticity in collocation processing

Regarding the first research question concerning the effects of repeated practice following declarative study on the development of automaticity in processing collocations, the results from the phrasal decision task showed that the treatments provided in this study were effective in improving collocation processing in terms of speed and automaticity. More specifically, the results of the RT and CV analyses showed that the participants improved their performance as a result of repeated practice following declarative study, becoming faster and less variable in processing the target expressions, as reflected in significant overall decreases in RT and CV. Moreover, significant positive correlations between RTs and CVs accompanied by decreases in RT and CV, observed for both groups at later sessions, indicate that both groups' improvement in collocation processing after the second and/or the third sessions was associated with automatization.

It is important to note that the learners in both groups started with almost zero knowledge of the target expressions at the beginning of the experiment but obtained significant gains in their speed of collocation processing, even after the first practice session (when learners had completed only four rounds of the fill-in-the-blank exercise

following declarative study), and kept performing significantly faster after every subsequent practice session. Thus, the repeated practice following declarative study was effective for promoting speed of lexical access of multiword items as well as single words as reported in previous studies (e.g., Akamatsu, 2008; Fukkink et al., 2005; Gatbonton & Segalowitz, 2005; Wood, 2009). More importantly, both groups also showed significant gains in automaticity in the later stage of learning (after the second and/or the third practice session). Note that overall RT and CV after the second and/or the third session significantly decreased from those of prior tests despite a relatively small amount of practice. Thus, the results seemed to indicate cumulative as well as immediate effects of the treatments on the development of automaticity of lexical access to the learned expressions. The results are consistent with findings from previous studies that reported robust effects of explicit vocabulary activities on developing automaticity of L2 lexical access (Akamatsu, 2008; Elgort, 2011; Fukkink et al., 2005; Pellicer-Sánchez, 2015). Using RT and CV measures, Pellicer-Sánchez (2015), for example, reported that various explicit vocabulary activities (including a gap-filling task) led to faster and more automatic processing of learned single words. The obtained effects remained robust even when retention was tested one month after the learning in her study.

The fact that CVs remained broadly constant during the initial stage of learning in the present study may seem inconsistent with the findings of Solovyeva and DeKeyser (2018) and Hui (2020), who reported a significant increase in CV during the initial stages of novel word learning as a result of the establishment of mental representations of the learned items. However, there are significant differences between the present study and the two previous ones. Both studies focused on the initial stage of novel word learning; learners were first tested on the target words with little or no prior knowledge in the pretraining test. However, in the present study, the declarative collocation study phase was given first, followed by the pretraining test. Thus, the participants already had sufficient declarative knowledge of the target items before taking the pretraining test. Moreover, the previous studies focused on single words unknown to the participants, while the present study focused on unknown multiword units that were made of known constituent words. However, the absence of noticeable CV changes in the early stages of learning (followed by its significant decrease in the later sessions) found in the present study seems to provide a potential explanation for the absence of an expected decrease in CV reported in some of the previous studies (e.g., Hulstijn et al., 2009).

Role of L1 collocational knowledge

Regarding the second research question concerning the role of L1 collocational knowledge in the development of automaticity in processing collocations, the RT results revealed a significant processing advantage for the L1-only items compared to the L2-only (incongruent) items in both groups at the early stages of learning. Thus, the existence of corresponding L1 collocations led to facilitation in processing speed. This result is in line with the robust facilitation effects found for processing congruent collocations over incongruent collocations (e.g., Sonbul & El-Dakhs 2020; Wolter & Gyllstad, 2011, 2013; Wolter & Yamashita, 2018; Yamashita & Jiang, 2010). The processing advantage observed in the L1-only items, however, disappeared with repeated practice. That is, performance with the L2-only items approached performance with the L1-only items in terms of processing time after the first or second practice session in both groups. Thus, the processing advantage offered by the L1

