Studies in Second Language Acquisition (2023), **45**, 370–392
doi:10.1017/S0272263122000237

RESEARCH ARTICLE

"Bread and butter" or "butter and bread"? Nonnatives' processing of novel lexical patterns in context

Suhad Sonbul[1*], Dina Abdel Salam El-Dakhs[2], Kathy Conklin[3] and Gareth Carrol[4]

[1]Umm Al-Qura University, Saudi Arabia; [2]Prince Sultan University, Saudi Arabia; [3]University of Nottingham, UK; [4]University of Birmingham, UK
*Corresponding author. E-mail: sssonbul@uqu.edu.sa

(Received 29 August 2021; Revised 20 May 2022; Accepted 27 May 2022)

Abstract

Little is known about how nonnative speakers process novel language patterns in the input they encounter. The present study examines whether nonnatives develop a sensitivity to novel binomials and their ordering preference from context. Thirty-nine nonnative speakers of English (L1 Arabic) read three short stories seeded with existing binomials (*black and white*) and novel ones (*bags and coats*) while their eye movements were monitored. The existing binomials appeared once in their forward (conventional) form and once in their reversed form. The novel binomials appeared in their experimentally defined forward form in different frequency conditions (two vs. four encounters) and once in the reversed form. Results showed no advantage for existing binomials over their reversed forms. For the novel binomials, the nonnative speakers read subsequent encounters significantly faster than initial ones for both frequency conditions. More importantly, the final reversed form also led to faster reading, suggesting that L2 speakers process the reversed form of a novel binomial as another encounter, ignoring the established order.

Introduction

English speakers say that things go together like *bread and butter*, not like *butter and bread.* Lexical patterns like these (often referred to as formulaic language or multiword sequences) account for up to half spoken discourse (Erman & Warren, 2000; Pawley & Syder, 1983). One example of such lexical patterns is binomials, defined as "coordinated word pairs whose lexical elements share the same word class" (Mollin, 2014, p. 1). Binomials abound in English (e.g., *aches and pains*, *fair and square*, *high and low*, *life and death*) and vary in their degree of reversibility along a cline (Malkiel, 1959). At one end of the cline are frozen binomials, which are very clearly irreversible (e.g., *hit and run*, *chalk and cheese*) because of their highly idiomatic meaning. However, the focus of the present study is not on such idiomatic binomials, but rather on semantically compositional (i.e., transparent) binomials. Transparent binomials exist along a continuum of fixedness, where components may have a preferred sequence even when the order could in theory be reversed without fundamentally changing the meaning (e.g., *public and private*, *mother and father*). Therefore, binomials have two important properties: co-occurrence restrictions (like other lexical patterns) and configuration restrictions, and this unique nature has led to interest in how they are processed and acquired.

One line of research has explored the factors that determine the preferred order of binomials, particularly in terms of diachronic changes (e.g., Goldberg & Lee, 2021; Mollin, 2014). Goldberg and Lee (2021) found that binomials like *uncles and aunts/ nephews and nieces*, which were in common use prior to the 1930s, have more recently reversed their preferred order to *aunts and uncles/nieces and nephews.* Goldberg and Lee proposed several cognitive explanations for the change, including the accessibility of the individual components of binomials in memory and their cluster strength. Others have attributed the ordering preferences in binomials to sociocultural factors (e.g., Mollin, 2013) and factors such as the semantic, phonological and lexical properties of component words, as well as how much experience an individual has with the binomial in question (Morgan & Levy, 2016).

Another research area is concerned with how binomials are acquired by native and nonnative speakers. Some studies have addressed this question using post-treatment tasks (Alotaibi et al., 2022) or using eye-tracking to examine the processing of novel binomials as it unfolds in real time (Alotaibi, 2020; Conklin & Carrol, 2021). The current study aims to contribute to research on the processing of novel binomials (i.e., infrequent phrases which do not have a conventionalized word order) by extending the work of Conklin and Carrol (2021), who examined the processing of novel binomials in native speakers, to a population of nonnative speakers of English. The study investigates whether non-native speakers show a processing sensitivity (i.e., speeded recognition) to novel binomials in a natural reading context, as was the case with native speakers in the original study. More specifically, we examine whether nonnative speakers can associate pairs of words in memory and register their preferred word order (e.g., *wires and pipes* instead of *pipes and wires*) over the course of reading short texts seeded with the novel binomials. Thus, the current study examines L2 processing of novel binomials rather than their acquisition (i.e., there is no post-treatment measure of knowledge).

The following sections will review two relevant strands of literature to situate the current study: the processing of lexical patterns by native and nonnative speakers and the acquisition of single-word vocabulary and lexical patterns by nonnative speakers.

Literature review

Processing of lexical patterns

It is well-established that native speakers recognize lexical patterns faster and process their phrase-level meaning more easily than other nonrecurrent combinations of words that do not show any significant degree of cohesion or fixedness. This has been shown for idioms (e.g., Carrol & Conklin, 2020; Conklin & Schmitt, 2008; Libben & Titone, 2008; Rommers et al., 2013), phrasal verbs (e.g., Blais & Gonnerman, 2013; Matlock & Heredia, 2002; Tiv et al., 2019) and binomials (e.g., Arcara et al., 2012; Carrol & Conklin, 2020). The processing advantage for lexical patterns by native speakers has prompted researchers to explore these patterns in nonnative speakers.

