languages

Article

The Effect of Congruency and Frequency of Exposures on the Learning of L2 Binomials

Abdulaziz Altamimi [1] and Kathy Conklin [2,*]

[1] Department of English, Imam Mohammad Ibn Saud Islamic University, Riyadh 11432, Saudi Arabia; aitamimi@imamu.edu.sa
[2] School of English, University of Nottingham, Nottingham NG7 2RD, UK
* Correspondence: k.conklin@nottingham.ac.uk

Abstract: Although extensive research has been carried out on opaque formulaic language where the meaning is not the sum of the individual words (i.e., idioms and many collocations), it is still not clear how cross-language congruency and frequency of exposure influence the learning of transparent formulaic language in an L2. In the current study, self-paced reading along with offline word order recognition tasks were used to investigate the role of cross-language congruency and the frequency of exposure on the learning and processing of fully transparent binomials. In the self-paced reading, Arabic second language learners of English and native English speakers encountered three types of binomial phrases either two or five times in English texts: English-only binomials, Arabic-only binomials that were translated into English, and congruent binomials (acceptable in English and Arabic). A subsequent offline task revealed that both native and non-native speakers developed knowledge of the 'correct' order of binomials (i.e., *fish and chips*, not *chips and fish*) after only two exposures in the self-paced reading. Native speakers were more accurate on congruent and English-only items than Arabic-only items, while non-natives speakers exhibited no differences in accuracy across the binomial types. The offline task showed that native speakers responded faster to congruent and English-only items than Arabic-only, and non-native speakers responded faster to congruent items than English-only and Arabic-only. The frequency of exposure had no effect, with no difference in response times and accuracy between two and five exposures.

Keywords: formulaic language; multiword sequences; binomials; self-paced reading; second language learning

Citation: Altamimi, Abdulaziz, and Kathy Conklin. 2024. The Effect of Congruency and Frequency of Exposures on the Learning of L2 Binomials. *Languages* 9: 9. https://doi.org/10.3390/languages9010009

Academic Editor: Gareth Carrol

Received: 1 August 2023
Revised: 30 November 2023
Accepted: 30 November 2023
Published: 21 December 2023

1. Introduction

While both *fish and chips* and *chips and fish* are perfectly grammatical phrases in English, native speakers are much more likely to say the former than the latter. For second language learners, this poses a considerable challenge: they need to learn what grammatical utterances are more probable (Pawley and Syder 1983). Lexical patterns like *fish and chips* are often referred to as formulaic language or multiword sequences and are ubiquitous in language. Corpus-based research shows that between 30% to 50% of language use is formulaic (e.g., Erman and Warren 2000; Kuiper 2004). Thus, L2 speakers need to learn a considerable amount of formulaic language.

Formulaic language differs in how much overlap there is across languages. For example, formulaic language can (1) overlap in meaning and form (e.g., the English collocation *warm reception* which has a word-for-word translation with the same meaning as an Arabic collocation *istigbal har*); (2) overlap in meaning but not in form (e.g., an idiom describing something as expensive in English is *costs an arm and leg*, and in Arabic, *kallaf dam qalbi* '*costs the blood of my heart*'); or (3) have no formulaic equivalent in the other language (e.g., the binomial *safe and sound* in English does not correspond to a binomial phrase in Arabic). In the present study, a formulaic sequence is considered congruent only if it has a word-for-word translation with the same meaning in the other language (i.e., the first category: overlap in meaning and form).

Investigations of L1-L2 congruency have become an increasingly important area in formulaic language research (for an overview, see Conklin and Carrol 2019). Previous research examining congruency has focused mainly on idioms (e.g., Beck and Weber 2016; Carrol and Conklin 2014, 2017; Carrol et al. 2016) and collocations (e.g., Wolter and Gyllstad 2011, 2013; Wolter and Yamashita 2017). Studies on these two types of formulaic language have demonstrated that when there is overlap between the two languages, processing is facilitated (Wolter and Gyllstad 2011, 2013; Yamashita and Jiang 2010; Carrol et al. 2016). On the other hand, when there is no L1-L2 correspondence, processing tends be disrupted (Wolter and Yamashita 2017), and learning becomes challenging (Peters 2016). However, little research has investigated the role of congruency in the learning and processing of binomial phrases (e.g., *salt and pepper*), which is the focus of the current research. Thus, the present study adds to our understanding of the learning and processing of binomial phrases, a less investigated type of formulaic language, by examining the effect of L1-L2 congruency.

Binomials are "recurrent (frequent), familiar (conventional) expressions" consisting of two words from the same lexical class joined by a conjunction, where one word order is more frequent (e.g., *king and queen* not *queen and king*, and *fish and chips* not *chips and fish*) (Siyanova-Chanturia et al. 2011, p. 2). However, developing a sensitivity to the word order preference of binomials (e.g., *salt and pepper* not *pepper and salt*) can be challenging for L2 learners. Binomials differ from idioms and other formulaic language in many ways. For example, binomials are structurally more predictable and more frequent than idioms. Further, binomials are generally transparent such that the overall meaning can be computed from the parts, with transparent binomials being the focus of the current investigation. Notably, an important facet of binomials, in contrast to other types of formulaic sequences, is that a change of word order does not alter the meaning or result in syntactic irregularity. The flexibility in word order, without introducing confounds (e.g., unigram frequencies, syntactic and semantic properties), makes binomials a good choice for investigating the role of phrasal frequency in L2 acquisition (Siyanova-Chanturia et al. 2011).

2. Literature Background

2.1. Congruency Effects

It is thought that L1-L2 congruency impacts formulaic language learning and processing. The majority of studies that investigated congruency have focused on idioms and collocations. Thus, studies on idioms and collocations will be reviewed to provide a background on the influence of congruency on formulaic language. Yamashita and Jiang (2010) explain that processing congruent collocations involves using ready-made links to L1 knowledge, which assists comprehension, whereas processing incongruent collocations is either more compositional and/or requires contextual cues to infer meaning, which requires more effort. Yamashita and Jiang say that when learning congruent collocations, L2 learners use links to L1 collocational counterparts to map the L2 collocational meanings. This yields faster recognition of L2 collocations that have an equivalent L1 collocation. For idioms, Carrol et al. (2016) put forth a similar explanation, emphasizing the role of cross-language priming. Specifically, when L2 learners encounter an L2 word, which comprises an initial part of a congruent idiom, its L1 counterpart word receives activation. The activation of the L1 word passes activation to the L1 idiom, as well as its L2 counterpart, which speeds up the recognition and processing of congruent idioms.

The majority of studies investigating congruency in formulaic language have focused on idioms, which are opaque or non-compositional formulaic language (i.e., overall meaning is not the sum of the meaning of the individual words like *kick the bucket*). For example, the study by Carrol et al. (2016) used eye-tracking to look at the processing of English-only idioms (not found in Swedish), Swedish-only translated idioms (not found in English), and congruent idioms (found in both languages). For native English speakers, the findings demonstrated the expected processing advantage for English-only and congruent idioms, but that advantage was not found for unfamiliar Swedish-only items. For Swedish learners of English, they found a processing advantage for congruent and Swedish-only idioms, but not for the English-only idioms.

Titone et al. (2015) investigated the effect of congruency on the processing of English idioms that varied in their cross-language overlap with French. English–French bilinguals read English sentences containing idioms or matched control phrases (e.g., *she lived/told a lie*) in a word-by-word presentation, where the final word was either in English or in French (e.g., intact condition: *he played with fire* vs. code-switched condition: *he played with feu*). Participants were asked to judge the meaningfulness of each sentence. They found that the code-switched condition was more disruptive for idioms than for control phrases, as indicated by judgment time. Importantly, increased cross-language overlap reduced response times for idioms in the code-switched condition.

Pritchett et al. (2016) investigated the effect of idiom congruency in a cued recall task. Russian–English bilinguals were exposed to adjective–noun idioms in English and/or Russian (e.g., English only: *blue blood*; English–Russian: *blue moon*; Russian only: *blue distances*). Following an exposure phase, participants were instructed to write down any phrases they could remember. Recall was significantly better for idiomatic phrases that existed in both languages than for phrases that only existed in one language.

In addition to idioms, evidence for congruency effects has been demonstrated in collocations. For example, Yamashita and Jiang (2010) asked native English speakers (monolinguals), Japanese ESL participants, and Japanese EFL participants to judge if phrases were acceptable in English when they were congruent collocations, existing in both languages (e.g., *heavy stone*), or incongruent, existing only in English (e.g., *kill time*). Native English speakers showed no difference in response times and accuracy for congruent and incongruent items. For Japanese EFL learners, incongruent items had longer response times and resulted in more errors compared to congruent items. The ESL learners behaved similarly to EFL learners in that they had more errors for incongruent than congruent items; however, the ESL response times resembled those of the native speaker group, with no difference between the two item types.