collocational knowledge at the early stages of learning appears to fade away if sufficient practice with the target expressions is provided. This finding is comparable with Sonbul and El-Dakhs (2020), who reported a gradual decrease in the congruency effect on timed recognition of collocations as the L2 learners' proficiency increased. The results also support the hypothesis from Yamashita and Jiang (2010) that at the initial stage of learning collocations, learners rely on the L1 mediation process, which results in the processing advantage of congruent collocations. At the initial stage, the recognition of a congruent collocation is much faster due to its link with its existing L1 counterpart, which is linked to the concept. The recognition of an incongruent collocation, however, is slower because the process involves more steps due to the lack of L1 equivalent expression, such as word-for-word L1 translation, then to a corresponding L1 collocation that shares the same meaning, and finally to the concepts. They further claimed that "with increasing input, L2 collocations, both congruent and incongruent, may become multiword units that are no longer dependent on the L1 lexicon and that are directly connected to concepts" (p. 653). In line with their prediction, Yamashita and Jiang (2010) also found that advanced ESL learners showed no difference in the processing time of recognizing congruent and incongruent collocations while lower-proficiency EFL learners showed a congruency effect. Therefore, in the present study, recognition of the target expressions during the early stages of learning involved the mediation and activation of the corresponding L1 expressions. However, increasing amounts of practice appear to have enabled learners to process both types of items (L1-only and L2-only collocations) with little involvement of L1 due to the establishment of a direct link between the target expression and its concept.

However, by examining learners' initial form-meaning mappings of collocations, Peters (2016) reported results inconsistent with the present findings as well as those in the previous studies. She found an absence of the congruency effect on learners' recognition performance in the immediate posttest following four explicit vocabulary activities (including a fill-in-the-blank exercise) but found the congruency effect in the productive recall test. Based on the findings, she suggested that congruency may not influence the learning burden of collocations in recognition but in production only. The results of her study, however, should be interpreted with care. First, as the author also pointed out, the target collocations included some collocations that were similar in form and meaning, which may have affected the results. Next, she used a matching test as a recognition test, in which participants were asked to match the individual constituents of the collocations. The test can be easily completed with the knowledge of the forms (compositions of collocations at the surface level); thus, the test may not necessarily involve retrieval of meaning.

In terms of improving automaticity, the present CV results indicated that the L2-only items tended to be processed with more variability during the early stages of learning. However, the learners came to process the L2-only items with a similar degree of variability as those for the L1-only items in the later stages of learning. The results may indicate that the recognition of L2-only collocations during the early stages of learning may involve not only slower but also more variable and effortful processing due to the lack of L1 equivalent expressions.

Interestingly, however, there was a hint of advantage for the L2-only items at the later stages of learning in terms of cumulative gains in the automaticity of collocation processing. When comparing the performance of the later stages (Times 3 and 4) to that of Time 1, greater evidence of automatization (as reflected in RT and CV) was found at an earlier stage (Time 3) for the L2-only items than that obtained for the L1-only items in both groups. As predicted by Yamashita and Jiang (2010), this finding may suggest

that to establish direct connections between L2 collocations and their corresponding concepts and for that link to become robust such that access to the concept operates fully automatically and autonomously, more practice may be required for collocations that have equivalent counterparts in L1 than incongruent collocations due to the former's strong link with existing L1 collocations. Incongruent collocations might have an advantage in this process due to the lack of direct translational L1 equivalents. However, with no statistically significant differences in CVs observed between the two item types in the later stages of learning, these predictions remain largely tentative and should be verified in future studies involving more practice over an extended period.