Several studies have explored the determinants of nonnative processing of lexical patterns. One important factor is first language (L1)–second language (L2) similarity.

https://doi.org/10.1017/S0272263122000237 Published online by Cambridge University Press

When single words are the same (or very similar) across languages (e.g., *piano* in both English and French), there is considerable evidence of cross-language activation, or a cognate effect, in nonnative lexical processing (for an overview, see van Hell & Tanner, 2012). More recently, researchers have started to examine the effect of L1-L2 similarity on the processing of lexical patterns. This has been referred to as the congruency effect (i.e., the availability of a literal translation equivalent). Studies have examined the congruency effect using different types of lexical patterns including collocations (e.g., Wolter & Gyllstad, 2011, 2013; Yamashita & Jiang, 2010) and idioms (e.g., Carrol & Conklin, 2014; Carrol et al., 2016; Irujo, 1986; Pritchett et al., 2016; Titone et al., 2015). The general finding is that congruency has a clear influence on the processing of lexical patterns in the L2 with an advantage for congruent items (L1 = L2) over incongruent items (L1 $\neq$ L2).[1]

Another important determinant of the processing of lexical patterns is transparency. One relevant study here is by Gyllstad and Wolter (2016), who employed a semantic judgment task to examine how advanced L1 Swedish–L2 English learners processed English free combinations and collocations. Reaction times and error rates showed a processing cost for collocations compared to free combinations, due to the semantically semitransparent nature of collocations. In the same vein, Yamashita (2018) explored the potential contribution of semantic transparency in explaining the congruency effect, and found that congruent items were dominated by transparency while incongruent items were generally characterized by opacity, indicating a clear overlap between these variables.

A third determinant of the processing of lexical patterns is frequency. A study by Sonbul (2015) explored the sensitivity of native and nonnative speakers of English to the corpus-derived frequency of collocations using both off-line (typicality rating task) and online (eye movements) measures. There was a clear sensitivity to corpus-derived frequency among both natives and nonnatives in the off-line task. The frequency effect was also notable in the early stages of reading but disappeared later for both groups. Wolter and Gyllstad (2013) examined the influence of frequency on the processing of congruent and incongruent collocations. They found that advanced Swedish learners of English were highly sensitive to the frequency of collocations, regardless of whether or not the collocations had a congruent form in the L1. Likewise, Wolter and Yamashita (2018) used an acceptability judgment task to examine the processing of collocations by intermediate and advanced Japanese speakers of English and native English speakers. They found effects of both word-level and collocation-level frequency among the three groups of participants. Such results support usage-based models of language acquisition, whereby experience with the language predicts language processing and acquisition (e.g., Bybee, 2006; Ellis, 2002).

In addition, L2 proficiency seems to influence the processing of lexical patterns. For example, Sonbul (2015) found that the effect of corpus-derived frequency on the processing of collocations was greater among nonnative speakers of English as their proficiency increased. Similarly, Ding and Reynolds (2019) found that the influence of congruency was clearer among highly proficient than less proficient Chinese EFL (English as a foreign language) learners. Sonbul and El-Dakhs (2020) showed that the estimated proficiency of Arab learners of English influenced the processing of collocations, with congruency effects slowly diminishing as

[1]It should be noted that congruency is not a main factor in the present study but, due to its central role in processing lexical patterns, it will be considered in item development and data analysis.

https://doi.org/10.1017/S0272263122000237 Published online by Cambridge University Press

proficiency increased (and see similar effects of increasing proficiency from Yamashita & Jiang, 2010).

The research reviewed thus far has mainly focused on the processing of collocations and idioms. To the best of our knowledge, only one study has examined the processing of binomials by nonnatives. Siyanova-Chanturia et al. (2011) employed eye-tracking to examine how native and non-native English speakers, of varied levels of proficiency, processed binomials that differed in phrasal frequency. The participants read sentences containing binomials in their preferred, frequent, order (*bride and groom*) or their reversed, less frequent, form (*groom and bride*). The results showed that both natives and nonnatives were generally sensitive to the frequency of occurrence of binomials, but only natives and higher proficiency nonnatives also exhibited a sensitivity to the canonical configuration.

Incidental acquisition of L2 vocabulary

Most of the available evidence on the incidental acquisition of L2 vocabulary from context has focused on single words (e.g., Pellicer-Sánchez & Schmitt, 2010; Pitts et al., 1989; Waring & Takaki, 2003). These studies mainly relied on off-line tests to assess gains and showed that L2 speakers do retain vocabulary from exposure, but the rate might be relatively low. Only recently have studies used eye-tracking, which allows for the examination of on-line processing as it unfolds in real time. Among the earliest eye-tracking studies to examine the incidental acquisition of word knowledge is Godfroid et al. (2013). In their study, Godfroid et al. (2013) had advanced Dutch-speaking learners of English read short English extracts that contained target known words and unknown pseudowords. Their results showed that participants spent more time processing the unknown pseudowords than the known words, and that longer fixations to the pseudowords were associated with better scores in an unannounced vocabulary posttest. Similarly, Pellicer-Sánchez (2016) combined off-line (paper-and-pencil) and online (eye-tracking) measures to examine the incidental acquisition of unknown words by nonnative English learners. Notably, the participants read a full story, not simply short extracts. The story contained pseudowords, each repeated eight times. The reading time (RT) for pseudowords decreased significantly after three to four encounters, and pseudowords were read in a similar manner to known real words after eight encounters. The paper-and-pencil tests showed that incidental acquisition of unknown words is possible. However, the acquisition of word meaning lagged behind the acquisition of word form.