Wolter and Yamashita (2017) examined how knowledge of collocations in the L1 (Japanese) influenced the processing of collocations in the L2 (English), using an acceptability judgment task. The non-native speakers responded faster to Japanese–English congruent collocations than English-only incongruent collocations, while native speakers demonstrated equivalent performance across the conditions. Wolter and Gyllstad (2011, 2013) found a similar effect of congruency on collocation processing. They presented native English speakers and Swedish learners of English with congruent and incongruent collocations in a lexical/phrasal decision task. L2 learners had a processing advantage for congruent items compared to incongruent ones, while native English speakers demonstrated equivalent performance across those two conditions. It is important to note that not all studies have found an effect of congruency. For example, Leśniewska and Witalisz (2007) showed that Polish L2 learners of English did not have a preference for congruent collocations over incongruent ones in an acceptability judgment task.

Overwhelmingly, research has demonstrated that L1-L2 congruency facilitates the processing of collocations and idioms. It is important to note that this research has primarily focused on the effect of congruency on the L2 processing of formulaic language and not on its acquisition. Thus, whether congruency affects the learning of formulaic language remains largely unknown.

2.2. Reading Exposure and Frequency Effects

Studies on individual word learning have shown that a number of factors influence learning gains from reading, such as contextual cues (e.g., Webb 2008), the spacing of encounters (e.g., Nakata and Elgort 2021), L2 proficiency (e.g., Tekmen and Daloglu 2006) and the frequency of encounters of the new item (e.g., Pellicer-Sánchez and Schmitt 2010). A number of studies have also shown that L2 learners can acquire formulaic sequences during reading, with the frequency of encounter being an important factor in learning gains (e.g., Alotaibi et al. 2022; Pellicer-Sánchez 2017; Sonbul and Schmitt 2013; Webb et al. 2013). For example, Durrant and Schmitt (2010) showed that two exposures to a collocation in a read-aloud task generated better performance in a subsequent recall task than a single exposure by ESL learners. Furthermore, Webb et al. (2013), who modified a short story by incorporating multiple instances of the same collocation (i.e., from 1 to 15), found that more encounters resulted in better learning gains for their EFL participants. Much of language learning research describes the frequency of encounter as playing a key role in learning (Uchihara et al. 2019).

In sum, studies on formulaic language have demonstrated the effect of congruency on processing and the benefit of frequency of encounter on learning. However, little is known about the combined effect of congruency and the frequency of encounter on transparent formulaic language such as binomials.

2.3. Binomials Phrases

A key type of formulaic language that has been understudied is binomials. Studies on binomial processing have focused mainly on the effect of phrasal frequency on processing by comparing binomials to their reversed forms (Siyanova-Chanturia et al. 2017; Siyanova-Chanturia et al. 2011). Overall, such studies show that binomials elicit faster processing than their reversed forms (e.g., *time and money* vs. *money and time*) for native English speakers and higher proficiency non-native speakers. The advantage is thought to be due to phrasal frequency and entrenchment in memory of the lexical pattern.

Conklin and Carrol (2021) used eye tracking to investigate the processing of novel binomials by native speakers, which were presented one to five times in reading passages in an experimentally defined forward form (e.g., *wires and pipes*), and subsequently were presented in the reversed form (e.g., *pipes and wires*). Forward forms demonstrated a processing advantage over their reversed forms after four exposures, suggesting that native speakers developed sensitivity to the conventional word order of binomials during reading. Sonbul et al. (2023) aimed to replicate the Conklin and Carrol study with non-native speakers but found no advantage of the forward forms over their reversed ones. The two studies suggest that non-native speakers may require more exposures than native speakers to develop sensitivity to the preferred word order of binomials.

Alotaibi et al. (2022) investigated whether different input modes (reading-only, listening-only, and reading-while-listening) and the number of exposure (2, 4, 5, and 6 occurrences) had an effect on binomial learning for non-native speakers. The results demonstrated that reading-only and reading-while-listening led to greater gains than listening-only. There was also some evidence for an effect of the number of exposures: novel binomials that occurred more were perceived as more familiar.

To date, a study by Du et al. (2021) is the only one to examine the effect of congruency on binomial processing. In this study, L1-Chinese L2-English and L1-English L2-Chinese bilinguals (both immersed in an L2 environment), and English monolinguals took part in a lexical decision task to examine priming effects for congruent binomials (e.g., *knife and fork*), English-only binomials (e.g., *salt and pepper*), and Chinese-only translated binomials (e.g., *wisdom and strength*). L1-Chinese L2-English bilinguals showed priming for congruent but not English-only binomials, while L1-English L2-Chinese bilinguals demonstrated no priming effects. English monolinguals showed priming effects for the two types of English binomials. Notably, none of the groups showed a processing advantage for Chinese-only binomials. The authors argued that the L1 may be inhibited in immersion contexts, and this might explain the weak congruency effect and absence of an L1 effect for English–Chinese bilinguals who were studying Chinese in China.

In sum, although the reviewed studies add to our understanding of binomial learning and processing, they had limitations. The Alotaibi et al. (2022) and Sonbul et al. (2023) studies did not address the congruency effect, and the Du et al. (2021) study did not address the frequency effect on binomial learning and processing. It remains an open question whether congruency and frequency of encounter affect binomial learning, which is the focus of the current research.

2.4. The Present Study

Previous research suggests that congruency facilitates the processing of formulaic sequences. However, there are notable gaps in the literature. First, previous studies have generally focused on opaque formulaic language, where the overall meaning is not the sum of the meaning of the individual words (i.e., idioms such as *pull my leg* and many collocations with partially figurative meaning such as *kill time* or *broken heart*). Little is known about the effect of congruency on transparent formulaic language (e.g., binomials) learning and processing. Further, findings about idioms and collocations may not generalize to binomials since learning binomials requires a different type of knowledge: developing a sensitivity to the "correct" word order (e.g., *salt and pepper* not *pepper and salt*). Second, as noted by Conklin and Carrol (2019), figurativeness has often been confounded with congruency in many previous studies (e.g., Wolter and Gyllstad 2011, 2013; Yamashita and Jiang 2010). Thus, research needs to carefully consider and/or control transparency alongside congruency. Third, although frequency of encounter has been demonstrated to play an important role in learning (Webb et al. 2013), no study has investigated the combined effect of frequency of encounter and congruency on learning. The present study aimed to address these gaps by examining the learning and processing of transparent binomial phrases while manipulating congruency and the frequency of encounters.

The present study investigated the role of congruency and the frequency of encounter on binomial learning and processing in English native speakers and Arabic learners of English. It investigated three types of binomials: 1. English-only, 2. Arabic-only (translated into English), and congruent (i.e., acceptable in both languages with the same conventional word order preference). Participants engaged in four tasks: 1. pre-test; 2. reading passages with embedded binomials; 3. immediate post-test; and 4. delayed test. The study aimed to address the following questions:

(1) Does the number of exposures while reading impact the learning and processing of binomials?
(2) Does the cross-language congruency of binomials in a reading task impact their learning and processing?

3. Methods

3.1. Participants

Native Arabic speakers learning English as a foreign language (EFL) in Saudi Arabia (n = 47, all male) and British native speakers of English, who were all monolinguals (n = 18, female = 10, male = 9) with no knowledge of Arabic, took part in the study. The native English speakers provide a baseline for nativelike judgment of word order preference of the binomial phrases. Participants were undergraduate students who participated in exchange for course credit or for GBP 20. The proficiency level of both groups of participants was estimated using two objective tests: the Lexical Test for Advanced Learners of English (LexTALE; Lemhöfer and Broersma 2012) and a short multiple-choice vocabulary test (Carrol and Conklin 2014, 2017; Carrol et al. 2016) that was modified from Nation and Beglar's (2007) Vocabulary Size Test (i.e., 20 vocabulary items from different frequency bands). Participants also filled out a questionnaire, where they provided information about their language background and were asked to self-rate their knowledge of English on a seven-point scale (1 = very low, 7 = native-like). Table 1 summarises the demographic and language proficiency data.

Table 1. Means with standard deviations in parentheses for the demographic data and self-ratings of proficiency on a seven-point scale (1 = very low; 7 = native).

