RESEARCH ARTICLE

The congruency effect in L2 collocational processing: The underlying mechanism and moderating factors

Yingzhao Chen

Michigan State University, East Lansing, MI, United States
Email: chenyi38@msu.edu

Abstract

The congruency effect—that is, faster and more accurate processing of congruent multiword units, has been demonstrated in multiple studies. It is still unclear, however, what its underlying mechanism is, and how congruency may interact with other factors. Using an acceptability judgement task, this study examined the congruency effect in immersive (Experiment 1) and nonimmersive (Experiment 2) L2 learners' collocational processing while taking into account L2 collocation frequency, immersive learners' L2 use, their length and starting age of immersion, nonimmersive learners' length of instruction, and their L2 proficiency. The study also tested whether L1 counterparts of words in L2 collocations were activated. Nonmmersive learners showed a congruency effect in both processing speed and accuracy. In contrast, immersive learners were affected by congruency only in processing accuracy. Higher L2 collocation frequency, greater length of instruction, and higher L2 proficiency did not reduce the congruency effect, whereas longer duration of immersion improved the processing of incongruent items. An effect of L1 lexical frequency was found, an indication of L1 activation. Results were discussed in light of how L2 proficiency and experiences changed the amount of L1 influence in L2 collocational processing.

Introduction

First language (L1) plays a significant role in second language (L2) learning and processing. Studies have demonstrated L1 influence in the processing of L2 multiword units (MWUs) through the congruency effect. MWUs, often used interchangeably with other terms such as formulaic sequences/language, may be defined as word sequences that frequently co-occur, that together often represent a single concept (Wood, 2019), and that are considered conventional by speakers of a language (Siyanova-Chanturia & Pellicer-Sánchez, 2019). Some main subcategories of MWUs include idioms (e.g., *spill the beans*), collocations (e.g., *single parents*), and phrasal verbs (e.g., *pick up*), among others. The congruency effect in MWU processing occurs when congruent MWUs— that is, those that have word-for-word translations between the L1 and the L2—are

processed faster and more accurately than incongruent ones, those without word-for-word L1 translations (e.g., Wolter & Gyllstad, 2013; Wolter & Yamashita, 2018; Yamashita & Jiang, 2010). An example of a congruent MWU between Chinese and English is *strong wind*, the direct Chinese translation of which is felicitous in Chinese. *Strong tea*, on the other hand, is incongruent, because its word-for-word Chinese translation does not constitute a felicitous expression in Chinese.

Little is known about the mechanism underlying the congruency effect, with some attributing the effect to L1 activation in L2 processing (e.g., Conklin & Carrol, 2018), and others to differences in the order of acquisition of congruent and incongruent MWUs (e.g., Wolter & Gyllstad, 2013). Few studies to date have directly tested these explanations. In addition, few studies on the congruency effect in MWU processing have considered how learner-related factors such as the learning context and item-related factors such as frequency (see Wolter & Gyllstad, 2013, for an exception) may moderate the congruency effect. This study intended to fill the above research gaps by taking into account L2 MWU frequency and by looking into MWU processing of learners in two different contexts—that is, English-as-a-foreign-language (EFL) and English-as-a-second-language (ESL). These two groups of learners mostly differ in their L2 experiences—for example, the amount and quality of exposure—and their comparison will shed light on how experiences shape L2 MWU representations and the bilingual lexicon. The study also directly tested whether L1 was activated in L2 MWU processing by examining whether L1 lexical frequencies of individual words in an L2 MWU affected learners' processing.

The underlying mechanism of the congruency effect in MWU processing

The congruency effect has been demonstrated across multiple subcategories of MWUs, including idioms (e.g., Carrol & Conklin, 2014, 2017), binomials (Du et al., 2021), and collocations (e.g., Sonbul & El-Dakhs, 2020; Wolter & Gyllstad, 2011, 2013; Wolter & Yamashita, 2018; Yamashita & Jiang, 2010). In general, data collected from eye-tracking and reaction-time (RT) tasks have suggested that an L2 MWU with a word-for-word translation in the L1 is processed faster and more accurately than an L2 MWU without such a counterpart in the L1. Conklin and Carrol (2018) attributed the congruency effect to the automatic activation of L1 translation equivalents of words in an L2 MWU. They argued that in the case of congruent MWUs, the L1 translation equivalents of individual words would then activate corresponding L1 MWUs, which leads to the activation of the underlying concepts; alternatively, the concepts can be directly activated by L2 MWUs without L1 mediation. For incongruent MWUs, on the other hand, there is no corresponding L1 MWUs to be activated, and L2 learners have to access the meaning of L2 MWUs directly without relying on the L1. They explained that in the initial stages of L2 learning, L2 MWU representations are likely to be weak and thus the L1-mediated route may be faster, resulting in an advantage of congruent MWUs. Conklin and Carrol's explanation is in line with the revised hierarchical model (Kroll & Stewart, 1994), which hypothesized that access to L2 word meanings is mediated through the L1 at least at the beginning of L2 learning. Similarly, Jiang (2000) proposed a bilingual lexicon model where L2 forms are linked to L1 semantics in the early stages of L2 lexical development. Applying Jiang's model to L2 MWU processing, congruent MWUs would be recognized faster because they are linked to existing L1 MWUs; the recognition of incongruent MWUs, in contrast, would be slower, because it requires rejecting the word-for-word L1 translations first and

matching L2 MWUs to the right L1 expressions that bear the same meanings (Yamashita & Jiang, 2010).

A few studies interpreted their findings as counterevidence against L1 activation as the mechanism underlying the congruency effect. Wolter and Yamashita (2018), using an acceptability judgement task (AJT), included three types of adjective-noun collocations: congruent, incongruent, and translated collocations (those that exist only in the L1, i.e., L1-only). Their results revealed that although Japanese learners of English showed a congruency effect, they did not show a processing difference between L1-only collocations and noncollocational items. Similarly, Wolter and Yamashita (2015) used a double lexical decision task and found no difference between the processing of L1-only collocations and noncollocates. The existence of the congruency effect and the lack of processing difference between L1-only collocations and noncollocates were used as evidence against the argument that L1 activation underlies the congruency effect. Further, Wolter and Gyllstad (2013), focusing on adjective-noun collocations, did not find an effect of L1 collocation frequency on L2 collocational processing, an indication that L1 collocations were not activated. Authors of these studies referred to the order of acquisition as an alternative explanation for the congruency effect. They argued that congruent MWUs are learned earlier than incongruent ones based on available L1 MWU knowledge. MWUs that are learned earlier then become more entrenched in the neural networks and thus enjoy a processing advantage.

However, it is also possible that the lack of difference between L1-only and noncollocational items in these studies did not come from the lack of L1 activation but was in fact due to the nature of the task. In Wolter and Yamashita (2015, 2018), participants had to make a Yes/No decision on the acceptability of a collocation as in an AJT or on the lexicality of the component words in a collocation as in a lexical decision task. It is possible that participants activated the L1 translations of individual words in an L1-only collocation and initially decided for a Yes based on their L1. But later participants were unable to find the L1-only collocation in their L2 mental lexicon to support their decision. They then had to revise their initial Yes decision before eventually pressing the response key, resulting in a slowdown and thus a null difference between L1-only and noncollocates. This potential explanation is supported by eye-tracking data from Carrol and Conklin (2017) and Carrol et al. (2016). In these studies, when participants did not have to make decisions, they read L1-only idioms faster than control items.

Regarding the null results of L1 collocation frequency effect in Wolter and Gyllstad (2013), it could well be the case that the L1 counterparts of individual words in a MWU rather than the whole L1 MWU are activated. In this case, L1 frequencies of individual words rather than of MWUs will have an effect on L2 processing. In the current study, I explored this possibility by examining whether L2 learners' processing of MWUs is influenced by the L1 translation frequencies of individual words in the MWUs. This approach is less likely to be subject to task effects than the use of L1-only MWUs and provides more direct evidence of L1 activation (Jiang et al., 2020).