Godfroid et al. (2018) conducted a study that involved participants reading an authentic novel. The participants (native and nonnative English speakers) read five chapters of an English novel that included foreign Dari (Farsi) words ranging in frequency (1–23 occurrences). After reading, the participants were given a comprehension test and surprise vocabulary tests. Using growth curve analysis to model form knowledge development, the results showed that both the quantity (number of exposures) and the quality (total RT) of lexical processing facilitated incidental vocabulary acquisition. The results showed a nonlinear, S-shaped pattern of RTs for newly acquired words with an initial speed up (one to four exposures) followed by a plateau and a slight increasing trend (7 to 10 exposures) before further decreases in RTs (11 to 23 exposures). Posttest scores suggested that the frequency of occurrence of the new words and how long the participants read them at each encounter predicted the acquisition of knowledge.

https://doi.org/10.1017/S0272263122000237 Published online by Cambridge University Press

The research reviewed in this section thus far has focused on the acquisition of individual words. In contrast to single words, lexical patterns are often claimed to be less noticed by nonnatives (e.g., Boers & Lindstromberg, 2012; Christiansen & Arnon, 2017; Wray, 2000). However, some studies have suggested that nonnative speakers are able to notice lexical patterns in context. For example, Durrant and Schmitt (2010) assigned nonnative speakers of English to one of three training conditions: single exposure, verbatim repetition, and varied repetition, followed by a naming task. Participants in both repetition conditions recalled the target collocations better than those in the single exposure condition. The authors concluded that adult nonnative learners retain information about the lexical patterns they are exposed to in input, in line with usage-based models of language acquisition (Bybee, 2006).

Evidence for the incidental acquisition of lexical patterns has accrued over the past few years. Some studies have focused on incidental vocabulary acquisition from television/video viewing (e.g., Majuddin et al., 2021; Puimège & Peters, 2020). More relevant to the present study is research investigating the incidental acquisition of lexical patterns from a reading context (Pellicer-Sánchez, 2017; Webb et al., 2013). Webb et al. (2013) investigated the effect of repetition on the incidental acquisition of collocations. Taiwanese EFL learners simultaneously read and listened to one of four versions of a modified graded reader that included target collocations, with 1, 5, 10, and 15 encounters. Immediate-posttest results indicated that encountering collocations repeatedly when reading while listening contributed to incidental acquisition of form and meaning, with collocation acquisition increasing as a factor of frequency. Pellicer-Sánchez (2017) examined the incidental acquisition of collocational knowledge, focusing on adjective-pseudoword collocations in reading. L2 learners read a story seeded with target collocations that were repeated either four or eight times. The scores on a 1-week delayed posttest lent support to the benefits of collocation acquisition from a reading context, even suggesting that ESL learners can incidentally develop collocational knowledge at a similar rate to individual words. However, there was not a significant effect of the frequency manipulation on collocation acquisition.

While most studies have focused on collocations, very few studies have examined the incidental acquisition of binomials. In one such study, Alotaibi et al. (2022) investigated the effect of input mode (reading-only, listening-only, and reading-while-listening) and frequency of exposure (two, four, five, and six occurrences) on declarative binomial knowledge. Based on performance on immediate paper-and-pencil tests, results indicated that it was possible for nonnative learners of English (L1 Arabic) to develop declarative knowledge of the preferred order of binomials from the various input modes; novel binomials encountered six times showed similar familiarity ratings as existing binomials.

While previous studies indicate that lexical patterns (including binomials) can be acquired incidentally in L2 speakers, they are limited in that they used post-hoc measures of knowledge and did not examine processing as it unfolds in real time. Only a few studies have employed eye-tracking to examine the online processing of lexical patterns (Alotaibi, 2020, Study 2; Choi, 2017). The aim of Alotaibi's (2020) Study 2, for example, was to examine how nonnative learners process novel binomials in different input modes. The findings showed that repeated exposure to novel binomials led to fewer fixations and shorter RTs. Additionally, with increased exposure, the processing of novel binomials gradually became comparable to existing ones. It should be noted that since the focus of Alotaibi's (2020, Study 2) was on mode of processing rather than acquisition per se, she did not include a reversed form of the novel binomials. Including a reversed form can help address the special nature of binomials (see "Introduction") that does not merely involve co-occurrence restrictions but also entails word order preferences.

The present study

As noted earlier, the current study is an extension of Conklin and Carrol (2021), who investigated the processing of novel binomials amongst native speakers, exploring sensitivity to co-occurrence information and canonical word order. They monitored the eye movements of 40 native English speakers while reading short stories that contained existing binomials in their common forward form (e.g., *time and money*), seen once, and novel binomials (e.g., *wires and pipes*), seen one to five times in their experimentally defined forward form. Then, the readers saw the existing and novel patterns in the reversed order (e.g., *money and time*, *pipes and wires*). The results revealed an initial co-occurrence memory effect for the components of novel binomials (i.e., "wires" and "pipes") regardless of direction whereby the last "reversed" form was processed similar to or even significantly faster than the first forward occurrence. However, when frequency of encounter was considered, an advantage emerged for forward novel patterns over subsequently encountered reversed forms after four to five exposures, suggesting that natives could develop a sensitivity to the order of novel binomials rapidly from exposure.