	Native Speakers (L1 English)	**Non-Native Speakers (L1 Arabic)**
Age	19.83(1.25)	20.19 (2.15)
Age of English acquisition	0 (0.0)	11.46 (2.88)
Years lived in an English-speaking country	19.44 (1.53)	0.24 (0.92)
Short vocabulary test score (max = 20)	18.77 (0.87)	9.29 (3.25)
LexTALE score (%)	92.29 (6.72)	56.84 (8.21)
Self-rating (1 to 7)		
Overall English proficiency	7 (0)	4.44 (1.47)
Proficiency in speaking	7 (0)	4.17 (0.91)
Proficiency in understanding	7 (0)	5.57 (0.94)
Proficiency in reading	7 (0)	5.10 (1.28)
Proficiency in writing	6.72 (0.55)	3.91 (1.50)

3.2. Materials

Item Selection. The items were all transparent binomial phrases of the form *noun and noun* and were of three types: English-only binomials (occur frequently in English but not Arabic, such as *fish and chips*), Arabic-only binomials (occur frequently in Arabic but not English and are translated into English such as *coffee and dates*), and congruent binomials (occur frequently in both English and Arabic with the same conventional word order preference such as *food and drink*).

The following procedure was followed for selecting items. First, English binomials were extracted from the British National Corpus (BNC), Arabic binomials were extracted from the KACST (Al-Thubaity 2015) (i.e., an Arabic corpus of over one billion Arabic words), and congruent binomials were identified in both corpora. A binomial phrase was considered congruent if it had a literal translation equivalent in English and Arabic with the same word order preference. All binomials had a frequency greater than 50 in the corpus/corpora from which they were extracted. Second, only binomial phrases whose forward (e.g., *food and drink)* phrasal frequency was at least three times larger than their backward (e.g., *drink and food*) frequency were selected. Third, for Arabic-only binomials, there needed to be a good translation, which was confirmed in a rating task (i.e., four native Arabic speakers with an advanced knowledge of English rated the translations on a five-point scale with 5 = "excellent translation" to 1 = "poor translation"). Only phrases that achieved a rating score of four or five were selected (M = 4.66, SD = 0.51). Fourth, the binomial items were all reversible, meaning that both orders were semantically and logically possible. Lastly, none of the selected binomials violated the semantic and phonological constraints thought to bias binomial word order preference (Benor and Levy 2006).[1] This left us with 14 experimental items per binomial type. An additional 14 English only items were included only in the testing sessions (i.e., pre-test, post-test, and delayed test) to measure the effect of the testing sessions themselves. These are referred to as "control items". The target items are presented in Table S1 in the Supplementary Materials.

A set of *t*-tests based on the BNC frequencies showed that the English items (English-only, congruent and control), had similar frequencies, p's > 0.05, while the Arabic-only items (translated into English) had a significantly lower English frequency, p's < 0.05. A further set of *t*-tests based on the Arabic corpus confirmed no significant differences between Arabic-only and congruent items, $p > 0.05$, and between English-only items when translated into Arabic and control items, $p > 0.05$, but Arabic-only and congruent items' frequencies were significantly higher than English-only and control items (p's < 0.05). In other words, the frequency analysis confirmed the item categories, in that English-only and congruent items were matched in their BNC frequencies, and Arabic-only and congruent items were matched in their Arabic corpus frequencies.[2]

With regard to individual word frequencies, a set of *t*-tests confirmed no differences in frequencies between the nouns making up the binomial phrases across all item categories (all p's > 0.05). This was important to control for because it is thought that word frequency could bias binomial ordering preference, such that the more frequent word often precedes the less frequent one (Benor and Levy 2006). We ensured that word frequency of the nouns was well matched in an item.

Norming Studies. The norming of the items was carried out across three studies involving native English speakers ($n = 34$) and Arabic non-native speakers of English ($n = 82$). Participants were different in each norming study and did not take part in the main study. Results from norming studies were considered as covariates in later analyses to control for any minor differences between the items in terms of subjective familiarity. A summary of the norming data can be found in the Supplementary Materials (see Table S2).

The first norming study was a word order recognition task, in which participants indicated which order of the binomial phrase sounded more natural/familiar to them (e.g., is it *gold and silver* or *silver and gold*). For the native speakers, *t*-tests revealed that congruent and English-only items had greater accuracy than Arabic-only items (p's < 0.05), with no difference between the English-only and congruent items ($p > 0.05$). For the non-native speakers, *t*-tests showed that the congruent items had greater accuracy than Arabic-only and English-only items, and that Arabic-only items had greater accuracy than English-only items (p's < 0.05).

The second norming study was a "forward" completion task. Participants were given the first noun and the conjunction as a prompt and were asked to fill in the blank with the first word that comes to mind (e.g., *time and* _____). The third norming study was a "backward" completion task, where participants were given the conjunction and the final noun and were asked to fill in the blank (e.g., _____ *and money*). Several independent sample *t*-tests were carried out to compare completion rates. The non-native speakers were more accurate on the forward completion task than the backward task in all item types (p's < 0.05). The native speakers were more accurate on the forward completion task than the backward task in the congruent phrases ($p < 0.05$), with no differences between forward and backward Arabic binomial phrases, and between forward and backward English-only phrases, p's > 0.05.

Reading Passages. In the reading treatment, target items were embedded in English passages that participants were asked to read for comprehension. Items occurred in the middle of a sentence within short paragraphs, with the same binomial item presented either two or five times in the paragraph. Paragraphs were either two sentences or five sentences long. Only one item occurred in a paragraph. The number of repetitions for each binomial item was counterbalanced across two lists, such that if an item appeared twice in a passage in List 1, the same item appeared five times in List 2. In each list, the order of presentation of the passages was randomized. A sample of passages in which the number of occurrences of items was counterbalanced across different lists is presented in Table S3 in the Supplementary Materials.

It was important to ensure that the passages would not be too difficult to understand for non-native speakers. According to the VocabProfile tool (Cobb n.d.), after proper nouns were excluded, 80.04% of the words in the texts came from the first thousand frequency band, 6.68% from the second thousand, and 6.35% came from the Academic Word List. The vocabulary profile analysis showed that low-frequency words were kept to a minimum, which would facilitate participants' comprehension of the passages.

In the reading treatment, participants encountered the passages in a non-accumulating, region-by-region, self-paced reading task. Self-paced reading allowed participants' reading time to be measured, and the non-accumulating text ensured that they did not revisit previously read items, which could have confounded the repetition manipulation. Participants' reading times in the self-paced reading task were recorded using Inquisit 5 from Millisecond Software (2015).

3.3. Pre-, Post-, and Delayed Tests

Participants' knowledge of the binomial phrases was assessed in a word order recognition task in which they were instructed to choose which form of a binomial phrase sounded more natural in English (e.g., *salt and pepper* vs. *pepper and salt*). They were encouraged to respond as quickly and accurately as possible. Participants' accuracy and response times in the word order recognition task were recorded using Inquisit 5. Accuracy was used as a measure of binomial knowledge, and response times were used to estimate the speed of access of that knowledge.

3.4. Procedure and Apparatus

The study was carried out in accordance with the research ethics protocols at the University of Nottingham. The study was carried out in three sessions. In the first session (pre-testing session), participants signed the consent form upon arrival. Then, they took part in the binomial word order recognition task (pre-test). Subsequently, participants completed the two proficiency tests (LexTALE and the multiple-choice vocabulary test), and the language background questionnaire.

Participants took part in the second session 10–14 days after the pre-testing session, which involved the reading treatment and the immediate post-test. First, participants engaged in the self-paced reading task in which they were exposed to the target items embedded in passages. Each trial began by displaying rows of dashes and blank spaces across a monitor's screen. The dashes corresponded to all of the non-white-space characters in a paragraph. To reveal the first phrase region, participants pressed the spacebar, causing the dashes corresponding to this region to be replaced by text. Each subsequent press of the spacebar caused the just-read region to revert to dashes, while simultaneously revealing text of the next region. The binomials always occurred towards the centre of a line in a single region of text. Reading times were collected for each phrase region. Once participants reached the end of the paragraph, a simple yes–no question appeared on the screen to ensure that participants read for comprehension. An analysis of the questions revealed good comprehension of the text (native speakers = 93.0% accuracy; non-native speakers = 79.5% accuracy). Once the reading task was completed, an unannounced word order recognition task (e.g., *salt and pepper* vs. *pepper and salt*) was carried out (i.e., the immediate post-test). The post-test was identical to the pre-test. The whole session (i.e., reading treatment + immediate post-test) took around 40 min.