Factors moderating the congruency effect

This study focused on the item-related variable of L2 MWU frequency and the learner-related variable of learning context and investigated how they interacted with congruency. These two variables were chosen to represent two aspects of language learning experiences. In particular, MWU frequency indexes the more specific micro-L2

experiences that learners have with individual MWUs, and the learning context represents the macro aspect of one's L2 experiences—that is, being immersed in the L2 environment or not. Research into how these two factors moderate the congruency effect may tell a more complete story about how experiences influence the bilingual lexicon.

L2 frequency of multiword units

Frequency represents the prevalence of a linguistic unit in the input and indicates the amount of experience one has with the unit (Tremblay & Tucker, 2011). The more frequent the unit, the more likely one encounters it repeatedly. Repeated encounter of a structure makes its representation stronger and more readily accessible in memory; and particularly related to MWUs, high frequency also leads to structure autonomy—that is, being processed as a unit rather than as individual parts (Bybee, 1985, 2006). Empirical research has suggested that more experiences with a MWU, or higher MWU frequency, lead to better learning (e.g., Pellicer-Sánchez et al., 2022; Puimège & Peters, 2020), faster processing (e.g., Ellis et al., 2008; Yi, 2018), and faster production (e.g., Janssen & Barber, 2012)

   Although studies on L2 MWU processing and learning largely confirmed the effects of L2 MWU frequency and congruency, not much is known about how congruency interacts with frequency. Yamashita and Jiang (2010) hypothesized that congruent MWUs may be less influenced by L2 MWU frequency because they can be accepted based on learners' L1, whereas for incongruent MWUs, L2 frequency may play a bigger role, as they receive little L1 support. Another possibility is that the congruency effect may attenuate among MWUs of high L2 frequency because as these MWUs are encountered repeatedly, their L2 representations become stronger and are subject to less L1 influence. So far, only Wolter and Gyllstad (2013) have examined this potential interaction. Counter to Yamashita and Jiang's predictions, their results revealed that the congruency effect applied to both high- and low-frequency collocations, and that both congruent and incongruent collocations were subject to L2 frequency influence. Further investigation is needed to understand in what circumstances this interaction may or may not occur. Several studies have suggested that L2 learners' sensitivity to L2 frequency may vary based on learners' amount of L2 exposure: more L2 exposure is associated with reduced sensitivity to L2 frequency (Cop et al., 2015; Yi, 2018; Yi et al., 2017). If learners with different amounts of L2 experiences vary in their sensitivity to L2 MWU frequency, it is possible that these learners' reaction to congruency is also influenced by L2 MWU frequency to different extents.

Learning context

The previously mentioned bilingual lexicon models by Kroll and Stewart (1994) and Jiang (2000), in addition to hypothesizing how the L1 and the L2 interact, also provided predictions on how the L1-L2 interaction develops as the L2 learner progresses. Kroll and Stewart (1994) argued that although access to L2 word meanings is mediated by the L1 at the beginning of L2 learning, the connections between L2 words and their meanings will grow stronger and thus reliance on the L1 would decrease when (1) the L2 becomes dominant possibly when learners are in an immersive environment and/or (2) L2 proficiency improves. From a developmental perspective, Jiang (2000) also argued that as learners' L2 exposure and proficiency

increase, words in their L2 mental lexicon might connect to concepts directly without L1 mediation.

Findings regarding whether an immersive environment reduces L1 influence in L2 learning and processing in general have been mixed so far. In Linck et al. (2009), English learners of Spanish, after three months of studying abroad, showed attenuated L1 access in comprehension and production compared with students who learned the L2 in a classroom setting in their home country. Jiang et al. (2020) included Chinese learners of English who resided in the US and China. The study revealed that when processing L2 words, only learners in China but not those immersed in the United States exhibited L1 activation. Several studies, in contrast, demonstrated that learners who were immersed in the L2 environment still activated their L1 during L2 processing. Carrol and Conklin (2017), using eye tracking, showed activation of L1 forms but not meanings in L2 idiom processing by Chinese learners of English in the UK. In Thierry and Wu (2007) and Wu and Thierry (2010), Chinese learners of English in the UK activated L1 translations in the L2 processing of single words as shown by ERP data; however, such L1 activation was not detected in behavior RT data.

With respect to L2 MWU processing, most studies recruited a homogenous group of L2 learners who were residing in their home country or in the target language country. These studies were therefore not positioned to explore the potential moderating effects of learning context on the congruency effect. One exception, Yamashita and Jiang (2010), compared EFL and ESL learners, who resided in China and the US, respectively, in their processing of adjective-noun and verb-noun collocations. The authors in fact found that only EFL but not ESL learners showed a congruency effect in processing speed, which was interpreted as an indication of ESL learners' L2 collocation representations being independent of the L1. However, it is not known what aspects—for example the length of residence, age of arrival, or the amount of L2 use—in the immersive environment contribute to the development of L2 MWU representations. Although these variables have been found to affect language learning to different extents in an immersive context (length of residence: e.g., Amuzie & Winke, 2009; age of arrival: e.g., Baker, 2010; L2 use: e.g., Baker-Smemoe et al., 2014; Cubillos & Ilvento, 2012), they have not been studied in L2 MWU processing. For L2 learners residing in their home countries, it is unclear whether more instruction and/or higher proficiency levels will allow them to eventually develop strong L2 MWU representations like their counterparts in the immersive environment. For the variable of length of instruction, Muñoz (2014) suggested that in terms of oral proficiency, nonmmersive learners' syntactic complexity but not fluency or lexical diversity was predicted by the number of years of formal instruction. Again, this variable has not been explored in MWU processing studies. Regarding L2 proficiency, three studies on L2 collocational processing so far (Sonbul & El-Dakhs, 2020; Wolter & Gyllstad, 2013; Wolter & Yamashita, 2018) have examined its effect and showed mixed findings. In Sonbul and El-Dakhs (2020), participants reacted to adjective-noun and verb-noun collocations in a timed AJT task. The authors found an interaction between L2 proficiency and congruency, suggesting that the processing difference in RT between congruent and incongruent collocations was smaller for more advanced learners. This interaction, however, was not found in the untimed multiple-choice test. Wolter and Gyllstad (2013), also using an AJT, uncovered an L2 proficiency effect in the accuracy but not RT data: the effect of L2 proficiency was more pronounced on the processing accuracy of incongruent than congruent collocations. Wolter and Yamashita (2018), on the other hand, did not reveal an L2 proficiency effect: learners displayed a congruency effect regardless of their proficiency. As Sonbul and El-Dakhs (2020) pointed out, the reason

why Wolter and Yamashita (2018) did not find an effect of L2 proficiency may be because they included proficiency as a categorical rather than a continuous variable; further, the proficiency range included in previous studies was somewhat narrow. Although Sonbul and El-Dakhs (2020) filled the abovementioned two gaps, their proficiency measure only assessed 1k and 2k levels of vocabulary, which may not have fully revealed participants' proficiency. The inconsistent findings regarding the effect of L2 proficiency on congruency warrant further research with more comprehensive proficiency measures.

The current study

The goal of the current study was to test whether L1 activation of individual words in a MWU is underlying the congruency effect and how the strength of the congruency effect may vary in relation to learners' learning context and L2 MWU frequency. Two experiments are reported. Experiment 1 involved ESL learners residing in the US. Experiment 2 was intended to see whether the results of ESL learners also apply to EFL learners in China. ESL learners' length of residence and age of arrival in the target L2 country, the amount of L2 use, and EFL learners' L2 proficiency and length of instruction were taken into account to further gauge the effects of learning context and language experiences.

The current study focused on the processing of adjective-noun collocations, a subcategory of MWUs. Collocations are more likely to be subject to L1 influence because unlike other types of MWUs such as idioms, an L1 collocation usually has a corresponding expression in the L2, though sometimes with different word combinations (Wolter, 2006; Yamashita & Jiang, 2010). Adjective-noun collocations were chosen in this study because the lack of variability in whether a determiner is present before the noun in this type of collocations makes it easier to develop comparable items used in the experiments (Wolter & Gyllstad, 2013; Yi, 2018). Because collocation frequency was included as a variable in the study, the frequency-based approach, which defines collocations as words that frequently co-occur (e.g., Gyllstad & Wolter, 2016; Nesselhauf, 2003; Sinclair, 1991; Webb et al., 2013), was followed when choosing experimental items. The following research questions were addressed:

1. Does the congruency effect show in the L2 collocational processing of ESL (Experiment 1) and EFL (Experiment 2) learners?
2. Are the L1 translation equivalents of individual words in an L2 collocation activated during ESL (Experiment 1) and EFL (Experiment 2) learners' L2 processing?
3. To what extent does L2 collocation frequency (Experiments 1 and 2), ESL learners' length of residence, age of arrival, and L2 use (Experiment 1), and EFL learners' L2 proficiency and length of instruction (Experiment 2) moderate the congruency effect?