Role of practice condition

With regard to the third research question concerning the role of two different types of practice conditions (practice in identical or varied contexts), the results showed no significant group differences for any of the outcome measures (RT and CV) at any time point. The learners in both practice conditions showed significantly improved collocation processing in terms of speed and automaticity over time, to a comparable degree. This finding largely contrasts with the results of Durrant and Schmitt (2010), who reported the superiority of verbatim repetition over varied repetition in the initial stage of learning collocations. The current study, however, differs from the study carried out by Durrant and Schmitt (2010) in many aspects, which may have contributed to the different results. For example, they examined the development of procedural knowledge of the surface form of collocations (form recall of adjective-noun associations) under an incidental learning condition and used a naming test to assess learning. In this test, the adjective and the first two letters of the noun were provided as hints, which may have significantly primed the recall of the target pairs learners had encountered during the treatment especially considering the small number of the target items included (that were already assumed to be known by the participants). Moreover, the number of encounters (only twice) may not have been sufficient to result in development of any type of durable procedural knowledge. Another potential reason for the absence of a group effect in the current study may be that amount of practice was not sufficient. As the participants seemed to have only engaged in the relatively early stage of automatization due to the relatively short period of practice, more practice might have been needed for the two practice conditions to exert any differential effects.

At the same time, it is worth noting that there was a slight hint of an advantage for the identical-sentence condition over varied-sentence condition in the initial stage of learning. For example, learners in the identical-sentence condition significantly improved in their speed of collocation processing for both items even after completing the first session while the varied-sentence group did not make statistically significant gains for the L1-only items during the first session. Moreover, the analysis of gain scores (where gain is defined as each participant's RT and CV scores for the posttraining tests subtracted from those of the pretraining test) showed that greater gains for the L1-only items (although only marginally significant with the FDR correction) in both RT and CV (RT: $t(70) = 1.734, p = .087$; CV: $t(70) = 1.847, p = .069$) were made at Time 2 in the identical-sentence group compared to the varied-sentence group.¹ No indication of such an advantage for the identical-sentence condition was observed in the later stages

¹Note that the two groups did not statistically differ in their performance with both types of collocations in terms of processing speed and automaticity at Time 1.

of learning. Interestingly, however, when comparing the performance at Time 4 with that of Time 1 (the pretraining test), an indication of advantage for the varied-sentence condition over the identical-sentence condition was found in the later stage of learning. That is, greater evidence of automatization (as reflected in RT and CV) was observed in the varied-sentence group at Time 4 for both L1-only and L2-only items. This indication might be partly explained by the desirable difficulty framework (e.g., Bjork & Kroll, 2015; Brown et al., 2014; Suzuki et al., 2019). In the varied-sentence condition, participants had to be attentive to each new sentential context and process new language input in every round of practice because the sentential contexts were varied. Thus, this condition creates more challenging conditions, which may have required more active and deeper processing of the target items. However, identical sentential contexts in the other group provided more constrained and less challenging conditions, likely allowing easier retrieval of the target knowledge from memory during practice. Bjork and Bjork (2011) also suggested that varying the conditions of practice instead of keeping them constant and predictable can enhance learning. The results also seem to suggest that different stages of learning may play an important role in determining desirable difficulty that facilitates learning. It should be noted, however, that such tendencies should be interpreted with caution as the current study failed to find significant group differences at any time point.

Conclusion

The present study demonstrated that increasing practice with the target collocations following declarative study led to significantly faster and more automatic collocation processing. Both groups started with little to no declarative knowledge of the target collocations, but they acquired and retained a great degree of representational and functional aspects of collocation knowledge through the three practice sessions. The findings have important implications for second language pedagogy as the learners were able to develop automaticity of access to the representations of the newly learned collocations with a relatively small amount of simple practice following declarative collocation study, which can be easily provided in the classroom. As in the present study, it would be useful to take advantage of technologies to implement the exercises to enhance learning for fluency development.

The results also shed some light on the amount and conditions of practice that are necessary to start seeing significant gains in the speed and automaticity of collocation processing. Empirical evidence had thus far been lacking on this point, despite the need for more attention on fluency development in language instruction (Nation, 2007; Segalowitz, 2010). L1 collocational knowledge was found to play a facilitative role in the speed of lexical access in the early stages of learning. However, such facilitation effects disappeared with increasing practice with the target expressions. The findings highlight the differential role of L1 knowledge in different stages of learning collocations, as suggested by previous studies (Sonbul & El-Dakhs, 2020; Yamashita & Jang, 2010). As for the learning trajectory of collocations that have equivalent counterparts in L1 and that have no such counterparts, similar developmental stages of proceduralization and automatization took place for both types of collocations throughout the learning phase in both groups. However, the learners specifically improved in terms of processing stability in the later stages of learning as a result of automatization.