Once participants were finished, they were invited to participate in the delayed post-test session, which occurred exactly seven days later. In the delayed task, participants again engaged in the same word order recognition task (e.g., *salt and pepper* vs. *pepper and salt*).

3.5. Analysis

Data were trimmed by deleting data points that fell above or below 2.5 standard deviations for each condition in each language group separately. This led to a loss of 1.91% of the response time data for the word order recognition task and 3.08% of reading time data for the self-paced reading task.[3] A few data points from the self-paced reading were removed due to technical issues related to the recording of reading time for these trials (0.25% of the data).

The data were analysed using mixed-effects modelling in *R* version 3.6.1 (R Core Team 2019). For the accuracy data in the word order recognition task (a binary variable, correct = 1 or incorrect = 0), logistic mixed-effects models were fitted (Jaeger 2008), while for response time data, linear mixed-effects models were fitted using the lme4 package (Bates et al. 2014). The *p*-values were estimated using the lmerTest package (Kuznetsova et al. 2015), and interactions were inspected using the *emmeans* package (Lenth 2019).[4] *p*-values for all pairwise comparisons were adjusted using Bonferroni correction.

Response/reading time data, from both the word order recognition task and self-paced reading task, were log-transformed before the analysis to reduce skewness. All of the other continuous variables (e.g., proficiency scores and variables from norming data) were also log-transformed to ensure that variables were on the same scale. Items' frequencies were transformed using the Zipf scale (Van Heuven et al. 2014).

The final (reported) models were chosen based on likelihood ratio tests and AIC scores for model comparisons. The results of model comparisons are reported in Table S4 in the Supplementary Materials. Models always included random intercepts for subjects and for items. Random effect structures were not kept maximal because this may result in uninterpretable and overspecified models that are difficult to estimate (Matuschek et al. 2017).

Models always included the following main predictors: item type (English-only, Arabic-only, or congruent), participant group (native or non-native), session (pre-test, immediate post-test, or delayed test), and repetition (two or five).[5] The inclusion of interaction effects between the predictors was added when it improved the model fit. In addition, the following variables were added as covariates *only* when they improved the model fit: proficiency scores (LexTALE scores and multiple-choice vocabulary test scores), the length of the binomial phrase, frequency (individual word frequency and binomial phrase frequency), binomial forward/backward ratio (i.e., forward binomial frequency divided by the backward binomial frequency), pre-test scores (i.e., accuracy scores in pre-test to account for the effect of previous knowledge on reading times), and familiarity ratings from the norming study.[6]

The proficiency tests (the multiple-choice vocabulary test and the LexTALE) were strongly correlated with each other ($r = 0.76$, $p < 0.05$). The likelihood ratio tests showed that the multiple-choice vocabulary test made a greater contribution to the models than the LexTALE; thus, it was considered in the analyses. Since proficiency scores for the native speakers were at ceiling, the proficiency analysis was limited to the non-native speakers to avoid the possibility that the proficiency effect would be partialled out by the effect of language group.

4. Results

4.1. Analysis of Accuracy Data

Table 2 presents a summary of the native and non-natives speakers' accuracy in the different testing sessions across the item types. The mixed-effects logistic models for accuracy, summarised in Table 3, showed a significant improvement in both the immediate and delayed post-tests for the experimental items. The control items that only occurred in the tests but not the reading passages did not exhibit a significant improvement ($p = 0.319$), suggesting that learning did not occur from repeated testing. Therefore, an analysis of the control items was not considered further. The familiarity norming ratings provided by the native speaker group predicted accuracy, while those from the Arabic speakers did not.

Table 2. Mean accuracy (out of a maximum of 14) and standard deviation.

	Item Type							
	Arabic		**Congruent**		**English**		**Control**	
	M	***SD***	***M***	***SD***	***M***	***SD***	***M***	***SD***
Non-native speakers (L1 Arabic)								
Pre-test	8.11	2.10	10.80	1.80	8.74	2.06	9.26	1.68
Post-test	10.40	1.74	11.00	1.98	9.81	2.03	8.94	1.71
Delayed test	9.49	2.10	11.10	1.68	9.62	1.99	8.85	1.75
Native speakers (L1 English)								
Pre-test	8.11	1.29	13.20	1.12	13.30	1.05	11.70	1.25
Post-test	10.60	1.17	13.20	0.91	13.60	0.67	11.50	1.54
Delayed test	9.61	1.83	12.90	1.13	13.40	0.88	11.40	1.01

Note. The maximum score was 14.

Table 3. Model outcome for accuracy scores.

	Accuracy for Experimental Items			
Predictors	*Estimates*	*Std. Error*	*Statistic*	*p*
(Intercept)	−3.44	2.17	−1.58	0.113
Group [Natives]	0.02	0.14	0.15	0.878
Item Type [Congruent]	0.76	0.25	3.09	0.002
Item Type [English]	0.09	0.27	0.35	0.725
Session [Post-test]	0.85	0.11	7.98	<0.001
Session [Delayed test]	0.49	0.10	4.71	<0.001
Familiarity ratings for natives	0.32	0.13	2.52	0.012
Group [Natives] * Item Type [Congruent]	1.43	0.19	7.65	<0.001
Group [Natives] * Item Type [English]	2.58	0.22	11.87	<0.001
Item Type [Congruent] * Session [Post-test]	−0.80	0.17	−4.76	<0.001
Item Type [English] * Session [Post-test]	−0.46	0.16	−2.89	0.004
Item Type [Congruent] * Session [Delayed test]	−0.41	0.16	−2.49	0.013
Item Type [English] * Session [Delayed test]	−0.18	0.16	−1.15	0.249
Random Effects				
σ^2	3.29			
τ_{00}	$0.14_{subject}$ 0.27_{Item}			
ICC	0.11			

As reported in Table 3, there was a significant interaction between group and binomial type. A pairwise test revealed that the native speakers outperformed the non-native speakers on congruent (odds ratio = 0.23, $p < 0.001$) and English-only items (odds ratio = 0.07, $p < 0.001$), but no difference emerged between the groups in the Arabic-only items (odds ratio = 0.97, $p = 0.877$). An analysis of the interaction showed that while non-natives speakers exhibited no differences in accuracy across the binomial types, the native speakers were more accurate on congruent and English-only items than Arabic-only items. There was also a significant interaction between binomial type and session. Pairwise tests of this interaction showed that congruent items were the only item type that did not exhibit significant differences in accuracy across the three testing sessions (all p's > 0.05). There was a significant increase in accuracy for the Arabic-only items (for both groups) and English-only items (for non-native native speakers) between the pre-test and immediate post-test, as well as between the pre-test and the delayed test (p's < 0.05). There was no interaction between group and session, suggesting that both groups experienced the same improvement from the pre-tests to post-tests.

In another model the influence of proficiency on non-native speakers was considered. The analysis indicated that higher proficiency resulted in greater accuracy ($\beta = 0.46$, SE = 0.15, z = 3.17, $p < 0.001$). There was no interaction with testing session ($\chi^2 = 0.23$, $p = 0.891$) nor with binomial type ($\chi^2 = 0.68$, $p = 0.710$), indicating that there was an overall increase in accuracy with increased proficiency regardless of the time of the test or the binomial type.

To examine the effect of repetitions (two repetitions vs. five repetitions), we fitted a model for repetition for the immediate and delayed post-tests, where repetition was a factor. A summary of the descriptive statistics and model outputs are presented in the Supplementary Materials (see Table S5). In sum, the analysis indicated that the effect of the number of repetitions on accuracy was not significant ($\beta = 0.08$, $SE = 0.07$, $z = 1.13$, $p = 0.256$), and that there was no interaction with the testing session (immediate or delayed) nor with item type.

4.2. Analysis of Response Time Data

Analyses of response time data were carried out on the correct responses. Table 4 presents the descriptive statistics of native and non-native speakers' response times across the three testing sessions for the binomial types. A linear mixed-effects model was fitted to compare differences in response time for experimental and control items (see Table S6 in the Supplementary Materials). The model revealed that there was a decrease in response times over time (i.e., pre-test vs. immediate post-test; immediate post-test vs. delayed test), and that there was a significant interaction between testing session and binomial type, indicating that the decrease in response times was larger for the experimental items than for the control items. This suggests that the reading treatment indeed affected the processing speed, and this effect was beyond the effect of repetition (i.e., practice effects).

Table 4. Mean RTs (in ms) and standard deviation from the word order recognition task.