Materials

Using the Phrases in English database (Fletcher, 2011), I retrieved a preliminary list of 66,285 adjective-noun collocations with the minimum frequency of 10 from the British National Corpus (BNC). The raw frequencies were then standardized using the Zipf scale (Van Heuven et al., 2014). The Zipf scale is logarithmic and does not contain negative values. Zipf frequency in this study was calculated by the formula log10

(frequency per million words) + 3. The conversion resulted in a Zipf-frequency range of 2 to 4.44. Following Siyanova-Chanturia and Spina (2015) and Yi (2018), I divided the frequency range into four bins: 2–2.3, 2.5–2.8, 2.9–3.2, 3.3–4.44.

Next, I went through items in each bin to determine which were congruent and incongruent. A congruent collocation has a word-for-word translation from English to Chinese. That is, the most typical Chinese translations of the English collocation's components constitute an acceptable Chinese word or collocation that expresses the same meaning as the English collocation. For example, *good* and *idea* in *good idea* are typically translated into Chinese as 好 (*hǎo*) and 主意 (*zhǔ yì*), respectively; 好主意 is a legitimate Chinese word and corresponds to the meaning of *good idea.* An incongruent collocation is one without such word-for-word translation between the two languages. An example is *heavy rain*, where *heavy* and *rain* are typically translated into 重 (*zhòng*) and 雨 (*yǔ*), respectively. The combination of the two characters—that is, 重雨, is infelicitous in Chinese and does not correspond to the meaning of *heavy rain*, which is expressed in Chinese as 大 (*dà*; *big*) 雨 (*yǔ*; *rain*).

Pilot tests following Yamashita and Jiang's (2010) procedure were run to make sure that the typical translations of the collocations and their constituents I assumed were also agreed upon by other L1 speakers of Chinese. Chinese learners of English who had a background similar to those of prospective ESL participants in the study were asked to provide Chinese translations for (1) the constituent words in the English collocations in the first pilot (*n* = 4) and (2) the English collocations in the second pilot (*n* = 5). Collocations in the study fulfilled the following criteria, as shown by responses from at least three out of four participants in the first pilot and four out of five in the second pilot. First, the meanings of the collocations and their components should be known. Second, for congruent items, the translations of words in the first pilot should be equal to individual Chinese words in the Chinese translations of collocations in the second pilot. For example, *white* and *paper* were translated into 白 (*bái*) *and* 纸 (*zhǐ*), respectively, in the first pilot, each corresponding to the words in the translation of the congruent collocation *white paper* (白纸). Third, for incongruent items, the translations of words in the first pilot should not completely overlap with Chinese words in the Chinese translations of collocations in the second pilot. *Living rooms*, for example, was translated to 客 (*kè*; *guest*) 厅 (*tīng*; *hall*) in the second pilot; neither 客 nor 厅 corresponded to the translations provided for *living* (生活 *shēng huó*) and *rooms* (房间 *fáng jiān)* in the first pilot.

The congruent and incongruent items were then matched for length (number of letters) of the whole collocation as well as its individual words, mutual information (MI; the strength of co-occurrence between words in a collocation), L2 collocation frequency, L2 Word1 frequency, and L2 Word2 frequency, using data from BNC. The final materials included 40 congruent and 40 incongruent collocations, with 10 congruent and 10 incongruent items in each of the four frequency bins; 99% of the words used in the stimuli were from the most frequent 4,000 word families in the BNC/COCA word frequency list (Nation, 2012). Words within this frequency band were likely to be known by target EFL participants in China, who would have passed the College English Test Band 4 (CET4), a standardized proficiency test in China (see Zhao & Ji, 2018, for the relationship between vocabulary size and CET4 scores). In addition, five EFL students in China, who were similar to prospective EFL participants in the study, rated the items for familiarity—that is, how well the participants knew the meanings of the items, on a 7-point Likert-type scale. The average familiarity rating was 6.70 (*SD* = .28).

Eighty noncollocates, serving as control items, were created by randomly combining adjectives and nouns from the congruent and incongruent collocations. The control

Table 1. Summary of item characteristics

	Congruent M (SD)	Incongruent M (SD)	Control M (SD)
Item length	10.82 (2.31)	10.72 (2.44)	10.76 (2.21)
Word1 length	5.30 (1.26)	5.10 (1.57)	5.20 (1.42)
Word2 length	5.53 (1.57)	5.62 (2.01)	5.56 (1.81)
L2 collocation frequency	2.89 (.58)	2.90 (.61)	NA
L2 Word1 frequency	5.10 (.43)	4.97 (.52)	5.04 (.48)
L2 Word2 frequency	5.11 (.39)	4.93 (.49)	5.02 (.45)
MI	5.56 (1.77)	6.65 (2.74)	NA
L1 Word1 frequency	4.93 (.75)	4.79 (.81)	4.85 (.78)
L1 Word2 frequency	4.91 (.58)	4.83 (.71)	4.88 (.66)

Note. All frequencies were standardized to Zipf values. Word frequencies were word form frequencies—that is, the number of times a word is counted in corpus. Length referred to the number of letters.

items did not appear in BNC, except for *black meal*, which has a frequency of 1 per 100 million and a low MI of .15. The L1 (Chinese) translation frequencies of Word1 and Word2 were obtained from SUBTLEX-CH (Cai & Brysbaert, 2010). Table 1 presents a summary of item characteristics. Mann–Whitney U tests revealed no significant difference between the congruent and incongruent collocations in terms of L2 collocation frequency ($W = 794.50$, $p = 1.00$, $r = -.006$) and MI ($W = 953.00$, $p = .14$, $r = .16$). No significant difference was found between congruent, incongruent, and control items as suggested by Kruskal–Wallis tests for L2 Word1 frequency, $\chi^2 (2) = .64$, $p = .73$, effect size in $\epsilon^2 = .004$; L2 Word2 frequency, $\chi^2 (2) = .4.43$, $p = .11$, effect size in $\epsilon^2 = .028$; Word1 length, $\chi^2 (2) = .08$, $p = .96$, effect size in $\epsilon^2 = .001$; Word2 length, $\chi^2 (2) = .01$, $p = 1.00$, effect size in $\epsilon^2 = 4.89e-05$; and item length, $\chi^2 (2) = .07$, $p = .96$, effect size in $\epsilon^2 = .0005$. Both Experiments 1 and 2 used the same set of items. For the complete list of stimuli, see Appendix A. Twenty-four L1 speakers of English were recruited to respond to the stimuli in an AJT. This manipulation check confirmed that the stimuli were working as intended: L1 speakers of English (1) responded significantly faster and more accurately to collocations than to control items, (2) did not show a congruency effect, and (3) did not show effects of L1 Chinese Word1 or Word2 frequencies. For L1 English speakers' RT and accuracy and details of the analyses, see Appendix B.

The acceptability judgement task

An AJT was used in Experiments 1 and 2 to examine collocational processing. This task taps into participants' processing of meaning (Wolter & Yamashita, 2018). In this task, participants were asked to decide whether or not a word combination was acceptable as quickly and accurately as possible (see Appendix C for task instructions). The instructions were written in both English and Chinese. The task was created and administered online using Gorilla, an online experiment builder (Anwyl-Irvine et al., 2019).

The items in the AJT were presented one at a time in a randomized order at the center of a screen. Each trial began with a fixation cross for 500 ms, followed by the item, which remained on the screen until response or disappeared and was replaced by the next trial after 5,000 ms. Before the presentation of the 160 test items, there was a handedness questionnaire and 20 practice items. Left-handers pressed *s* for YES responses and *k* for NO responses. For those who were right-handed, YES responses were indicated by pressing *k* and NO responses by pressing *s*. Participants took a break after 90 items and resumed by pressing the space bar.