Some limitations and suggestions for future study should be noted. First, the lack of a control group in the current study does not allow for generalizations about the effect of the practice. As one of the reviewers pointed out, it might be possible that the learners in

the present study may have benefitted from repeating the phrasal decision task, for example. Although CV is reported not to be sensitive to practice effects due to repeated testing (Flehmig et al., 2007), further studies with a control group are necessary. Second, given that developing automaticity is a slow process that requires repeated practice over an extended period, future research with a greater amount of practice over a longer period is also necessary to corroborate the present findings. The effects of L1 collocational knowledge and practice conditions on developing automaticity in collocation processing might have been reliably observed if the participants had been given more practice over a longer period. At the same time, we stress that some of the accounts on the effects of L1 collocational knowledge and practice conditions on developing automaticity in collocation processing remain tentative and should be verified in future studies. Finally, in future studies, it would be useful to examine the development of automaticity in lexical processing by focusing on different types of collocations (e.g., adjective-noun collocations) with actual congruent expressions that exist in the target language as well as other types of multiword units. Comparing different types of linguistic knowledge (productive knowledge vs. receptive knowledge) would also provide a more comprehensive understanding of collocation learning. Considering the limitations noted above, further studies along these lines may yield greater generalizability of the findings.

Supplementary Materials. To view supplementary material for this article, please visit <http://doi.org/10.1017/S0272263122000547>.

Acknowledgement. This study was supported by the National Federation of Modern Language Teachers Association (NFMLTA) Dissertation Support Grant and the Ph.D. program in Second Language Acquisition at the University of Maryland.

Competing interests. We declare none.

References

- Abrahamsson, N., & Hyltenstam, K. (2009). Age of onset and natelikeness in a second language: Listener perception versus linguistic scrutiny. *Language Learning*, 59, 249–306.
- Akamatsu, N. (2008). The effects of training on automatization of word recognition in English as a foreign language. *Applied Psycholinguistics*, 29, 175–193.
- Anderson, J. R. (1992). Automaticity and the ACT theory. *The American Journal of Psychology*, 105, 165–180.
- Baayen, R. H. (2008). *Analyzing linguistic data: A practical introduction to statistics using R*. Cambridge University Press.
- Bahns, J. (1993). Lexical collocations: A contrastive view. *ELT Journal*, 47, 56–63.
- Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, 68, 255–278.
- Barton, K. (2016). MuMIn: Multi-model inference. R package version 1.40.4. Accessed at <https://CRAN.R-project.org/package=MuMIn>.
- Bates, D., Kliegl, R., Vasishth, S., & Baayen, H. (2015). Parsimonious mixed models. *arXiv preprint arXiv:1506.04967*.
- Bates, D., Maechler, M., Bolker, B., and Walker, S. (2015). lme4: Linear mixed effects models using Eigen and S4. R Package Version 1.1–9. <http://CRAN.Rproject.org/package=lme4>
- Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B (Methodological)*, 57, 289–300.
- Bjork, E. L., & Bjork, R. A. (2011). Making things hard on yourself, but in a good way: Creating desirable difficulties to enhance learning. *Psychology and the Real World: Essays Illustrating Fundamental Contributions to Society*, 2, 59–68.