	Item Type							
	Arabic		Congruent		English		Control	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Non-native speakers (L1 Arabic)								
Pre-test	3787	1569	3108	1459	3640	1470	3758	1582
Post-test	2565	1048	2207	949	2627	1037	3064	1273
Delayed test	2541	1022	2076	972	2536	1097	2612	1139
Native speakers (L1 English)								
Pre-test	1889	813	1431	574	1375	487	1741	832
Post-test	1564	750	1193	403	1124	317	1387	503
Delayed test	1382	551	1117	356	1075	302	1267	474

Two additional linear mixed-effects models were fitted to explore the effect of reading treatment on the response times for the experimental items. As can be seen in Table 5, native speakers required significantly less time to respond to binomial phrases in the post-tests than the non-native speakers. There was a significant effect of testing session, with both the immediate and delayed post-tests eliciting significantly less time than the pre-test. Pairwise comparisons indicated that the delayed test response times were significantly shorter than immediate post-test response times. The significant interaction between group and testing session (see Figure S1 in the Supplementary Materials) indicates that while the testing session had an effect on reducing both groups' response times, the magnitude of this effect was different across the groups. In particular, the difference in response times between the pre-test and post-test was larger for non-native speakers than for native speakers. The significant interaction between group and binomial type (see Figure S2 in the Supplementary Materials) indicates that the effect of binomial type on response time was different for the two groups. For the non-native group, pairwise comparisons show that while congruent items elicited shorter response times than either English-only ($p < 0.001$) or Arabic-only items ($p < 0.001$), both English-only and Arabic-only items had equivalent response times ($p < 0.001$). For the native speaker group, there was no difference in response times for the congruent and English-only items ($p = 0.760$), but Arabic-only items elicited longer response times than both congruent and English-only items (p's < 0.001). Length was a significant predictor, with shorter phrases being recognized faster.

In order examine the effect of proficiency on response time for the non-native speaker group, proficiency was included in a model of their data only. Proficiency scores significantly predicted response times, with higher proficiency scores eliciting faster response times ($\beta = -0.18$, $SE = 0.07$, $t = -2.46$, $p = 0.017$). There was a significant interaction between proficiency and testing session ($\chi^2 = 20.24$, $p < 0.001$), indicating that higher proficiency yielded shorter response times in the delayed test. However, proficiency did not have an effect on the immediate post-test.

Table 5. Model outcome for RTs from the word order recognition task.

	RTs for Experimental Items			
Predictors	***Estimates***	***Std. Error***	***Statistic***	***p***
(Intercept)	7.19	0.17	41.71	<0.001
Group [Natives]	−0.71	0.06	−11.64	<0.001
Session [Post-test]	−0.33	0.01	−25.75	<0.001
Session [Delayed test]	−0.39	0.01	−30.06	<0.001
Item Type [Congruent]	−0.17	0.03	−5.11	<0.001
Item Type [English]	−0.00	0.03	−0.10	0.922
Length (log)	0.37	0.06	5.82	<0.001
Group [Natives] * Session [Post-test]	0.15	0.02	6.55	<0.001
Group [Natives] * Session [Delayed test]	0.15	0.02	6.64	<0.001
Group [Natives] * Item Type [Congruent]	−0.08	0.02	−3.42	0.001
Group [Natives] * Item Type [English]	−0.28	0.02	−11.96	<0.001
σ^2	0.11			
τ_{00}	0.04_{subject} 0.01_{Item}			
ICC	0.30			

The influence of repetition (two vs. five) on response time was considered in a model of the immediate and delayed post-test data (see Table S7 in the Supplementary Materials). There was no effect of repetition ($\beta = -0.01$, $t = -1.34$, $p = 0.179$) nor any interactions with repetition.

4.3. Analysis of Self-Paced Reading Times

The final analyses considered the reading time for the self-paced reading task. As can be seen in Table 6 and the model summary in Table 7, native speakers' reading times were faster than the non-natives'. There was a significant speed up across the repetitions. Trial number was significant, with all binomial types eliciting faster reading times as the experiment progressed. Further, shorter words were read faster. Pre-test performance was not significant, suggesting that previous knowledge of the binomials did not affect reading times.

The main variables were involved in interactions. There was a significant interaction between group and item type (see Figure S3 in the Supplementary Materials). Pairwise comparisons indicated that native speakers read all binomial types at the same speed (p's > 0.05), but non-native speakers read congruent items significantly faster than English-only items ($p = 0.004$). For non-native speakers, no difference was found between Arabic-only items and congruent items ($p = 0.052$) or between Arabic-only and English-only items ($p = 1.000$). There was a significant interaction between the number of repetitions (1, 2, 3, 4, and 5) and group (see Figure S4 in the Supplementary Materials). An analysis of the interaction showed that the difference in reading times between number of repetitions was larger for the non-native speakers than the native speakers. For non-native speakers, each additional exposure resulted in significantly faster reading times. However, for native speakers, it was only the second exposure that demonstrated a speed up (second exposure faster than the first). Lastly, there was a significant interaction between the number of repetitions and item type. Pairwise comparisons showed that Arabic-only phrases had a much steeper decrease in reading times with increased repetitions than the other binomial types (i.e., for Arabic-only phrases, each repetition resulted in a significant decrease in reading times).

Table 6. Mean self-paced RTs (in ms) across repetitions (Rep) for each item type.

Item Type	Rep 1	Rep 2	Rep 3	Rep 4	Rep 5
Non-native speakers (L1 Arabic)					
Arabic-only	1687	1240	1168	1130	1002
Congruent	1407	1167	1079	1019	946
English-Only	1658	1293	1092	1127	1055
Native speakers (L1 English)					
Arabic-only	618	552	565	504	490
Congruent	532	498	515	463	458
English-only	602	500	491	486	487

Table 7. Model outcome for reading times from self-paced reading task.

	RTs			
Predictors	*Estimates*	*Std. Error*	*Statistic*	*p*
(Intercept)	6.43	0.16	39.50	<0.001
Group [Natives]	−0.99	0.06	−15.31	<0.001
Item Type [Congruent]	−0.12	0.03	−3.65	<0.001
Item Type [English]	0.02	0.03	0.69	0.492
Reps [2]	−0.26	0.02	−16.07	<0.001
Reps [3]	−0.33	0.02	−16.35	<0.001
Reps [4]	−0.37	0.02	−18.72	<0.001
Reps [5]	−0.47	0.02	−23.72	<0.001
Length (log)	0.51	0.06	8.61	<0.001
Trial number (log)	−0.07	0.00	−19.99	<0.001
Group [Natives] * Item Type [Congruent]	0.03	0.02	1.39	0.165
Group [Natives] * Item Type [English]	−0.06	0.02	−3.01	0.003
Group [Natives] * Reps [2]	0.15	0.02	7.85	<0.001
Group [Natives] * Reps [3]	0.23	0.02	9.73	<0.001
Group [Natives] * Reps [4]	0.21	0.02	8.73	<0.001
Group [Natives] * Reps [5]	0.27	0.02	11.13	<0.001
Item Type [Congruent] * Reps [2]	0.07	0.02	3.02	0.003
Item Type [English] * Reps [2]	0.01	0.02	0.37	0.714
Item Type [Congruent] * Reps [3]	0.07	0.03	2.46	0.014
Item Type [English] * Reps [3]	−0.06	0.03	−2.38	0.017
Item Type [Congruent] * Reps [4]	0.04	0.03	1.56	0.119
Item Type [English] * Reps [4]	0.00	0.03	0.15	0.877
Item Type [Congruent] * Reps [5]	0.08	0.03	3.09	0.002
Item Type [English] * Reps [5]	0.06	0.03	2.35	0.019
Random Effects				
σ^2	0.10			
$\tau_{00subject}$	0.05			
τ_{00Item}	0.01			
ICC	0.35			

5. Discussion

The goal of this study was to assess whether learners can acquire the word order preference of binomial phrases (e.g., it is *fish and chips* and not *chips and fish*) from reading. To explore this question, participants read English passages in which the target binomial phrases were embedded and repeated either two or five times. Participants encountered three types of binomials: English-only, translated Arabic-only, and congruent. A pre-test was administered two weeks before the main task to establish baseline knowledge. After the reading treatment, in which reading times were measured, the task was administered again (immediate post-test), and participants returned seven days later to take it a final time (delayed test). The tests assessed participants' recognition of the "correct" order of binomial phrases (*salt and pepper* vs. *pepper and salt*), recording their reaction times and accuracy.