Procedure

In both Experiments 1 and 2, participants first completed a language background questionnaire on Qualtrics (https://www.qualtrics.com) before doing the AJT on Gorilla. It was emphasized to the participants that they should do the task (1) on a desktop or laptop, (2) in a quiet environment with good internet connection, and (3) after reading the instructions carefully. For the first requirement, a constraint was also set on Gorilla such that the AJT was only accessible on desktops or laptops. The duration of each experiment was around 20 min.

Experiment 1

Participants

Thirty-three ESL learners in the US participated in the study. Two of them, who arrived in the US at the age of 6 and 11 years, respectively, and were studying at high school were excluded to make the ESL group more homogenous. The remaining ESL learners all spoke Mandarin as their L1, were studying at university in the US or had graduated from a US university, and were residing in the US at the time of participation. Their mean age was 25.52 years (SD = 5.38, range = 20–42). They had been in the US for at least a year (M = 4.82, SD = 2.13, range = 1.5–12) and came to the US at a mean age of 20.84 years (SD = 5.47, range = 13–36). On a 100-point scale in the language background questionnaire, the ESL learners reported using English 38.10% of their day on average (SD = 22.10, range = 0–81). The proficiency level of the ESL learners was estimated to be at least intermediate-low, given that a TOEFL score of 60 or IETLS of 6.0 is at the lower end of the admission scores required by US universities.

Analysis

All statistical analyses were performed in R (version 3.6.1; R Core Team, 2019). The RT and accuracy data collected from the AJT were analyzed with mixed-effects modeling. Mixed models were built using the lme4 package (version 1.1-21; Bates et al., 2015). The p values were calculated by lmerTest (version 3.1-0; Kuznetsova et al., 2017). Responses with RTs shorter than 400 ms or longer than 4,000 ms were trimmed using the trimr package (version 1.0.1; Grange, 2015). These responses were considered too slow or fast to accurately reflect the genuine recognition process (see also Gyllstad & Wolter, 2016; Öksüz et al., 2021; Wolter & Yamashita, 2018, for similar outlier removal criteria). RT trimming affected 0.52% of the ESL data. Trimmed RT data were then log transformed to bring the variable closer to normal distribution.

Log-transformed RTs were analyzed with linear mixed-effects models and only correct responses were included. Generalized linear mixed-effects models were built for accuracy data in which correct responses were coded as 1 and incorrect ones as 0. A maximal model was first built, which included (1) main effects of theoretical interest: L2 collocation frequency, condition (congruent and incongruent), L1 Word1 and Word2 frequencies, age of arrival, length of residence, and L2 use; (2) interactions between the main effects of interest; (3) covariates: collocation length (number of letters), L2 Word1 and Word2 frequencies, and MI; and (4) maximal random-effects structure justified by the data (Barr et al., 2013). Condition was treatment coded with the congruent condition serving as the reference. Three items (one congruent and two incongruent) were excluded from the analysis because the L1 frequencies of these items were not

found in the corpus. In case of nonconvergence, the random-effects structure was simplified by dropping by-subject random slopes first (Barr et al., 2013), followed by by-item random slopes. The backward stepwise modeling procedure was adopted in which insignificant interactions were first excluded, followed by insignificant covariates. Interactions and covariates that did not improve model fit were then also removed. Models were compared based on the Akaike Information Criterion (AIC), and the model with the lowest AIC was chosen. All continuous independent variables were mean centered to reduce collinearity. The variance inflation factors of the final models were checked using the performance package (version 0.4.0; Lüdecke et al., 2019) and were all under 10 (Hair et al., 1995). Residuals of the final models were also examined, and observations with a residual greater than 2.5 SD away from the mean were removed. The 95% Wald confidence intervals for estimates were obtained through the confint.merMod() function.

Results and Discussion

Table 2 includes the descriptive statistics of ESL learners' trimmed RTs (in milliseconds) and accuracy (in percentage) by item type. In response to RQ1 on the congruency effect, the RT analysis (Table 3) showed that overall, ESL learners responded to congruent and incongruent collocations at a similar speed—that is, no congruency effect; in contrast, the accuracy analysis (Table 4) showed that that ESL learners made more errors when responding to incongruent than congruent items. In terms of RQ2 on L1 activation, ESL learners' processing speed but not accuracy was affected by the L1 frequencies of component words in the L2 collocations. Further, the RT data showed no significant

Table 2. Descriptive statistics for ESL learners' RT and accuracy

Item type	RT (SD)	Accuracy
Congruent	1,302.21 (539.75)	88.29
Incongruent	1,299.75 (507.13)	66.99
Control	1,501.32 (563.30)	77.89

Table 3. Mixed-effects model for ESL learners' RT

	Fixed effects				Random effects			
					By participant		By item	
	Estimate [95% CI]	SE	t	p	Variance	SD	Variance	SD
Intercept	7.11 [7.03, 7.17]	.03	207.85	<.001***	.02	.15	.02	.12
L2 collocation frequency	−.11 [−.17, −.06]	.03	−4.31	<.001***				
Incongruent	.04 [−.02, .10]	.03	1.11	.27				
L1 Word1 frequency	−.05 [−.09, −.01]	.02	−2.46	.02*				
L1 Word2 frequency	−.06 [−.11. −.01]	.02	−2.54	.01*				
Age of arrival	.003 [−.01, .01]	.005	.55	.59				
Length of residence	.008 [−.02, .04]	.01	.57	.57				
Percentage of L2 use	−.002 [−.004, .001]	.001	−1.18	.25				
Incongruent: Length of residence	−.02 [−.03. −.003]	.006	−2.49	.01*				
Length	.03 [.01, .04]	.007	3.84	<.001***				

*p < 0.5; **p < .01; ***p < .001

Table 4. Mixed-effects model for ESL learners' accuracy

| | Fixed effects | | | | Random effects | | | |
| | | | | | By participant | | By item | |
	Estimate [95% CI]	SE	z	p	Variance	SD	Variance	SD
Intercept	3.11 [2.49, 3.73]	.32	9.82	<.001***	.49	.70	1.99	1.41
L2 collocation frequency	1.65 [.99, 2.31]	.34	4.90	<.001***				
Incongruent	−1.95 [−2.69, −1.21]	.38	−5.20	<.001***				
L1 Word1 frequency	.44 [−.04, .92]	.24	1.81	.07				
L1 Word2 frequency	.32 [−.24, .87]	.28	1.12	.33				
Age of arrival	.01 [−.04, .06]	.03	.34	.74				
Length of residence	−.08 [−.25, .08]	.08	−.96	.34				
Percentage of L2 use	.01 [−.01, .02]	.01	1.11	.27				
Incongruent: Length of residence	.15 [.03, .27]	.06	2.54	.01*				