The current study aims to examine whether the effect found for natives by Conklin and Carrol (2021) emerges for nonnative speakers. More specifically, we are interested in whether nonnative speakers rapidly develop a sensitivity to the preferred word order of novel binomials through natural reading. The current study addresses the following questions:

1. Is the language processing system sensitive to novel binomials in L2 input that simulates a real-world context?
2. What is the effect of frequency of exposure on nonnatives' sensitivity to novel binomials in a real-world context?

Similar to Conklin and Carrol (2021), the current study presented existing binomials only once in their forward form followed by once in their reversed form. However, unlike Conklin and Carrol (2021) who included five frequency levels for novel binomials, the design of the present study included two frequency categories only (2-repetition vs. 4-repetition) to increase item power. In Conklin and Carrol (2021) as the number of repetitions increased, the number of items per frequency level decreased. For example, 25 items were read once, but only five items were read five times. Thus, there was much less item power at the fifth occurrence versus the first. By only looking at two frequency levels, we were able to include the same number of items for both categories.[2] We selected two versus four repetitions for our frequency categories based on Conklin and Carrol's (2021) finding that natives showed a clear sensitivity to a given configuration after four exposures but not after two exposures. Thus, the novel

[2]Another option would have been to follow Conklin and Carrol's (2021) design with five frequency levels, but just increase the number of items at each level. Doing this would have increased item power but would have also resulted in passages becoming saturated with binomials, making the repetition manipulation clearly marked.

https://doi.org/10.1017/S0272263122000237 Published online by Cambridge University Press

binomials in the current study involved two main factors. The first factor was *Category* with two levels: 2-repetition versus 4-repetition, and the second factor was *Iteration* with three levels: first, last (i.e., second occurrence for 2-repetition items and fourth occurrence for 4-repetition items), and reversed.

To evaluate an emerging sensitivity to novel binomials (RQ1), we will compare RTs of existing and novel binomials in the forward (first) and reversed iterations only. When examining the effect of frequency on sensitivity to novel binomials (RQ2), we will include the first, last, and reversed forms of both frequency categories.

To establish any similarities or differences between native and nonnative processing of novel binomials, the current findings from nonnative speakers will be discussed in relation to those of the native speakers from Conklin and Carrol (2021). It is important to point out, that while we have introduced certain changes to the design (see preceding text), the experiments are very similar (i.e., use the same items and the same texts with minor modifications). Thus, comparing the findings of the current study to that of Conklin and Carrol (2021) should be justified.[3]

We do not expect nonnatives in the present study to necessarily have a sensitivity to the canonical order of existing binomials across all proficiency levels. Similar to Siyanova-Chanturia et al. (2011), we anticipate that sensitivity to binomials' word order should only emerge for participants with high L2 proficiency. For the novel binomials, based on Alotaibi (2020, Study 2), we expect our nonnatives to show online memory effects for the co-occurrence of novel binomials' components, that is, shorter RTs for the last over the first encounter in the forward form. However, nonnatives might or might not develop sensitivity to the canonical order of novel binomials (*wires and pipes* vs. *pipes and wires*). If nonnatives follow the same pattern as natives (Conklin & Carrol, 2021), they should show a processing advantage (i.e., shorter RTs) for novel binomials in the forward form over the backward form after four exposures, but likely not after as few as two encounters. If, however, nonnative speakers are not sensitive to word order differences in lexical patterns, our nonnative participants might not show such a processing advantage even after four exposures. Rather, they might treat the backward form as another co-occurrence of the components and overlook the direction preference. In that case, the final backward occurrence should demonstrate an additional processing advantage over the last encounter with the forward form.

Experiment

Methods

Participants

Initially, 40 participants took part in the experiment. One participant was excluded as her score in the V_YesNo vocabulary test was below 4,000 word families, suggesting that she might not know all words comprising the target binomials (see the following text for more details).

The final pool of 39 participants were all nonnative speakers of English who were academic and administrative staff at a university in Saudi Arabia (L1 Arabic; 30 females,

[3]We should note that the native group in Conklin and Carrol (2021) was reasonably large, and we expect less variation in a group of native speakers than in a group of nonnatives, especially when the native group is drawn from a similar pool (native speaker undergraduates at a British university). Thus, their pattern of results should be robust. However, researchers may wish to replicate their novel research.

https://doi.org/10.1017/S0272263122000237 Published online by Cambridge University Press

average age = 34.39, *SD* = 10.61).[4] They started learning English at an average age of 6.5 years (*M* = 6.46; *SD* = 5.19). Their self-reported proficiency scores (on a scale from 1 = very poor to 5 = excellent) were: reading *M* = 4.63, *SD* = 0.55; writing *M* = 4.60, *SD* = 0.55; speaking *M* = 4.63, *SD* = 0.60; and listening *M* = 4.57, *SD* = 0.61.