- Bjork, R. A., & Kroll, J. F. (2015). Desirable difficulties in vocabulary learning. *The American Journal of Psychology*, 128, 241.
- Boers, F. (2014). A reappraisal of the 4/3/2 activity. *RELC Journal*, 45, 221–235.
- Boers, F., Dang, T. C. T., & Strong, B. (2017). Comparing the effectiveness of phrase-focused exercises: A partial replication of Boers, Demecheleer, Coxhead, and Webb (2014). *Language Teaching Research*, 21, 362–380.
- Boers, F., & Lindstromberg, S. (2012). Experimental and intervention studies on formulaic sequences in a second language. *Annual Review of Applied Linguistics*, 32, 83–110.
- Brown, P. C., Roediger, H. L. I., & McDaniel, M. A. (2014). *Make It Stick: The Science of Successful Learning*. Belknap/Harvard.
- Browne, C., Culligan, B., & Phillips, J. (2013). *The new general service list*. <http://www.newgeneralservicelist.org>
- Bulut, T., & Çelik-Yazici, İ. (2004). Idiom processing in L2: Through rose-colored glasses. *The Reading Matrix*, 4, 105–116.
- Carrol, G., & Conklin, K. (2014). Getting your wires crossed: Evidence for fast processing of L1 idioms in an L2. *Bilingualism: Language and Cognition*, 17, 784–797.
- Carrol, G., & Conklin, K. (2017). Cross language lexical priming extends to formulaic units: Evidence from eye-tracking suggests that this idea “has legs.” *Bilingualism: Language and Cognition*, 20, 299–317.
- Council of Europe. (2001). *Common European Framework of reference for languages: Learning, teaching, assessment*. Cambridge University Press.
- Cunnings, I., & Finlayson, I. (2015). Mixed effects modeling and longitudinal data analysis. In L. Plonsky (Ed.), *Advancing quantitative methods in second language research* (pp. 159–181). Routledge.
- De Jong, N., & Perfetti, C. A. (2011). Fluency training in the ESL classroom: An experimental study of fluency development and proceduralization. *Language Learning*, 61, 533–568.
- DeKeyser, R. (2015). Skill acquisition theory. In B. VanPatten & J. Williams (Eds.), *Theories in second language acquisition: An introduction* (pp. 94–112). Routledge.
- Durrant, P., & Schmitt, N. (2010). Adult learners’ retention of collocations from exposure. *Second Language Research*, 26, 163–188.
- Ehri, L. C. (1991). Development of the ability to read words. In R. Barr, M. Kamil, P. Mosenthal, & P. D. Pearson (Eds.), *Handbook of reading research* (Vol. 2, pp. 383–417). Longman.
- Elgort, I. (2011). Deliberate learning and vocabulary acquisition in a second language. *Language Learning*, 61, 367–413.
- Erman, B., & Warren, B. (2000). The idiom principle and the open choice principle. *Text Interdisciplinary Journal for the Study of Discourse*, 20, 29–62.
- Favreau, M., & Segalowitz, N. S. (1983). Automatic and controlled processes in the first-and second-language reading of fluent bilinguals. *Memory & Cognition*, 11, 565–574.
- Flehmig, H. C., Steinborn, M., Langner, R., Scholz, A., & Westhoff, K. (2007). Assessing intraindividual variability in sustained attention: Reliability, relation to speed and accuracy, and practice effects. *Psychology Science*, 49, 132–149.
- Forster, K. I., & Forster, J. C. (2003). DMDX: A Windows display program with millisecond accuracy. *Behavior Research Methods, Instruments, & Computers*, 35, 116–124.
- Fukink, R. G., Hulstijn, J., & Simis, A. (2005). Does training in second-language word recognition skills affect reading comprehension? An experimental study. *The Modern Language Journal*, 89, 54–75.
- Gatbonton, E., & Segalowitz, N. (2005). Rethinking communicative language teaching: A focus on access to fluency. *Canadian Modern Language Review*, 61, 325–353.
- Grabe, W. (2009). *Reading in a second language: Moving from theory to practice*. Cambridge University Press.
- Granena, G., & Long, M. H. (2013). Age of onset, length of residence, language aptitude, and ultimate L2 attainment in three linguistic domains. *Second Language Research*, 29, 311–343.
- Harrington, M. (2006). The lexical decision task as a measure of L2 lexical proficiency. *EUROSLA Yearbook*, 6, 147–168.
- Harrington, M., & Jiang, W. (2013). Focus on the forms. *Australian Review of Applied Linguistics*, 36, 132–145.
- Hui, B. (2020). Processing variability in intentional and incidental word learning: An extension of Solovyeva and DeKeyser (2018). *Studies in Second Language Acquisition*, 42, 327–357.