*p < 0.5; **p < .01; ***p < .001

interaction between condition (congruent vs. incongruent) and L1 word frequencies, suggesting that L1 was activated in the processing of both congruent and incongruent collocations. As for the moderating variables on the congruency effect (RQ3), the interaction between L2 collocation frequency and congruency did not survive model selection, suggesting that ESL learners were sensitive to frequency regardless of congruency and that, more importantly, the congruency effect in accuracy or the lack of it in RT remained regardless of collocation frequency. ESL learners' reaction to congruency in L2 collocational processing was affected by their length of residence in the US. Longer residence in the US was associated with faster and more accurate processing of incongruent items (see Figure 1). Age of arrival in the country or percentage of daily L2 use did not moderate the congruency effect. The lack of age effect may be attributed to the fact that most ESL participants arrived in the US after age 18 (DeKeyser, 2000; DeKeyser et al., 2010). For the null effect of L2 use, one explanation could be that percentage of L2 use did
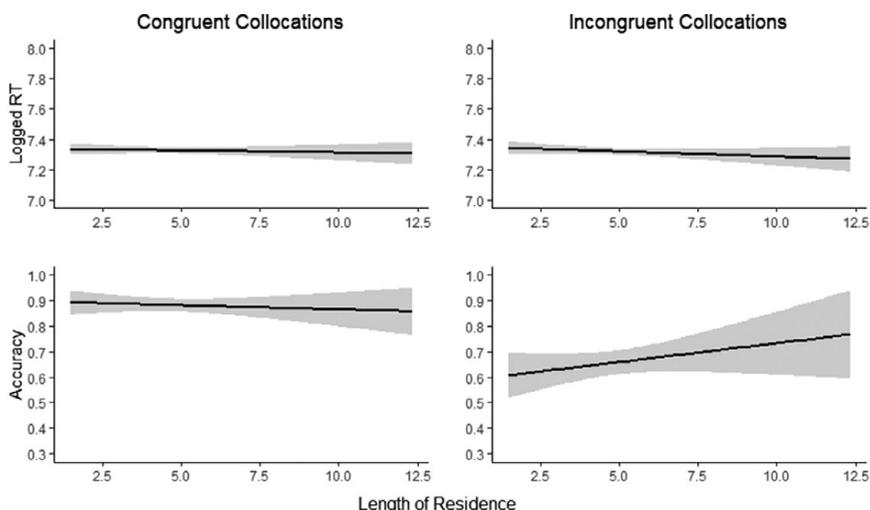

Figure 1. Interactions in ESL learners' RT and accuracy.

not reflect the quality of L2 use, which may contribute to language development more than quantity (Baker-Smemoe et al., 2014).

Experiment 2

Participants

Thirty-three EFL learners were recruited. They were non-English majors in their sophomore or junior years in college in mainland China, with an average age of 19.52 years (SD = 0.94, range = 18–22). They had never lived or studied in an English-speaking country and had received English instruction for an average of 10.39 years (SD = 2.36, range = 6–15). Self-reported CET4 scores were used as a proxy of EFL learners' proficiency level. All but four participants had taken the test, and those who did took the test less than 2 years ago from the time of data collection. The CET4 has a maximum score of 710, and the average score of the EFL participants was 494.97 (SD = 57.59, range = 372–593). According to China's National Education Examinations Authority (n.d.), the participants' mean score was around the 44th percentile, with a range from the 4th to 90th percentile. The EFL learners' CET scores suggested that they were of intermediate-low to intermediate proficiency.

Analysis

Mixed-effects models were built to analyze RT and accuracy data. The RT trimming and transformation, data coding, modeling procedure, model comparison, and model criticism followed those in Experiment 1. RT trimming affected 2.50% of the EFL data. The initial maximal models for RT and accuracy included (1) main effects of theoretical interest: L2 collocation frequency, condition (congruent and incongruent), L1 Word1 and Word2 frequencies, L2 proficiency (CET4 scores), and length of instruction; (2) interactions between the main effects of interest; (3) covariates: collocation length (number of letters), L2 Word1 and Word2 frequencies, and MI; (4) maximal random-effects structure justified by the data (Barr et al., 2013). Four EFL participants were excluded from analysis due to missing CET4 scores.

Results and discussion

Table 5 presents the descriptive statistics for EFL learners' RT and accuracy. Mixed-effects models (see Tables 6 and 7) showed that unlike ESL learners, EFL learners showed a congruency effect in both processing speed and accuracy (RQ1), reacting significantly faster and more accurately to congruent than incongruent items. Similar to ESL learners, EFL learners were affected by the L1 frequencies of component words only in RT but not in accuracy (RQ2). L1 word frequencies did not interact with condition, an indication of L1 activation in both congruent and incongruent collocation

Table 5. Descriptive statistics for EFL learners' RT and accuracy

Item type	RT *(SD)*	Accuracy
Congruent	1561.82 (663.93)	79.09
Incongruent	1650.46 (725.73)	56.84
Control	1785.03 (683.71)	66.44

Table 6. Mixed-effects model for EFL learners' RT

| | Fixed effects | | | | Random effects | | | |
| | | | | | By participant | | By item | |
	Estimate [95% CI]	SE	t	p	Variance	SD	Variance	SD
Intercept	7. 27 [7.19, 7.36]	.04	164.45	<.001***	.04	.21	.01	.12
L2 collocation frequency	−.17 [−.25, −.09]	.04	−4.22	<.001***				
Incongruent	.07 [.01, .14]	.03	2.24	.03*				
L1 Word1 frequency	−.04 [−.09, −.0002]	.02	−1.97	.05*				
L1 Word2 frequency	−.05 [−.10, −.003]	.03	−2.07	.04*				
Proficiency	−.001 [−.002, .0003]	.001	−1.48	.15				
Length of instruction	−.02 [−.05, .02]	.02	−.95	.35				
L2 collocation frequency: Incongruent	.10 [−.01, .21]	.06	1.71	.09				
L1 Word1 frequency: Length of instruction	−.009 [−.02., .0001]	.005	−1.93	.05				
Length	.02 [.002, .03]	.007	2.23	.03*				

*$p < 0.5$; **$p < .01$; ***$p < .001$

Table 7. Mixed-effects model for EFL learners' RT

| | Fixed effects | | | | Random effects | | | |
| | | | | | By participant | | By item | |
	Estimate [95% CI]	SE	z	p	Variance	SD	Variance	SD
Intercept	1.72 [1.20, 2.24]	.27	6.47	<.001***	.98	.99	1.05	1.02
L2 collocation frequency	1.02 [.56, 1.48]	.23	4.41	<.001***				
Incongruent	−1.35 [−1.87, −.83]	.27	−5.08	<.001***				
L1 Word1 frequency	.33 [−.01, .67]	.17	1.90	.06				
L1 Word2 frequency	.25 [−.15, .65]	.21	1.22	.22				
Proficiency	.04 [−.37, .44]	.21	.17	.86				
Length of instruction	.26 [−.14, .65]	.20	1.27	.20				
Incongruent: Proficiency	−.33 [−.55, −.11]	.11	−2.91	.003**				
L2 collocation frequency: Proficiency	.27 [.06, .48]	.11	2.53	.01*				

*$p < 0.5$; **$p < .01$; ***$p < .001$

processing. The interaction between L1 Word1 frequency and length of instruction showed a trend toward significance. This interaction suggested that L1 Word1 frequency effect would be slightly larger for learners with longer length of instruction.

In terms of variables that moderated the congruency effect (RQ3), there was no significant interaction between L2 collocation frequency and congruency, indicating that EFL learners showed a congruency effect even for high-frequency collocations. Length of instruction did not interact with congruency, meaning that more classroom instruction did not reduce L1 influence in EFL learners' collocational processing. The RT data showed that L2 proficiency did not moderate congruency. This is in line with Wolter and Gyllstad (2013) and Wolter and Yamashita (2018), which revealed null results of L2 proficiency, but stands in contrast with Sonbul and El-Dakhs (2020), who found that advanced learners showed a smaller congruency effect. As mentioned in the literature review, Sonbul and El-Dakhs (2020) attributed the null results in Wolter and Gyllstad (2013) and Wolter and Yamashita (2018) to the treating of L2 proficiency as a categorical variable and to including participants of a limited range of proficiency levels.

In the current study, even when L2 proficiency was analyzed as a continuous variable and the range of proficiency was fairly wide (i.e., from 4th to 90th percentile), L2 proficiency did not come out as a significant moderating variable of congruency. In the accuracy analysis, L2 proficiency interacted with L2 collocational frequency. It seems that EFL learners with higher proficiency level benefitted more from greater collocational frequency. Proficiency interacted with congruency in an unexpected direction such that the higher the L2 proficiency, the more errors EFL learners made when responding to incongruent collocations. Although such results seemed counterintuitive, one speculation was that EFL learners of lower proficiency lacked confidence in their ability and tended to think that items they did not know were not necessarily unacceptable. Thus, even when these lower level learners did not know the incongruent items, they accepted the items as collocations, leading to higher accuracy. If this was the case, lower level learners should also have made more errors for noncollocational control items, thinking that those items were collocations that they have not learned. In contrast, learners of higher proficiency would be more certain and reject items that they were not familiar with—namely, incongruent collocations as well as noncollocates. To evaluate this explanation, a follow-up analysis on the relationship between proficiency and accuracy for noncollocational control items was conducted. Proficiency turned out to be significantly and positively related to accuracy for noncollocational control items ($z = 4.60$, $p < .001$; see Appendix D for the full model), suggesting that EFL learners with lower proficiency were indeed more likely to accept a control item as a collocation. The lack of proficiency and congruency interaction in the RT analysis, which only involved correct responses, also supported this explanation: Once the EFL learners have learned the collocations, proficiency had little influence on the learners' processing of congruent and incongruent collocations.