To obtain a rough estimate of their proficiency in English, the participants completed the V_YesNo online vocabulary test (Meara & Miralpeix, 2017; maximum score = 10,000). The test presents participants with 200 items (half real words and half imaginary pseudowords) and instructs them to press "Yes" if they know the meaning of the presented form and "Next" if they do not. The score is adjusted downward based on guessing (i.e., pressing "Yes" for pseudowords) using an equation (see Meara & Miralpeix, 2017, p. 120). Uchihara and Clenton (2020) found a significant association between the V_YesNo test scores and speaking ability. Moreover, previous versions of the Yes/No vocabulary test format demonstrated a medium to strong correlation with proficiency measures (e.g., Meara & Jones, 1988; Miralpeix & Muñoz, 2018). The scores of our participants in the V_YesNo test ranged between 4,000 and 9,302 (*M* = 6712.05, *SD* = 1323.96), which roughly indicates a good to high level of proficiency (Meara & Miralpeix, 2017). The vocabulary test score was added as a covariate in all mixed-effects models (see "Analysis") to control for the effect of proficiency on RTs.

Materials

Thirty-two "noun-and-noun" binomials were selected for the present study, taken from Conklin and Carrol (2021), to represent two categories of binomials: existing ($n = 12$) and novel ($n = 20$). The full list of 32 items and their features is presented in Appendix S1 (Online Supplementary Materials). All constituent words belonged to the most frequent 4,000 word families in English (BNC/COCA List with 25 1,000-word bands; Nation, 2012), meaning that our participants should be familiar with them.

The existing binomials were highly frequent phrases (BNC frequency per million) and had a conventionalized order: forward *M* = 351.58, *SD* = 305.74; reversed *M* = 23.25, *SD* = 27.00; $t(11) = 3.71$, 95% CI [144.59, 512.08], $p < .001$, $d = 1.51$. The novel binomials were infrequent phrases (1–11 occurrences in the BNC) constructed using two common nouns (most frequent 4,000 word families). They did not have a typical configuration (forward *M* = 3.90, *SD* = 2.75; reversed *M* = 3.35, *SD* = 2.91; $t(19) = 0.82$, 95% CI [−0.86, 1.96], $p = .43$, $d = 0.19$). More details on item selection and categorization can be found in Conklin and Carrol (2021). Appendix S2 (Online Supplementary Materials) presents characteristics of the target stimuli.

Because participants in the present study were nonnatives from the same L1 background (Arabic), we also considered L1-L2 congruency of the existing and novel binomials. We operationalized congruency in two steps: existence (exists as a common binomial in both language vs. only exists in one language) and configuration or direction (same in both languages vs. different in the two languages). The two steps are explained in detail in Appendix S3 (Online Supplementary Materials). Based on this operationalization, 18 out of the 20 novel English binomials were not common in Arabic (two existed in Arabic in the opposite direction) but only 6 out of the 12 existing English binomials had the same direction in Arabic (six had a different direction in Arabic). To check whether the pattern of results reported in the following text

[4]Four of the 39 participants reported learning a language other than Arabic (English or French) at an early age. They can thus be considered balanced bilinguals. We conducted the analyses excluding them, and the pattern of results remained the same.

https://doi.org/10.1017/S0272263122000237 Published online by Cambridge University Press

(see "Results") was influenced by existence and direction in Arabic, we fit all models with and without the eight nonmatching items. The pattern of results remained the same.[5]

We also considered how familiar our participants are likely to be with the English form of the binomials. To test this, a familiarity rating task was administered to 23 L1 Arabic–L2 English speakers who were comparable to our main participant pool.[6] They were instructed to rate both existing and novel binomials (intermixed in a list) for familiarity on a scale from 1 = very unfamiliar to 7 = very familiar. The results showed significantly higher familiarity ratings for existing binomials ($M = 6.51$, $SD = 0.79$) than novel binomials ($M = 4.86$, $SD = 0.84$; t (30) = 5.52, 95% CI [1.05, 2.27], $p < .001$, $d = 2.03$). To examine the potential effect of familiarity on the pattern of results, we fit all models (see "Results)" with the average familiarity rating score as a covariate and found no significant effect of rating scores. More importantly, including familiarity as a covariate in the analysis did not alter the pattern of results. It should be noted that the novel binomials were rated toward the middle of the scale. We will return to this point in "Discussion."

Three stories of approximately 1,100 words each were adapted from Conklin and Carrol (2021) to include the 32 target items. The passages were simplified to ensure suitability for our nonnative participants (99% of words belonged to the most frequent 4,000 word families in English). All target "existing" and "novel" binomials were presented once in the forward form and once in the reversed form. Half the novel binomials ($n = 10$) were then presented one more time in the forward form to make a total of two exposures and the other half ($n = 10$) were presented three more times in the forward form to make a total of four exposures. The reversed form for both existing and novel binomials occurred once after all occurrences of the corresponding forward form. Conklin and Carrol (2021) conducted a predictability norming task with native speakers of English. We included their predictability scores as a potential covariate in the analysis. The passages, full data, and R codes are available through Open Science Framework at https://osf.io/kymsp/?view_only=ec6b2f7e9ac74f15be7afcd864373f07.