5.1. Learning of Binomials

Accuracy and response time in the word order recognition task indicated that binomial phrases can be learned by both native and non-native speakers when they are exposed to them in a reading task. Native speakers had better performance in the immediate post-test than the pre-test for Arabic-only items, and non-native speakers had better performance in the post-test for English-only and Arabic-only items. Participants were already relatively familiar with the congruent items as demonstrated in the pre-test, which likely accounts for the small improvement observed for the congruent items. Seven days after the reading treatment, there was no significant decrease in accuracy, which demonstrates that the learning gains were durable. Overall, the results demonstrate that knowledge about word order preferences for binomials can be acquired during reading.

While the learning in this study can be described as incidental because participants were not instructed to learn the binomials nor were they forewarned about the post-tests, intentional learning cannot be ruled out. Taking part in the pre-test two weeks previously might have drawn participants' attention to the binomials, as could repetition of the items in the passages. Thus, the current study may be better characterized as semi-incidental (Pellicer-Sánchez and Boers 2019) because the target items could have been salient and garnered increased attention due to their prior exposure and repetitions.

The finding for binomials expands upon those of Sonbul and Schmitt (2013) and Webb et al. (2013) for collocations, demonstrating that other types of formulaic sequences can be learned from reading. The present study contributes to the limited literature on binomials. It aligns with Alotaibi et al.'s (2022) study, which showed that reading-only exposures yielded learning gains for binomials. The present study expands their findings by demonstrating the effect of congruency on binomial learning and processing. Overall, the ability to recognize the binomial order preference after having encountered binomial phrases only twice in one particular configuration supports usage-based theories (e.g., Bybee 1998; Goldberg 2006; Tomasello 2003), which highlight the role of language input in learning.

The effects of proficiency observed for the non-native speakers, such that increased proficiency yielded overall better and faster recognition performance on both pre-test and post-test, highlight the role of language use. These patterns of results support the view that with increased proficiency, L2 learners experience more phrasal frequency effects (e.g., Fernández and Schmitt 2015; Siyanova-Chanturia et al. 2011; Wolter and Gyllstad 2013). It suggests a progression towards a more native-like processing for formulaic language with increased exposure. In this study, the effect of proficiency was stronger on the delayed test than the immediate post-test for response time, indicating that recent/immediate exposure minimizes the gap in performance between lower and higher proficiency groups.

5.2. Repetition Effects

Both native and non-native speakers had durable learning gains after only two exposures. However, the study did not find an effect of repetition; increasing the number of occurrences from two to five did not yield better recognition, either in terms of accuracy or response times for either group. This finding conflicts with that of Webb et al.'s (2013) study which indicated that a greater number of exposures results in more collocational gains. The findings are consistent with the Alotaibi et al. (2022) study, suggesting no effect of the number of exposures (i.e., 2, 4, 5 or 6 occurrences) on recognition of the preferred word order of binomials for non-native speakers. They are also consistent with the Sonbul et al. (2023) study, showing that more exposures (i.e., two vs. four) did not lead to a processing advantage of forward forms of novel binomials over their reversed forms for non-native speakers. Further, the study aligns with Pellicer-Sánchez (2017), who showed that the frequency of exposure (i.e., 4 to 8) did not affect the acquisition of collocational knowledge. Pellicer-Sánchez (2017) speculated that a reading-only treatment might require a greater range of occurrences in order for frequency effects to emerge. That is, the limited range of repetitions in this study (i.e., two vs. five), as well as in Alotaibi et al. and Sonbul et al.'s studies, might have been too small for frequency effects to emerge.

The lack of a repetition effect could also be due to all of the repetitions of an item occurring in a single paragraph. Indeed, after the initial trials, it is likely that participants understood that items would reoccur. This could lead to less attention to the items, eliminating any advantage due to repetitions. It might be that the spacing of repetitions (i.e., occurrences distributed randomly across texts) would show frequency effects. Research on the effect of the spacing of repetitions on learning is still lacking (Webb 2014).

Another possibility that might account for the lack of a repetition effect relates to the type of knowledge assessed (i.e., the form recognition of word order). As indicated by Pellicer-Sánchez and Schmitt (2010), different aspects of word knowledge may require a different number of encounters. Notably, in their metanalysis on the effect of frequency of occurrence on vocabulary learning, Uchihara et al. (2019) found that frequency of occurrence was more indicative of performance in form recall than in form recognition. Since recognition knowledge is easier to acquire than recall knowledge (e.g., Laufer and Goldstein 2004; Webb et al. 2013), increased occurrences may not be needed for better recognition of an item, but they could benefit the recall knowledge of that item. Thus, more demanding tasks (i.e., recall) may show clearer frequency effects than less demanding ones (i.e., form recognition). The findings of Szudarski and Carter (2016) support such a view. They found that an increase in encounters (from 6 to 12) improved the form recall of collocations, while it did not reflect better performance on form recognition. The current study involved a word order recognition task (e.g., is it *fish and chips* or *chips and fish*), which is an aspect of form recognition, and is arguably less demanding than typical form recall formats, possibly masking any repetition effects.

Finally, it is important to acknowledge that the role of frequency is very complex, and that there are many methodological variables (e.g., learners' individual differences and items' distinctive characteristics) that could influence the relationship between repetition and vocabulary learning (Uchihara et al. 2019). Differences in such variables may account for the inconsistencies in the literature, which highlights the need for further research to disentangle them.

5.3. Congruency Effects

The study showed that when a binomial phrase overlapped between languages (i.e., congruent), it was responded to more quickly and accurately. More specifically, non-native speakers' response times for congruent phrases were faster than for English-only (recognition task and self-paced reading task) and Arabic-only items (recognition task). Further, the reading treatment had less benefit for congruent items, which were already better known at pre-test, than Arabic-only and English-only items for non-native speakers. Congruent items were also read faster in the self-paced reading task. Thus, the congruent items had less scope for improvement in a reading task with only a few exposures. Reading exposures led to more gains for the English-only (70.06%) and Arabic-only items (74.62%), showing that non-native speakers reached an accuracy level similar to the baseline of the congruent items (77.36%). A similar pattern was found in native speakers' (i.e., English-only and congruent items did not improve with the reading treatment; only Arabic-only items did). It is important to note that performance on the pre-test highlights that comparisons across the conditions on the post-tests are challenging because there was a higher baseline for congruent items.

Overall, there was an advantage in recognizing the specific order preference of the congruent items. This is compatible with other studies which found confirmatory evidence for the role of congruency on processing or learning. For example, Du et al. (2021) found a processing advantage for congruent over English-only and Chinese-only binomials for Chinese–English bilinguals. The current findings support evidence from previous research on collocations. For example, Wolter and Gyllstad's (2011, 2013) studies showed a processing advantage for congruent items over incongruent L2-only items, and Yamashita and Jiang (2010) demonstrated increased accuracy for L1-L2 collocations over L2-only collocations.

Notably, there was a clear influence of L1 knowledge on non-native speakers; however, this influence was limited to binomial phrases that were the same in L1 and L2, and not ones that had been translated from the L1 (Arabic-only items). Thus, L1 knowledge from translated L1 binomial phrases that are not part of the L2 lexicon behaved differently to congruent binomials, supporting the recent findings of Du et al. (2021). Non-native speakers' performance on the Arabic-only items was unexpected, generally not showing a difference to the English-only items and their accuracy not differing to that of the native English speakers. These findings contrast to the norming data (i.e., offline task), showing that non-native speakers were more accurate on Arabic-only than English-only items. One explanation is related to the nature of the task in the experiment. Participants were asked to choose which form of a binomial phrase sounded more natural in English (the L2), and they were instructed to respond as quickly as possible. The explicit instruction to focus on the L2 and the time pressure may have mitigated the influence from non-target language (the L1).

The current findings contrast with those of Carrol and Conklin (2014, 2017), who showed an L1 facilitative effect when non-native Chinese speakers of English read Chinese-only idioms translated into English. Similarly, Carrol et al. (2016) found that while native English speakers read translated Swedish-only idioms with more difficulty, Swedish participants did not experience that disruption when reading the translated Swedish-only idioms. The current study is incompatible with the cross-language priming account speculated by Carrol et al. (2016). This account predicts that when L2 input is encountered (e.g., the Arabic-only phrase *coffee and dates*), activation is automatically sent to L1 translation equivalents. When activation overlaps with an L1 configuration (i.e., binomial), a processing advantage emerges, which should benefit both processing and accuracy. However, this pattern did not emerge in the current study, as an advantage was only found when the phrase overlapped in both L1 and L2 (i.e., congruent but not Arabic-only condition). It might be that cross-language activation and the priming of L1 forms during L2 processing is more apparent for opaque and non-compositional formulaic language like idioms than for transparent and compositional formulaic language like the binomials in this study.