General discussion

The findings of the study can be summarized as follows. First, EFL and ESL learners reacted differently to congruency (RQ1): ESL learners only showed a congruency effect in processing accuracy but not speed, whereas EFL learners' accuracy and speed were both affected by congruency. In response to RQ2 on L1 activation, L1 lexical frequencies of individual words in an L2 collocation affected both ESL and EFL learners' processing speed, an indication that L1 was activated regardless of whether a learner showed a congruency effect. Finally, in terms of the moderating variables on the congruency effect (RQ3), L2 collocation frequency did not interact with congruency, meaning that repeated encounter of a structure or the lack of it did not affect how L2 learners were influenced by their L1 during L2 processing. ESL learners' length of residence in the US moderated the effect of congruency: The longer ESL learners resided in the US, the faster and more accurately they reacted to incongruent collocations. The congruency effect was not reduced by EFL learners' L2 proficiency or length of instruction. Below I discuss the results in relation to (1) the underlying mechanisms of the congruency effect (based on RQ2) and (2) how language experiences and L2 proficiency moderate L1 influence (based on RQs 1 and 3).

The underlying mechanism of the congruency effect

To reiterate, two explanations have been proposed in the literature to account for the congruency effect—namely, L1 activation and order of acquisition. The results of the

current study favor L1 activation for two reasons. First, the absence of a congruency effect in ESL learners contradicts the proposal that the congruency effect comes from earlier acquisition of congruent items. If it is the case that congruent items are processed faster because they are learned earlier, they are likely to always enjoy an advantage over their incongruent counterparts regardless of whether a learner has lived in the L2 environment. Second, the effect of L1 lexical frequency on L2 collocational processing constitutes direct evidence for L1 activation.[1] That is, the higher the L1 lexical frequencies of individual words in the L2 collocations, the faster L2 learners processed the collocations. More importantly, L1 was activated for both congruent and incongruent collocations and for EFL as well as ESL learners even when ESL learners did not show a congruency effect in RT. These findings indicate that, consistent with Conklin and Carrol's (2018) hypothesis, L1 translation equivalents of words in the L2 collocations are automatically activated. It is not known, however, what routes learners may take after the activation of L1 translation equivalents to access the concepts of the L2 collocations. Based on Conklin and Carrol (2018) and Jiang (2000), I propose some possible scenarios in the following paragraphs that could account for the presence of the congruency effect among EFL learners and the absence of it among ESL learners.

For incongruent collocations, it is most likely that after activation of the L1 translation equivalents of individual words, L2 learners need to suppress their L1 because the activated L1 translation equivalents do not constitute felicitous L1 expressions. L2 learners then switch to the L2 route, which involves the activation of L2 collocations to directly access the concepts. The processing speed for incongruent collocations, therefore, depends on the strength of the conceptual link between an L2 collocation and its meaning and/or how good L2 learners are at suppressing their L1. EFL learners may have a comparatively weak conceptual link and/or a more dominant L1 that is harder to suppress. ESL learners, in comparison, may have established a stronger link between L2 collocations and concepts, or their L1 might be inhibited by the L2 immersive environment (Linck et al., 2009).

For congruent collocations, there are three possible scenarios. The first one entails dual activation, where the activated L1 translation equivalents lead to the activation of the corresponding L1 collocation as well as the L2 collocation. The dual activation may explain EFL learners' faster processing of congruent collocations over incongruent ones (Wolter & Gyllstad, 2011). Such a route, however, seems less plausible for ESL learners, who did not show a congruency effect. In the second scenario, the L1 translation equivalents activate the L1 collocation, through which meaning is accessed. For ESL learners, strong representations have been established for L2 collocations, and thus there is little difference in speed between the L2 route for incongruent collocations and the L1-mediated route for congruent ones. EFL learners may also take this route. But

---

[1]A reviewer pointed out that to establish direct evidence for L1 activation, apart from significant effects of L1 lexical frequencies, it was also necessary to demonstrate that models with L1 frequencies provide a better fit than those with L2 frequencies. Following this suggestion, additional analyses were conducted. In the first analysis, I fit two models for RT data. One with L1 frequencies, the main variables, and covariates in the study (i.e., condition, L2 collocation frequency, length of instruction, L2 proficiency, length of residence, age of arrival, percentage of L2 use, collocation length, and MI). The other model included L2 frequencies, the main variables, and covariates. This analysis was done for both EFL and ESL RT data. In the second analysis, I used the final models for RT data selected in the original manuscript. I then replaced L1 frequencies in those models with L2 frequencies. Again, this was done for both EFL and ESL RT data. In both analyses, the models with L1 frequencies were better fitting (i.e., lower AIC) than the ones with L2 frequencies. Appendix E summarizes the specifications and AICs for each model in the two analyses.

because their L2 representations are still weak, the L2 route may be slower than the L1-mediated route, resulting in the congruency effect. Finally, in the third scenario, for ESL learners, it is possible that after the automatic activation of L1 translation equivalents of the individual words, the L2 collocation and its concept are directly activated, without the mediation of the L1 corresponding collocation. The strength of the conceptual links for both congruent and incongruent collocations is similar, and thus there was no congruency effect. This route, which does not involve activation of L1 collocations, may account for the lack of L1 collocation frequency effect in Wolter and Gyllstad (2013).

Language experiences, proficiency, and L1 influence

The congruency effect is an indication of the amount of L1 influence in L2 processing. Analyzing learners in two different learning contexts (ESL and EFL) and the moderating variables on the congruency effect (i.e., L2 collocation frequency, length of residence, age of arrival, and length of instruction) informs us about how language experiences may affect L1 influence. Frequency indicates the more specific micro-L2 experiences that learners have with individual linguistic units (Spätgens & Schoonen, 2019). Learning context and related variables (length of residence, age of arrival, and length of instruction) represent the macro aspect of one's L2 experiences—that is, time being immersed in the L2 environment or learning in a classroom setting.

For the micro aspect of experiences, the results for both RT and accuracy showed that regardless of learning context, ESL and EFL learners were tuned to L2 collocation frequency, a finding consistent with usage-based approaches to language acquisition that learning is driven by experiences (e.g., Bybee, 2006). More importantly, the effect of L2 collocation frequency did not interact with congruency, a result similar to that of Wolter and Gyllstad (2013). That is, the congruency effect existed for both high- and low-frequency collocations in EFL learners' RTs and accuracy and ESL learners' accuracy. This suggested that although repeated encounters can contribute to fluent and accurate processing of collocations, it does not seem to reduce the influence of the L1.

Learning context, the macro aspect of L2 experiences, on the other hand, affects the amount of L1 influence in L2 processing. ESL learners, different from EFL learners, did not show a congruency effect in terms of processing speed. This finding is in line with Yamashita and Jiang (2010), who attributed the discrepancy between ESL and EFL learners to their different degrees of dependence on the L1 lexicon as a result of learning context difference: When an L2 collocation is first learned, it is linked to its L1 counterpart, which serves as a mediation to meaning, and this dependence on L1 gradually subsides with more L2 exposure, leading to the lack of congruency effect in ESL learners' RTs. The effects of length of residence on ESL learners' collocational processing further support the role of learning context in reducing L1 influence. ESL learners' processing of incongruent collocations was faster with longer length of residence in the US. For processing accuracy, though ESL learners still showed a congruency effect, the longer they lived in the L2 environment, the smaller the congruency effect was. This implies that immersing in an L2 environment alleviates the negative effect of L1-L2 incongruency, making it less difficult to learn and accept L2 collocations that did not have counterparts in the L1. In contrast, for EFL learners in a classroom setting, having more language instruction or being higher in language proficiency did not seem to reduce L1 influence in L2 collocational processing in terms of either speed or accuracy.