It is important to note here that the novel binomials in Conklin and Carrol (2021) were counterbalanced across two lists, such that one order (*wires and pipes*) was the forward direction on one list and the other (*pipes and wires*) was the forward direction on the other list. This was done to account for any inherent word order preferences in the novel items. However, no list differences were found, thus, in the current study, items appeared in a single list with one version designated as the forward version.

Procedure

Upon arrival at the lab, the participant signed a consent form and completed the vocabulary test, then the eye-tracking experiment started. Eye movements were recorded monocularly using an SR Research EyeLink 1000+ eye-tracker. A desk-mounted chinrest was used to minimize head movement. A 9-point grid calibration procedure was conducted before the experiment and each screen was preceded by a

[5]Only one difference was found in Analysis 1 for the total RT word 1 measure where the Type × Direction interaction was not significant ($p = .56$).

[6]To check comparability between the two groups, we gave the participants who completed the familiarity rating task the same V_YesNo Vocabulary test. Their scores ($M = 6568.96$, $SD = 1127.03$) were similar to those of the main group (see "Participants") and the difference was not significant: (t (60) = −0.43, 95% CI [−803.27, 517.08], $p = .67$, $d = 0.12$).

https://doi.org/10.1017/S0272263122000237 Published online by Cambridge University Press

fixation point for drift correction. The eye-tracker was recalibrated before each story and whenever needed. The stories were presented in Courier New, 18-point font and were double-spaced. Participants were told to read the stories as naturally as possible for comprehension and to press the space bar to go to the next screen. In the texts, neither existing nor novel binomials appeared at the beginning or end of a line or across a line break. Each story was followed by five comprehension questions to ensure that participants attended to the text (average percentile score: 94.36%, *SD* = 6.41). Performance on the comprehension questions indicates that participants understood the stories.

Analysis

Data cleaning was done prior to the analysis according to the four-stage process in the DataViewer software. Single fixations shorter than 100 ms and longer than 800 ms were removed (5% of all fixations). Then, following Conklin and Carrol (2021), we excluded trials that were discontinued or where track loss was experienced. Any phrase that was completely skipped was also excluded from the analysis. In such cases, all subsequent occurrences of the item, including the reversed form, were also removed. This resulted in a loss of 1% of the data points in both analyses. We also conducted the analysis with the full data set and the resulting models were the same.

The analysis was conducted with R version 4.0.5. (R Core Team, 2021). Linear mixed-effects models were constructed and analyzed using the *lme4* (version 1.1-26; Bates et al., 2015) and *lmerTest* (version 3.1-3; Kuznetsova et al., 2017) packages. Three interest areas were analysed for each model: the whole phrase, word 1, and word 3. The middle word "and" was skipped more than 45% of the time, so it was excluded from the analysis. Separate models were constructed for two eye-movement measures: first-pass RT and total RT. RTs were log-transformed to reduce skewness in the data. All analyses adopted the maximal random effects structure justified by the design (see the following text for more details). Final models were checked for collinearity, and no issues were observed (all VIFs < 7). Words and phrases that received no fixations during first-pass reading were excluded from subsequent analyses.

We conducted two separate analyses to answer the research questions. First, we compared the RTs of forward and reversed forms for both the existing and novel binomials (RQ1). Then, RTs for novel binomials only were compared across three iterations (first, last, reversed) for both the 2-repetition and 4-repetition binomial categories (RQ2).

Results

Table 1 presents mean RTs for both existing and novel binomials for all occurrences in both directions. Existing binomials showed no clear pattern, with slower RTs for the forward form under some measures and for the reversed form under other measures. For novel binomials, however, the pattern was much clearer. RTs were shorter (i.e., faster processing) with more exposure to target items (first vs. last vs. reversed) for both repetition categories (2 vs. 4). It should be noted that for the 4-repetition novel binomials, occurrences in the middle (second, third, last) did not always exhibit this processing advantage as exposure increased; the last exposure for 4-repetition items had a longer average RT than the third exposure under all measures. We will return to this in "Discussion."

Table 1. Mean RTs in milliseconds with standard deviation in parentheses for existing binomials and novel binomials for first-pass and total RT

		Existing binomials (*n* = 12)		Novel binomials (2-repetition) (*n* = 10)			Novel binomials (4-repetition) (*n* = 10)				
		Forward	Reversed	First	Last	Reversed	First	Second	Third	Last	Reversed
First-pass RT	Whole phrase	675.40 (356.80)	680.05 (353.31)	763.44 (403.68)	694.87 (420.53)	621.97 (351.04)	714.18 (422.02)	703.98 (371.75)	634.39 (349.77)	665.51 (329.41)	617.69 (358.59)
	Word 1	321.64 (166.95)	306.31 (160.09)	333.47 (159.59)	310.78 (179.50)	294.71 (137.04)	323.37 (161.27)	338.05 (159.16)	312.62 (146.35)	317.13 (131.14)	303.75 (165.90)
	Word 3	294.30 (121.34)	300.83 (129.12)	383.40 (187.98)	340.34 (146.29)	316.84 (148.04)	364.65 (182.04)	321.26 (153.03)	303.80 (130.20)	322.79 (149.21)	301.58 (127.86)
Total RT	Whole phrase	834.17 (444.75)	836.53 (418.00)	1056.09 (516.29)	936.25 (509.15)	819.46 (464.86)	1022.63 (548.71)	879.49 (452.93)	812.03 (393.67)	848.79 (416.21)	783.62 (402.40)
	Word 1	386.85 (252.48)	368.86 (216.63)	426.54 (239.51)	403.96 (265.08)	369.10 (219.55)	431.44 (263.00)	403.85 (232.57)	372.54 (210.83)	395.85 (207.74)	370.15 (215.15)
	Word 3	349.14 (198.10)	356.97 (187.07)	486.21 (293.06)	410.89 (231.55)	373.87 (228.87)	452.22 (264.99)	375.41 (219.22)	335.21 (162.40)	368.65 (217.74)	339.19 (174.60)