Alternatively, Wolter and Yamashita (2017) explain congruency effects with the mapping hypothesis (Ellis and Lambon Ralph 2000). In this view, the L1 influences the L2 only when there is L1-L2 correspondence. This assumes that in the initial stages of L2 learning both L1-only and L1-L2 formulaic language are equally accessible to L2 learners (i.e., they are processed similarly). However, as L2 learners become more proficient, they should encounter L1-L2 items, not L1-only items, in the L2 input, thus providing more reinforcement for L1-L2 items and resulting in further entrenching them in memory. Under this account, due to the lack of exposure and reinforcement in the L2, L1-only items gradually become "a less prominent part of the network" until they eventually no longer trigger any collocational associations in the L2 lexical network (Wolter and Yamashita 2017, p. 16).

In sum, the current findings expand congruency effects to binomials as L2 learners had better word order recognition performance on congruent items (faster and more accurate) than incongruent ones (whether Arabic-only or English-only). Because congruent items are encountered more across the L1 and L2, they should be more strongly entrenched in memory, thereby improving their performance.

5.4. Limitations and Future Directions

While the current study explores the interaction between frequency of exposure and congruency on binomials' learning and processing, there are some limitations to bear in mind. Although using existing binomial phrases provides more ecological validity, participants' prior knowledge needed to be assessed (i.e., the pre-test). Efforts were made to eliminate an influence of the pre-test on the treatment by separating the two (i.e., they were 10–14 days apart). However, pre-testing could have alerted the participants to the purpose of the study and the items that were under investigation. A further issue is related to the baseline knowledge indicated by the pre-test (i.e., performance on the congruent condition was high already at the pre-test). This makes interpreting the results challenging with regard to input type. Second, the study used a self-paced reading task to measure online reading, which has been used widely within L2 research; however, the task is limited in its ecological validity (e.g., it does not provide a natural reading experience).[7] Another limitation is that the study only assessed word order recognition (an aspect of form recognition); future studies should investigate both receptive and productive knowledge. It might be that the number of repetitions plays a different role in recognition and recall tasks (Pellicer-Sánchez and Schmitt 2010; Uchihara et al. 2019). In addition, the current study focused on massed input. It would be important to consider the impact of spacing. Finally, it is unclear how incidental versus intentional tasks might impact learning of an aspect of formulaic language knowledge like word order (i.e., *salt and pepper* not *pepper and salt*). To better inform pedagogical practice, future studies need to evaluate the effect of different learning conditions for learning binomials.

6. Conclusions

The present study adds to our understanding by examining the effect of congruency and repetition on the learning and processing of a type of formulaic language that is transparent/compositional (i.e., binomial phrases), while most research has investigated opaque/non-compositional formulaic language. Results showed that reading exposure improved knowledge of the preferred binomial word order, but that knowledge was not enhanced by repetition. Response time and accuracy data suggest that congruency plays an influential role; the recognition of binomial phrases was only facilitated when the conventionalised word order of the phrases in the L1 and L2 was the same, which points to the key role of congruency in L2 formulaic language learning and processing. Thus, it can be concluded that incongruent formulaic sequences should receive more attention in L2 teaching than congruent ones because of the difficulty associated with learning them.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/languages9010009/s1, Figure S1: Interaction between Group and Session; Figure S2: Interaction between Group and Item; Figure S3: Interaction between Group and Item Type in the reading task; Figure S4: Interaction between Group and number of repetitions in the reading task; Table S1: The target binomial phrases selected in the study; Table S2: Average correct responses in the norming studies across the item types for NSs and NNSs; Table S3: Example for the item gold and silver when presented two and five times in the reading passages; Table S4: Re-sults of model comparisons; Table S5: Analysis of accuracy as a function of repetition; Table S6: Analysis of response time as a function of test session for experimental and control items; Table S7: Analysis of response time as a function of repetition.

Author Contributions: Conceptualization, A.A. and K.C.; methodology, A.A. and K.C.; software, A.A.; formal analysis, A.A.; investigation, A.A.; data curation, A.A.; writing—original draft preparation, A.A.; writing—review and editing, K.C.; supervision, K.C. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: The study was conducted in accordance with the Declaration of Helsinki, and approved by the Institutional Review Board (or Ethics Committee) of University of Nottingham on 18/12/2017.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The data presented in this study are openly available at https://osf.io/3kjn6/?view_only=a644834978744d9fb20815d0477e54dd (accessed on 30 November 2023).

Conflicts of Interest: The authors declare no conflict of interest.

Notes

1 An example of these constraints is markedness, where less marked words come before more marked ones (e.g., *good and bad*). Another example is frequency, where more frequent words are ordered before less frequent ones. However, there are exceptions for each constraint (e.g., *bride and groom* not conforming with the constraint "male before female").

2 While the frequency analysis and norming confirmed the categorisation of the items (English-only, congruent, Arabic-only), the Arabic binomials *heaven and hell*, *sun and moon*, *land and sea*, and *sunrise and sunset* might be familiar to native English speakers. Analyses were rerun with these items removed. Crucially, the pattern of results remained the same (i.e., significance level remained the same). Thus, we retained these items in the reported analyses.

3 Data trimming did not result in losing data unevenly from the conditions and groups.

4 VIF values were < 2.0 in all models, indicating that the multicollinearity assumption was not violated.

5 The repetition variable was only relevant after the reading treatment and, therefore, was included in analyses of the immediate and delayed post-tests.

6 The norming data from the forward and backward completion tasks were not included as covariates because (1) their addition did not make an improvement to the models, and (2) they were highly correlated with each other (all r's > 0.57 and all p's < 0.05).

7 For example, self-paced reading tasks require participants to engage in a secondary task (e.g., pressing a key), and in the present study, it did not allow for multiple readings of an item.

References

Alotaibi, Sara, Ana Pellicer-Sánchez, and Kathy Conklin. 2022. The effect of input modes and number of exposures on the learning of L2 binomials. *ITL-International Journal of Applied Linguistics* 173: 58–93. [CrossRef]

Al-Thubaity, Abdulmohsen O. 2015. A 700M + Arabic corpus: KACST Arabic corpus design and construction. *Language Resources and Evaluation* 49: 721–51. [CrossRef]

Bates, Douglas, Martin Maechler, Ben Bolker, and Steven Walker. 2014. lme4: Linear mixed-effects models using Eigen and S4. *R Package Version* 1: 1–23.

Beck, Sara D., and Andrea Weber. 2016. Bilingual and monolingual idiom processing is cut from the same cloth: The role of the L1 in literal and figurative meaning activation. *Frontiers in Psychology* 7: 1350. [CrossRef] [PubMed]

Benor, Sarah Bunin, and Roger Levy. 2006. The chicken or the egg? A probabilistic analysis of English binomials. *Language* 82: 233–78. [CrossRef]

Bybee, Joan. 1998. The emergent lexicon. *Chicago Linguistic Society* 34: 421–35.

Carrol, Gareth, and Kathy Conklin. 2014. Getting your wires crossed: Evidence for fast processing of L1 idioms in an L2. *Bilingualism: Language and Cognition* 17: 784–97. [CrossRef]

Carrol, Gareth, and Kathy Conklin. 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. [CrossRef]

Carrol, Gareth, Kathy Conklin, and Henrik Gyllstad. 2016. Found in translation: The influence of L1 on the processing of idioms in L2. *Studies in Second Language Acquisition* 38: 403–43. [CrossRef]

Cobb, Tom. n.d. Compleat Lexical Tutor (Version 2) [Online Application Software]. Available online: https://www.lextutor.ca (accessed on 30 November 2023).

Conklin, Kathy, and Gareth Carrol. 2019. First language influence on the processing of formulaic language in a second language. In *Understanding Formulaic Language: A Second Language Acquisition Perspective*. Edited by A. Siyanova-Chanturia and A. Pellicer-Sánchez. London: Routledge, pp. 62–77.