The differential effects of L2 collocation frequency, learning context and its related variables (i.e., length of residence, age of arrival, and length of instruction), and L2 proficiency on L1 influence suggest that frequent inhibition of the L1 may be the key to reducing L1 influence. Linck et al. (2009) showed that when learners were immersed in the L2 environment, L1 access was constantly suppressed, leading to attenuated L1 influence. In the current study, EFL learners may have encountered an L2 collocation repeatedly, had substantial classroom instruction, and achieved relatively high proficiency, but they were still in an L1 environment with active use of the L1. In contrast, ESL learners not only used the L1 less but also had to suppress the L1 when they used the L2. Reduced access to and frequent inhibition of the L1 may have resulted in lower level of L1 activation, which allowed the ESL learners to become more independent of the L1 in L2 collocational processing. This explanation also provides support for the argument in the previous discussion section that the processing of incongruent collocations depends on how well L2 learners are at L1 suppression.

Limitations

The findings of the study should be interpreted in light of several limitations. First, the L2 proficiency levels of ESL and EFL learners were not matched. The potential differences in proficiency levels between the two groups made it hard to attribute their processing differences solely to the learning context. Although it was found that L2 proficiency did not reduce the congruency effect, we still cannot rule out the possibility that the lack of congruency effect in ESL learners was a result of immersion and higher L2 proficiency combined. However, L2 proficiency and learning context are two constructs that are challenging to disentangle. Immersion might improve L2 proficiency in ways such as enhancing processing efficiency, which may not be captured solely by standardized tests. In fact, in Jiang et al. (2020), length of immersion was used as a proxy of L2 proficiency. When matching the proficiency levels of ESL and EFL learners, future studies can use both standardized tests and RT-based measures to provide more comprehensive proficiency profiles of their participants.

The second limitation is related to participants' familiarity with constituent words in the collocations. Participants' not knowing the constituent words may have changed the nature of "No" responses in the AJT: Participants indicated that an item was not a collocation because they didn't know some of the constituent words, not because they thought that the item was a noncollocate. Although the constituent words were estimated to be known to participants based on their proficiency level, future studies should directly check participants' knowledge of the words in an exit questionnaire.

Another limitation of the study concerns the accuracy of information about learners' L2 use. ESL participants' understanding of language use may have varied, with some only considering speaking the language as language use for example. The use of percentage as the unit of L2 use may have further contributed to variations in responses. Definitions of L2 use should be provided and L2 use should be measured in hours in future studies to obtain more accurate information.

Finally, as Yamashita (2018) pointed out, the semantic transparency effect may be involved in the congruency effect: Congruent collocations used in previous studies tended to be more transparent than incongruent collocations, and transparent collocations were processed faster (e.g., Gyllstad & Wolter, 2016). Although the manipulation check showed that L1 English speakers processed congruent and incongruent collocations used in the current study in a similar manner, indicating that these two

types of collocations were comparable in transparency, future studies should still obtain semantic transparency ratings of items to reduce the potential confounding effect of transparency.

Conclusion

The current study adds to our understanding of the mechanism underlying the congruency effect in L2 collocational processing and how L1 influence in L2 collocation processing may vary in relation to learners' L2 proficiency and linguistic experiences, indexed by L2 collocation frequency and the learning context. By examining the effect of L1 translation frequencies of individual words in an L2 collocation, the study provides direct evidence of L1 activation and pinpoints that the unit of L1 activation is individual words in an L2 collocation. The findings also showed that repeated encounter with an L2 collocation alone may not be sufficient to attenuate L1 influence in L2 processing. Being immersed in the L2 environment and the duration of immersion, on the other hand, may have a positive effect on reducing L1 influence. Further research is needed to elucidate how immersion attenuates L1 influence—for example, through L1 inhibition. Findings in this study should also be verified in other tasks, such as self-paced reading, and with other types of collocations.

Supplementary material. The supplementary material for this article can be found at http://doi.org/10.1017/S0272263123000281.

Data availability statement. The experiment in this article earned an Open Materials badges for transparent practices. The materials and data are available at https://osf.io/e4sng/?view_only=97246182b67546f08ce160b461399e92

Competing interest. We have no known conflict of interest to disclose.

References

Amuzie, G. L., & Winke, P. (2009). Changes in language learning beliefs as a result of study abroad. *System*, *37*, 366–379. https://doi.org/10.1016/j.system.2009.02.011

Anwyl-Irvine, A. L., Massonié J., Flitton, A., Kirkham, N. Z., & Evershed, J. K. (2019). Gorilla in our midst: An online behavioural experiment builder. *Behavior Research Methods*, *52*, 388–407, https://doi.org/10.3758/s13428-019-01237-x

Baker, W. (2010). Effects of age and experience on the production of English word-final stops by Korean speakers. *Bilingualism: Language and Cognition*, *13*, 263–278. https://doi.org/10.1017/S136672890999006X

Baker-Smemoe, W., Dewey, D. P., Bown, J., & Martinsen, R. A. (2014). Variables affecting L2 gains during study abroad. *Foreign Language Annals*, *47*, 464–486. https://doi-org.proxy2.cl.msu.edu/10.1111/flan.12093\

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. https://doi.org/10.1016/j.jml.2012.11.001

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*, 1–48. https://doi.org/10.18637/jss.v067.i01

Bybee, J. (1985). *Morphology: A study of the relation between meaning and form.* John Benjamins. https://doi.org/10.1075/tsl.9

Bybee, J. (2006). From usage to grammar: The mind's response to repetition. *Language*, *82*, 711–733. http://doi.org/10.1353/lan.2006.0186

Cai, Q., & Brysbaert, M. (2010). SUBTLEX-CH: Chinese word and character frequencies based on film subtitles. *Plos One*, *5*, Article e10729. https://doi.org/10.1371/journal.pone.0010729

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. https://doi.org/10.1017/S1366728913000795

Carrol, G., & Conklin, K. (2017). Cross language priming extends to formulaic units: Evidence from eye-tracking suggests that this idea "has legs." *Bilingualism: Language and Cognition*, *20*, 299–317. https://doi.org/10.1017/S1366728915000103

Carrol, G., Conklin, K., & Gyllstad, H. (2016). Found in translation: The influence of L1 on the processing of idioms in L2. *Studies in Second Language Acquisition*, *38*, 403–443. https://doi.org/10.1017/S0272263115000492

Conklin, K., & Carrol, G. (2018). First language influence on the processing of formulaic language in a second language. In A. Siyanova-Chanturia & A. Pellicer-Sánchez (Eds.), *Understanding formulaic language: A second language acquisition perspective* (pp. 62–77). Routledge. https://doi.org/10.4324/9781315206615

Cop, U., Keuleers, E., Drieghe, D., & Duyck, W. (2015). Frequency effects in monolingual and bilingual natural reading. *Psychonomic Bulletin & Review*, *22*, 1216–1234. https://doi-org.proxy2.cl.msu.edu/10.3758/s13423-015-0819-2

Cubillos, J. H., and Ilvento, T. (2012). The impact of study abroad on students' self-efficacy perceptions. *Foreign Language Annals*, *45*, 494–511. https://doi-org.proxy2.cl.msu.edu/10.1111/j.1944-9720.2013.12002.x

DeKeyser, R. M. (2000). The robustness of critical period effects in second language acquisition. *Studies in Second Language Acquisition*, *22*, 499–533. https://doi.org/10.1017/S0272263100004022

DeKeyser, R., Alfi-Shabtay, I., & Ravid, D. (2010). Cross-linguistic evidence for the nature of age effects in second language acquisition. *Applied Psycholinguistics*, *31*, 413–438. https://doi.org/10.1017/S0142716410000056

Du, L., Elgort, I., & Siyanova-Chanturia, A. (2021). Cross-language influences in the processing of multiword expressions: From a first language to second and back. *Frontiers in Psychology*, *12*, Article 666520. https://doi.org/10.3389/fpsyg.2021.666520

Ellis, N. C., Simpson-Vlach, R., & Maynard, C. (2008). Formulaic language in native and second language speakers: Psycholinguistics, corpus linguistics, and TESOL. *TESOL Quarterly*, *42*, 375–396. https://doi.org/10.1002/j.1545-7249.2008.tb00137.x

Fletcher, W. H. (2011). Phrases in English (PIE). Retrieved June 30, 2020, from http://phrasesinenglish.org/

Grange, J. (2015). trimr: An implementation of common response time trimming methods. R package version 1.0.1. https://CRAN.R-project.org/package=trimr

Gyllstad, H., & Wolter, B. (2016). Collocational processing in light of the phraseological continuum model: Does semantic transparency matter? *Language Learning*, *66*, 296–323. https://doi-org.proxy2.cl.msu.edu/10.1111/lang.12143

Hair, J. F., Jr., Anderson, R. E., Tatham, R. L., & Black, W. C. (1995). *Multivariate data analysis* (3rd ed.). Macmillan.