Notes: For word 1 and word 3 values, words that received no fixations are discounted; values for the whole phrase include trials where either word 1 or word 3 (but not both) was skipped.

RQ1: Is the language processing system sensitive to novel binomials in L2 input that simulates a real-world context?

To compare the mean RTs (first-pass and total) of existing and novel binomials, we constructed linear mixed-effects models for the whole phrase, word 1 and word 3. Following Conklin and Carrol (2021), we included Type (existing vs. novel) and Direction (forward vs. reverse) as fixed effects in all models. Word-level factors including length (in letters) and frequency (on the Zipf scale) were also included as fixed factors.[7] Additionally, number of repetitions (1, 2, 4) was added as a covariate in all models to control for the effect of variation in encounters on the subsequent reverse form of existing and novel binomials. All models included random intercepts for subjects and items. The *optimx* optimizer was used to avoid convergence errors when needed. Initially, by-subject random slopes for the effects of Type and Direction were included,[8] but this led to unavoidable convergence errors. As a result, only by-subject random slopes for Type were included in all models. Three other factors were then added stepwise one-by-one to each original model to examine their potential effect on RTs: phrase frequency on the Zipf scale, forward association strength, and cloze probability. Log-likelihood (X^2) tests and AIC values were used to compare the resulting model with the original model, and only factors that significantly improved the model fit were kept. All models also included the log-transformed vocabulary test score as a proxy for L2 proficiency.

All resulting models are presented in Table 2. Vocabulary Score was a significant predictor in all models but not for the word 3 first-pass measure, pointing to shorter RTs as proficiency increased. The Type × Direction interaction was significant for all first-pass and total reading measures except for the word 1 first-pass analysis. The *difflsmeans* function in the *lmerTest* package was used to compute pairwise comparisons (see Appendix S4, Online Supplementary Materials). We calculated Cohen's *d* for pairwise comparisons as a standardized measure of effect size based on the guidelines provided by Brysbaert and Stevens (2018). Results suggest that nonnatives were not sensitive to the established configuration of frequent existing binomials in English. There were no significant differences between the processing of forward and reversed forms for existing binomials (whole phrase) in first-pass RT ($\beta = -0.01$, $t = -0.24$, 95% CI [−0.08, 0.06], $p = .81$, $d = -0.01$) or total RT ($\beta = -0.01$, $t = -0.39$, 95% CI [−0.06, 0.04], $p = .69$, $d = -0.02$). These findings stand in sharp contrast with those reported in Conklin and Carrol (2021) for native speakers who showed a clear processing advantage for the forward form of existing binomials over their reversed forms. It seems that our nonnative speakers are not sensitive to the established word order of common binomials in English.

For the novel binomials, pairwise comparisons suggest significantly shorter RTs for the reversed forms in comparison to the forward forms for all measures with a small to medium effect. This is most clearly seen for the whole phrase, for both first-pass RT ($\beta = 0.15$, $t = 5.11$, 95% CI [0.09, 0.21], $p < .001$, $d = 0.25$) and total RT ($\beta = 0.28$, $t = 13.95$, 95% CI [0.24, 0.31], $p < .001$, $d = 0.62$). This result is similar to (but more robust than)

[7]For the first-pass RT measure for word 1 we did not include word 3 length or word 3 Zipf frequency as word 3 had not been encountered in reading yet at this point.

[8]The study employed a within-subject design (i.e., all participants saw items under all conditions: existing binomials, 2-repetition novel binomials, and 4-repetition novel binomials). Thus, we only included by-participant random slopes (no by-item slopes) to account for the random variation in repeated measures (see Linck & Cunnings, 2015).

https://doi.org/10.1017/S0272263122000237 Published online by Cambridge University Press

https://doi.org/10.1017/S0272263122000237 Published online by Cambridge University Press

Table 2. Linear mixed-effects model for existing versus novel binomials in forward and reversed forms for whole phrase RTs, word 1 and word 3