- Hulstijn, J. H. (2001). Intentional and incidental second language vocabulary learning: A reappraisal of elaboration, rehearsal and automaticity. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 258–287). Cambridge University Press.
- Hulstijn, J. H., Van Gelderen, A., & Schoonen, R. (2009). Automatization in second language acquisition: What does the coefficient of variation tell us? *Applied Psycholinguistics*, 30, 555–582.
- Ionin, T. (2003). Article semantics in second language acquisition (*Doctoral dissertation*). Massachusetts Institute of Technology, Cambridge, MA.
- Ionin, T., Ko, H. & Wexler, K. (2007). The role of semantic features in the acquisition of English articles by Russian and Korean speakers. In J. M. Liceras, H. Zobl, & H. Goodluck (Eds.), *The role of formal features in second language acquisition* (pp. 226–268). Routledge.
- Jiang, N. (2000). Lexical representation and development in a second language. *Applied Linguistics*, 21, 47–77.
- Jiang, N. (2004). Semantic transfer and its implications for vocabulary teaching in a second language. *The Modern Language Journal*, 88, 416–432.
- Kim, H. (2007). *Korean collocation dictionary for learners*. CommunicationBooks.
- Kim, J. W., Ritter, F. E., & Koubek, R. J. (2013). An integrated theory for improved skill acquisition and retention in the three stages of learning. *Theoretical Issues in Ergonomics Science*, 14, 22–37.
- Kim, K. M., & Godfroid, A. (2019). Should we listen or read? Modality effects in implicit and explicit knowledge. *The Modern Language Journal*, 103, 648–664.
- Kim, S., & Joo, G. (2016). *Learner's dictionary of Korean collocation*. Sotong.
- Kjellmer, G. (1994). *A dictionary of English collocations: Based on the Brown corpus*. Clarendon Press.
- Koda, K. (1996). L2 word recognition research: A critical review. *The Modern Language Journal*, 80, 450–460.
- Kremmel, B., Brunfaut, T., & Alderson, J. C. (2015). Exploring the role of phraseological knowledge in foreign language reading. *Applied Linguistics*, 38, 848–870.
- Kuznetsova, A., Brockhoff, P. B., Christensen, R. H. B., & Jensen, S. P. (2020). lmerTest: Tests in linear mixed effects models (R package version 3.1-3).
- Laufer, B. (1997). What's in a word that makes it hard or easy: Some intralexical factors that affect the learning of words. In N. Schmitt & M. McCarthy (Eds.), *Vocabulary: Description, acquisition and pedagogy* (pp. 140–155). Cambridge University Press.
- Lee Amuzie, G., & Spinner, P. (2012). Korean EFL learners' indefinite article use with four types of abstract nouns. *Applied Linguistics*, 34, 415–434.
- Lenth, R. (2020). emmeans: Estimated Marginal Means, aka Least-Squares Means. R package version 1.5.0.
- Li, M., & DeKeyser, R. (2019). Distribution of practice effects in the acquisition and retention of L2 Mandarin tonal word production. *The Modern Language Journal*, 103, 607–628.
- Lim, H., & Godfroid, A. (2015). Automatization in second language sentence processing: A partial, conceptual replication of Hulstijn, Van Gelderen, and Schoonen's 2009 study. *Applied Psycholinguistics*, 36, 1247–1282.
- Marian, V., Blumenfeld, H. K., & Kaushanskaya, M. (2007). The Language Experience and Proficiency Questionnaire (LEAP-Q): Assessing language profiles in bilinguals and multilinguals. *Journal of Speech, Language, and Hearing Research*, 50, 940–967.
- Martinez, R., & Murphy, V. A. (2011). Effect of frequency and idiomaticity on second language reading comprehension. *TESOL Quarterly*, 45, 267–290.
- Nation, P. (2007). The four strands. *International Journal of Innovation in Language Learning and Teaching*, 1, 2–13.
- Nesselhauf, N. (2003). The use of collocations by advanced learners of English and some implications for teaching. *Applied Linguistics*, 24, 223–242.
- Nesselhauf, N. (2005). Collocations in a learner corpus (Vol. 14). John Benjamins.
- O'Dell, F., & McCarthy, M. (2008). *English collocations in use: Advanced*. Cambridge University Press.
- Oxford University Press. (2009). *Oxford collocations dictionary: For students of English* (2nd ed.). Oxford University Press.
- Pellicer-Sánchez, A. (2015). Developing automaticity and speed of lexical access: The effects of incidental and explicit teaching approaches. *Journal of Spanish Language Teaching*, 2, 126–139.
- Perfetti, C. (2007). Reading ability: Lexical quality to comprehension. *Scientific Studies of Reading*, 11, 357–383.
- Peters, E. (2014). The effects of repetition and time of post-test administration on EFL learners' form recall of single words and collocations. *Language Teaching Research*, 18, 75–94.