Conklin, Kathy, and Gareth Carrol. 2021. Words go together like 'bread and butter': The rapid, automatic acquisition of lexical patterns. *Applied Linguistics* 42: 492–513. [CrossRef]

Du, Lingli, Irina Elgort, and Anna Siyanova-Chanturia. 2021. Cross-language influences in the processing of multiword expressions: From a first language to second and back. *Frontiers in Psychology* 12: 666520. [CrossRef] [PubMed]

Durrant, Philip, and Norbert Schmitt. 2010. Adult learners' retention of collocations from exposure. *Second Language Research* 26: 163–88. [CrossRef]

Ellis, Andrew W., and Matthew A. Lambon Ralph. 2000. Age of acquisition effects in adult lexical processing reflect loss of plasticity in maturing systems: Insights from connectionist networks. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 26: 1103–23. [CrossRef] [PubMed]

Erman, Britt, and Beatrice Warren. 2000. The idiom principle and the open choice principle. *Text-Interdisciplinary Journal for the Study of Discourse* 20: 29–62. [CrossRef]

Fernández, Beatriz González, and Norbert Schmitt. 2015. How much collocation knowledge do L2 learners have?: The effects of frequency and amount of exposure. *ITL-International Journal of Applied Linguistics* 166: 94–126. [CrossRef]

Goldberg, Adele. 2006. *Constructions at Work: The Nature of Generalization in Language*. Oxford: Oxford University Press.

Jaeger, T. Florian. 2008. Categorical Data Analysis: Away from ANOVAs (transformation or not) and towards Logit Mixed Models. *Journal of Memory and Language* 59: 434–46. [CrossRef] [PubMed]

Kuiper, Koenraad. 2004. Formulaic performance in conventionalised varieties of speech. In *Formulaic Sequences*. Edited by N. Schmitt. Amsterdam: John Benjamins, pp. 37–54.

Kuznetsova, Alexandra, Per B. Brockhoff, and Rune H. B. Christensen. 2015. lmerTest: Tests in Linear Mixed Effects Models. R Package Version 2.0-29. Available online: https://CRAN.Rproject.org/package5lmerTest (accessed on 3 March 2020).

Laufer, Batia, and Zahava Goldstein. 2004. Testing vocabulary knowledge: Size, strength, and computer adaptiveness. *Language Learning* 54: 399–436. [CrossRef]

Lemhöfer, Kristin, and Mirjam Broersma. 2012. Introducing LexTALE: A quick and valid Lexical Test for Advanced Learners of English. *Behavior Research Methods* 44: 325–43. [CrossRef]

Lenth, Russell V. 2019. Emmeans: Estimated Marginal Means, Aka Least-Squares Means. Available online: https://CRAN.R-project.org/package=emmeans (accessed on 3 March 2020).

Leśniewska, Justyna, and Ewa Witalisz. 2007. Cross-linguistic influence and acceptability judgments of L2 and L1 collocations: A study of advanced Polish learners of English. *EUROSLA Yearbook* 7: 27–48. [CrossRef]

Matuschek, Hannes, Reinhold Kliegl, Shravan Vasishth, Harald Baayen, and Douglas Bates. 2017. Balancing Type I error and power in linear mixed models. *Journal of Memory and Language* 94: 305–15. [CrossRef]

Millisecond Software. 2015. Inquisit 5. Available online: https://www.millisecond.com (accessed on 1 January 2019).

Nakata, Tatsuya, and Irina Elgort. 2021. Effects of spacing on contextual vocabulary learning: Spacing facilitates the acquisition of explicit, but not tacit, vocabulary knowledge. *Second Language Research* 37: 233–60. [CrossRef]

Nation, Paul, and D. Beglar. 2007. A vocabulary size test. *The Language Teacher* 31: 9–13.

Pawley, Andrew, and Frances Hodgetts Syder. 1983. Two puzzles for linguistic theory: Nativelike selection and nativelike fluency. *Language and Communication* 191: 225.

Pellicer-Sánchez, Ana. 2017. Learning L2 collocations incidentally from reading. *Language Teaching Research* 21: 381–402. [CrossRef]

Pellicer-Sánchez, Ana, and Frank Boers. 2019. Pedagogical perspectives to the teaching and learning of formulaic language. In *Understanding Formulaic Language: A Second Language Acquisition Perspective*. Edited by A. Siyanova-Chanturia and A. Pellicer-Sánchez. New York: Routledge, pp. 153–73.

Pellicer-Sánchez, Ana, and Norbert Schmitt. 2010. Incidental vocabulary acquisition from an authentic novel: Do things fall apart? *Reading in a Foreign Language* 22: 31–55.

Peters, Elke. 2016. The learning burden of collocations: The role of interlexical and intralexical factors. *Language Teaching Research* 20: 113–38. [CrossRef]

Pritchett, Lena K., Jyotsna Vaid, and Sumeyra Tosun. 2016. Of black sheep and white crows: Extending the bilingual dual coding theory to memory for idioms. *Cogent Psychology* 3: 1135512. [CrossRef]

R Core Team. 2019. *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing.

Siyanova-Chanturia, Anna, Kathy Conklin, and Walter J. B. Van Heuven. 2011. Seeing a phrase "time and again" matters: The role of phrasal frequency in the processing of multiword sequences. *Journal of Experimental Psychology: Language, Memory and Cognition* 37: 776–84. [CrossRef]

Siyanova-Chanturia, Anna, Kathy Conklin, Sendy Caffarra, Edith Kaan, and Walter J. B. van Heuven. 2017. Representation and processing of multi-word expressions in the brain. *Brain and Language* 175: 111–22. [CrossRef]

Sonbul, Suhad, and Norbert Schmitt. 2013. Explicit and implicit lexical knowledge: Acquisition of collocations under different input conditions. *Language Learning* 63: 121–59. [CrossRef]

Sonbul, Suhad, Dina Abdel Salam El-Dakhs, Kathy Conklin, and Gareth Carrol. 2023. "Bread and butter" or "butter and bread"? Nonnatives' processing of novel lexical patterns in context. *Studies in Second Language Acquisition* 45: 370–92. [CrossRef]

Szudarski, Paweł, and Ronald Carter. 2016. The role of input flood and input enhancement in EFL learners' acquisition of collocations. *International Journal of Applied Linguistics* 26: 245–65. [CrossRef]

Tekmen, E. Anne Ferrell, and Ayşegül Daloglu. 2006. An investigation of incidental vocabulary acquisition in relation to learner proficiency level and word frequency. *Foreign Language Annals* 39: 220–43. [CrossRef]

Titone, Debra, Georgie Columbus, Veronica Whitford, Julie Mercier, and Maya Libben. 2015. Contrasting bilingual and monolingual idiom processing. In *Bilingual Figurative Language Processing*. Edited by Roberto R. Heredia and Anna B. Cieślicka. Cambridge: Cambridge University Press, pp. 171–207.

Tomasello, Michael. 2003. *Constructing a Language: A Usage-Based Theory of Language Acquisition*. Cambridge: Harvard University Press.

Uchihara, Takumi, Stuart Webb, and Akifumi Yanagisawa. 2019. The effects of repetition on incidental vocabulary learning: A meta-analysis of correlational studies. *Language Learning* 69: 559–99. [CrossRef]

Van Heuven, Walter J. B., Pawel Mandera, Emmanuel Keuleers, and Marc Brysbaert. 2014. Subtlex-UK: A new and improved word frequency database for British English. *The Quarterly Journal of Experimental Psychology* 67: 1176–90. [CrossRef]

Webb, Stuart. 2008. The effects of context on incidental vocabulary learning. *Reading in a Foreign Language* 20: 232–45.

Webb, Stuart. 2014. Repetition in incidental vocabulary learning. In *The Encyclopedia of Applied Linguistics*. Edited by Carol A. Chapelle. Hoboken: Wiley, pp. 1–6.

Webb, Stuart, Jonathan Newton, and Anna Chang. 2013. Incidental learning of collocation. *Language Learning* 63: 91–120. [CrossRef]

Wolter, Brent, and Henrik Gyllstad. 2011. Collocational Links in the L2 mental lexicon and the influence of L1 intralexical knowledge. *Applied Linguistics* 32: 430–49. [CrossRef]

Wolter, Brent, and Henrik Gyllstad. 2013. Frequency of input and L2 collocational processing: A comparison of congruent and incongruent collocations. *Studies in Second Language Acquisition* 35: 451–82. [CrossRef]

Wolter, Brent, and Junko Yamashita. 2017. Word frequency, collocational frequency, L1 congruency, and proficiency in L2 collocational processing: What accounts for L2 performance? *Studies on Second Language Acquisition* 40: 395–416. [CrossRef]

Yamashita, Junko, and N. A. N. Jiang. 2010. L1 influence on the acquisition of L2 collocations: Japanese ESL users and EFL learners acquiring English collocations. *TESOL Quarterly* 44: 647–68. [CrossRef]