Janssen, N., & Barber, H. A. (2012). Phrase frequency effects in language production. *Plos One*, *7*, Article e33202. https://doi.org/10.1371/journal.pone.0033202

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

Jiang, N., Li, M., & Guo, T. (2020). A tale of two frequency effects: Toward a verification model of L2 word recognition. *Applied Psycholinguistics*, *41*, 215–236. https://doi.org/10.1017/S0142716419000481

Kroll, J. F., & Stewart, E. (1994). Category interference in translation and picture naming: Evidence for asymmetric connections between bilingual memory representations, *Journal of Memory and Language*, *33*, 149–174. https://doi.org/10.1006/jmla.1994.1008

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software*, *82*, 1–26. https://doi.org/10.18637/jss.v082.i13

Linck, J. A., Kroll, J. F., & Sunderman, G. (2009). Losing access to the native language while immersed in a second language: Evidence for the role of inhibition in second-language learning. *Psychological Science*, *20*, 1507–1515. https://doi.org/10.1111/j.1467-9280.2009.02480.x

Lüdecke, D., Makowski, D., & Waggoner, P. (2019). *Performance: Assessment of regression models performance* (R package version 0.4.0) [Computer software]. R Foundation for Statistical Computing. https://CRAN.R-project.org/package=performance

Muñoz, C. (2014). Contrasting effects of starting age and input on the oral performance of foreign language learners. *Applied Linguistics*, *35*, 463–482. https://doi_org.proxy2.cl.msu.edu/10.1093/applin/amu024

Nation, P. (2012). The BNC/COCA word family lists. https://www.wgtn.ac.nz/__data/assets/pdf_file/0005/1857641/about-bnc-coca-vocabulary-list.pdf

National Education Examinations Authority. (n.d.). 分数解释 [Explanation of scores]. http://cet.neea.edu.cn/html1/folder/19081/5124-1.htm

Nesselhauf, N. (2003). The use of collocations by advanced learners of English and some implications for teaching. *Applied Linguistics*, *24*, 223–242. https://doi-org.proxy2.cl.msu.edu/10.1093/applin/24.2.223

Öksüz, D., Brezina, V., & Rebuschat, P. (2021). Collocational processing in L1 and L2: The effects of word frequency, collocational frequency, and association. *Language Learning*, *71*, 55–98. https://doi-org.proxy2.cl.msu.edu/10.1111/lang.12427

Pellicer-Sánchez, A., Siyanova-Chanturia, A., & Parente, F. (2022). The effect of frequency of exposure on the processing and learning of collocations: A comparison of first and second language readers' eye movements. *Applied Psycholinguistics*, *43*, 727–756. https://doi.org/10.1017/S014271642200011X

Puimège, E., & Peters, E. (2020). Learning formulaic sequences through viewing L2 television and factors that affect learning. *Studies in Second Language Acquisition*, *42*, 525–549. https://doi.org/10.1017/S027226311900055X

R Core Team. (2019). *R: A language and environment for statistical computing* (version 3.6.1) [Computer software]. R Foundation for Statistical Computing.

Sinclair, J. (1991). *Corpus, concordance, collocation*. Oxford University Press.

Siyanova-Chanturia, A., & Pellicer Sánchez, A. (2019). Formulaic language: Setting the scene. In A. Siyanova-Chanturia & A. Pellicer-Sánchez (Eds.), *Understanding formulaic language: A second language acquisition perspective* (pp. 1–15). Routledge. https://doi-org.proxy2.cl.msu.edu/10.4324/9781315206615

Siyanova-Chanturia, A., & Spina, S. (2015). Investigation of native speaker and second language learner intuition of collocation frequency. *Language Learning*, *65*, 533–562. https://doi.org/10.1111/lang.12125

Sonbul, S., and El-Dakhs, D. (2020). Timed versus untimed recognition of L2 collocations: Does estimated proficiency modulate congruency effects? *Applied Psycholinguistics*, *41*, 1197–1222. https://doi.org/10.1017/S014271642000051X

Spätgens, T., & Schoonen, R. (2019). Single, but not unrelated: Key issues in researching single-word items. In S. Webb (Ed.), *The Routledge handbook of vocabulary studies* (pp. 511–528). Routledge. https://doi-org.proxy2.cl.msu.edu/10.4324/9780429291586

Thierry, G., & Wu, Y. J. (2007). Brain potentials reveal unconscious translation during foreign-language comprehension. *Proceedings of the National Academy of Sciences*, *104*, 12530–12535.

Tremblay, A., & Tucker, B. V. (2011). The effects of N-gram probabilistic measures on the recognition and production of four-word sequences. *The Mental Lexicon*, *6*, 302–324. https://doi.org/10.1075/ml.6.2.04tre

Van Heuven, W. J. B., Mandera, P., Keuleers, E., & Brysbaert, M. (2014). SUBTLEX-UK: A new and improved word frequency database for British English. *Quarterly Journal of Experimental Psychology*, *67*, 1176–1190. https://doi.org/10.1080/17470218.2013.850521

Webb, S., Newton, J., & Chang, A. (2013). Incidental learning of collocation. *Language Learning*, *63*, 91–120. https://doi-org.proxy2.cl.msu.edu/10.1111/j.1467-9922.2012.00729.x

Wolter, B. (2006). Lexical network structures and L2 vocabulary acquisition: The role of L1 lexical/conceptual knowledge. *Applied Linguistics*, *27*, 741–747. https://doi-org.proxy2.cl.msu.edu/10.1093/applin/aml036

Wolter, B., & Gyllstad, H. (2011). Collocational links in the L2 mental lexicon and the influence of L1 intralexical knowledge. *Applied Linguistics*, *32*, 430–449. https://doi.org/10.1093/applin/amr011

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*, 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*, 1193–1221. https://doi.org/10.1017/S0142716414000113

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. https://doi.org/10.1017/S0272263117000237

Wood, D. (2019). Classifying and identifying formulaic language. In S. Webb (Ed.), *The Routledge handbook of vocabulary studies* (pp. 511–528). Routledge. https://doi-org.proxy2.cl.msu.edu/10.4324/9780429291586

Wu, Y. J., & Thierry, G. (2010). Chinese–English bilinguals reading English hear Chinese. *Journal of Neuroscience*, *30*, 7646–7651. https://doi-org.proxy2.cl.msu.edu/10.1523/JNEUROSCI.1602-10.2010

Yamashita, J. (2018). Possibility of semantic involvement in the L1-L2 congruency effect in the processing of L2 collocations. *Journal of Second Language Studies*, *1*, 60–78. https://doi.org/10.1075/jsls.17024.yam

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

Yi, W., Lu, S., & Ma, G. (2017). Frequency, contingency and online processing of multiword sequences: An eye-tracking study. *Second Language Research*, *33*, 519–549. https://doi-org.proxy2.cl.msu.edu/10.1177/026765831770800

Yi, W. (2018). Statistical sensitivity, cognitive aptitudes, and processing of collocations. *Studies in Second Language Acquisition*, *40*, 831–856. https://doi.org/10.1017/S0272263118000141

Zhao, P., & Ji, X. (2018). Validation of the Mandarin version of the Vocabulary Size Test. *RELC Journal*, *49*, 308–321. https://doi-org.proxy2.cl.msu.edu/10.1177/0033688216639761