- Peters, E. (2016). The learning burden of collocations: The role of interlexical and intralexical factors. *Language Teaching Research*, 20, 113–138.
- Phillips, N. A., Segalowitz, N., O'Brien, I., & Yamasaki, N. (2004). Semantic priming in a first and second language: Evidence from reaction time variability and event-related brain potentials. *Journal of Neuro-linguistics*, 17, 237–262.
- Pulido, M., & Dussias, P. (2020). Desirable difficulties while learning collocations in a second language: Conditions that induce L1 interference improve learning. *Bilingualism: Language and Cognition*, 23, 652–667.
- R Core Team. (2018). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. <http://www.R-project.org/>.
- Schmidt, R. A., & Bjork, R. A. (1992). New conceptualizations of practice: Common principles in three paradigms suggest new concepts for training. *Psychological Science*, 3, 207–218.
- Schmitt, N. (2008). Instructed second language vocabulary learning. *Language Teaching Research*, 12, 329–363.
- Schoonen, R., Gelderen, A. van, Gloppe, K. de, Hulstijn, J., Simis, A., Snellings, P., & Stevenson, M. (2003). First language and second language writing: The role of linguistic knowledge, speed of processing, and metacognitive knowledge. *Language Learning*, 53, 165–202.
- Segalowitz, N. S. (2000). Automaticity and attentional skill in fluent performance. In H. Riggenbach (Ed.), *Perspectives on fluency* (pp. 200–219). University of Michigan Press.
- Segalowitz, N. S. (2010). *Cognitive bases of second language fluency*. Routledge.
- Segalowitz, N. S., & Hulstijn, J. (2005). Automaticity in bilingualism and second language learning. In J. F. Kroll & A. M. B. de Groot (Eds.), *Handbook of bilingualism: Psycholinguistic approaches* (pp. 371–388). Oxford University Press.
- Segalowitz, N. S., & Segalowitz, S. J. (1993). Skilled performance, practice, and the differentiation of speed-up from automatization effects: Evidence from second language word recognition. *Applied Psycholinguistics*, 14, 369–385.
- Segalowitz, S. J., Segalowitz, N. S., & Wood, A. G. (1998). Assessing the development of automaticity in second language word recognition. *Applied Psycholinguistics*, 19, 53–67.
- Snellings, P., Van Gelderen, A., & De Gloppe, K. (2004). The effect of enhanced lexical retrieval on second language writing: A classroom experiment. *Applied Psycholinguistics*, 25, 175–200.
- Solovyeva, K., & DeKeyser, R. (2018). Response time variability signatures of novel word learning. *Studies in Second Language Acquisition*, 40, 225–239.
- Sonbul, S., & El-Dakhs, D. (2020). Timed versus untimed recognition of L2 collocations: Does estimated proficiency modulate congruency effects? *Applied Psycholinguistics*, 41, 1197–1222.
- Sonbul, S., & Schmitt, N. (2013). Explicit and implicit lexical knowledge: Acquisition of collocations under different input conditions. *Language Learning*, 63, 121–159.
- Spadaro, K. (2013). Maturation constraints on lexical acquisition in a second language. In G. Granena & M. Long (Eds.), *Sensitive periods, language aptitude, and ultimate L2 attainment* (pp. 43–68). John Benjamins Publishing Company.
- Sternberg, R. J., Powell, J. S., & Kaye, D. B. (1983). Teaching vocabulary-building skills: A contextual approach. In A. C. Wilkinson (Ed.), *Classroom computers and cognitive science* (pp. 122–143). Academic Press.
- Strong, B., & Boers, F. (2019). Weighing up exercises on phrasal verbs: Retrieval versus trial-and-error practices. *The Modern Language Journal*, 103, 562–579.
- Suzuki, Y. (2017). The Optimal Distribution of Practice for the Acquisition of L2 Morphology: A Conceptual Replication and Extension. *Language Learning*, 67, 512–545.
- Suzuki, Y., Nakata, T., & DeKeyser, R. (2019). The desirable difficulty framework as a theoretical foundation for optimizing and researching second language practice. *The Modern Language Journal*, 103, 713–720.
- Suzuki, Y., & Sunada, M. (2018). Automatization in second language sentence processing: Relationship between elicited imitation and maze tasks. *Bilingualism: Language and Cognition*, 21, 32–46.
- Suzuki, Y., & Sunada, M. (2020). Dynamic interplay between practice type and practice schedule in a second language: The potential and limits of skill transfer and practice schedule. *Studies in Second Language Acquisition*, 42, 169–197.
- Thai, C., & Boers, F. (2016). Repeating a monologue under increasing time pressure: Effects on fluency, complexity, and accuracy. *TESOL Quarterly*, 50, 369–393.

- Van Gelderen, A., Schoonen, R., De Glopper, K., Hulstijn, J., Simis, A., Snellings, P., & Stevenson, M. (2004). Linguistic knowledge, processing speed, and metacognitive knowledge in first-and second-language reading comprehension: A componential analysis. *Journal of Educational Psychology*, 96, 19–30.
- Webb, S., Newton, J., & Chang, A. (2013). Incidental learning of collocation. *Language Learning*, 63, 91–120.
- Wolter, B., & Gyllstad, H. (2011). Collocational links in the L2 mental lexicon and the influence of L1 intralexical knowledge. *Applied Linguistics*, 32, 430–449.
- Wolter, B., & Gyllstad, H. (2013). Frequency of input and L2 collocational processing. *Studies in Second Language Acquisition*, 35, 451–482.
- Wolter, B., & Yamashita, J. (2015). Processing collocations in a second language: A case of first language activation? *Applied Psycholinguistics*, 36, 1193–1221.
- Wolter, B., & Yamashita, J. (2018). Word frequency, collocational frequency, L1 congruency, and proficiency in L2 collocational processing: What accounts for L2 performance? *Studies in Second Language Acquisition*, 40, 395–416.
- Wood, D. (2009). Effects of focused instruction of formulaic sequences on fluent expression in second language narratives: A case study. *The Canadian Journal of Applied Linguistics*, 12, 39–57.
- Wray, A. (2002). *Formulaic language and the lexicon*. Cambridge University Press.
- Wray, A., & Perkins, M. R. (2000). The functions of formulaic language: An integrated model. *Language & Communication*, 20, 1–28.
- 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.

Cite this article: Jeong, H. and DeKeyser, R. (2023). Development of automaticity in processing L2 collocations: The roles of L1 collocational knowledge and practice condition. *Studies in Second Language Acquisition*, 45: 930–954. <https://doi.org/10.1017/S0272263122000547>