	First pass RT				Total RT			
	Intercept	*SE*	*t*	*p*	*Intercept*	*SE*	*t*	*p*
Whole phrase								
(Intercept)	12.34	1.33	9.31	<.001***	15.47	1.41	10.96	<.001***
Type (novel)	0.05	0.07	0.71	.48	0.19	0.09	2.07	.048*
Direction (reverse)	0.01	0.04	0.24	.81	0.01	0.03	0.39	.69
Type × Direction	−0.16	0.05	−3.32	<.001***	−0.29	0.03	−8.86	<.001***
Repetitions	−0.02	0.02	−0.77	.45	−0.01	0.03	−0.39	.70
W1 Length	0.02	0.02	0.96	.35	0.02	0.02	0.80	.43
W3 Length	−0.02	0.02	−1.12	.27	0.00	0.02	0.06	.95
W1 Zipf	−0.06	0.05	−1.28	.21	−0.10	0.06	−1.64	.11
W3 Zipf	0.05	0.06	0.96	.35	−0.02	0.07	−0.24	.81
Vocabulary Test Score (log)	−0.67	0.15	−4.55	<.001***	−0.95	0.15	−6.128	<.001***
Random effects:	**Variance**	**SD**			**Variance**	**SD**		
Subject	0.02	0.15			0.03	0.18		
Subject|Type	0.00	0.04			0.00	0.03		
Item	0.01	0.08			0.02	0.13		
Residual	0.33	0.57			0.15	0.39		
Word 1								
(Intercept)	10.27	0.96	10.68	<.001***	12.90	1.14	11.34	<.001***
Type (novel)	0.01	0.05	0.15	.88	0.02	0.07	0.26	.80
Direction (reverse)	−0.03	0.03	−0.99	.32	−0.02	0.03	−0.50	.62
Type × Direction	−0.07	0.04	−1.82	.07	−0.14	0.04	−3.50	<.001***
Repetitions	−0.00	0.02	−0.15	.89	0.01	0.02	0.53	.60
W1 Length	0.01	0.01	0.72	.48	0.02	0.02	0.82	.42
W3 Length	–	–	–	–	0.00	0.02	0.20	.84
W1 Zipf	−0.07	0.03	−2.33	.03*	−0.15	0.05	−2.93	.007**
W3 Zipf	–	–	–	–	−0.07	0.06	−1.18	.25
Vocabulary Test Score (log)	−0.49	0.11	−4.53	<.001***	−0.70	0.12	−5.60	<.001***
Random effects:	**Variance**	**SD**			**Variance**	**SD**		
Subject	0.02	0.13			0.02	0.14		
Subject|Type	0.00	0.01			0.00	0.02		
Item	0.00	0.06			0.01	0.09		
Residual	0.17	0.41			0.21	0.46		

(*Continued*)

Table 2. (*Continued*)

	First pass RT				Total RT			
	Intercept	*SE*	*t*	*p*	*Intercept*	*SE*	*t*	*p*
Word 3								
(Intercept)	7.44	0.82	9.07	<.001***	10.01	1.07	9.36	<.001***
Type (novel)	0.18	0.05	3.34	.002**	0.28	0.09	3.21	.003**
Direction (reverse)	0.02	0.03	0.84	.40	0.03	0.03	1.08	.28
Type × Direction	−0.18	0.03	−5.54	<.001***	−0.28	0.04	−7.36	<.001***
Repetitions	−0.02	0.02	−0.93	.36	−0.03	0.03	−1.02	.32
W1 Length	0.02	0.01	1.48	.15	0.02	0.02	0.87	.39
W3 Length	0.00	0.01	0.37	.72	0.00	0.02	0.05	.96
W1 Zipf	−0.08	0.04	−2.31	.03*	−0.10	0.06	−1.74	.09
W3 Zipf	−0.04	0.04	−0.90	.38	−0.00	0.07	-0.07	.95
Vocabulary Test Score (log)	−0.15	0.09	−1.68	.10	−0.43	0.11	−3.77	<.001***
Random effects:	**Variance**	**SD**			**Variance**	**SD**		
Subject	0.01	0.08			0.02	0.13		
Subject|Type	0.01	0.08			0.01	0.08		
Item	0.00	0.06			0.01	0.11		
Residual	0.15	0.38			0.20	0.44		

Note: p-values are estimated using the *lmerTest* package in R (Kuznetsova et al. 2017).
*p <0.05; **p <0.01; ***p <.001.

Table 3. Linear mixed–effects model for novel binomials in different iterations (first, last and reversed) for whole phrase RTs, word 1 and word 3

	First pass RT				Total RT			
	Intercept	*SE*	*t*	*p*	*Intercept*	*SE*	*t*	*p*
Whole phrase								
(Intercept)	11.99	1.45	8.25	<.001***	15.26	1.49	10.25	<.001***
Category (4–repetition)	−0.12	0.05	−2.40	.02*	−0.05	0.07	−0.73	.47
Iteration (last)	−0.12	0.04	−2.78	.006**	−0.14	0.03	−5.14	<.001***
Iteration (reversed)	−0.20	0.04	−4.73	<.001***	−0.28	0.03	−10.22	<.001***
Category (4–repetition) × Iteration (last)	0.12	0.06	2.04	.04*	−0.03	0.04	−0.73	.47
Category (4–repetition) × Iteration (reversed)	0.10	0.06	1.70	.09	0.01	0.04	0.33	.74
W1 Length	0.04	0.02	2.42	.03*	0.03	0.03	0.78	.45
W3 Length	−0.02	0.02	−1.29	.22	0.01	0.03	0.40	.69
W1 Zipf	−0.02	0.06	−0.36	.72	−0.21	0.11	−1.95	.07
W3 Zipf	0.04	0.06	0.72	.49	0.03	0.09	0.34	.74
Vocabulary Test Score (log)	−0.69	0.16	−4.28	<.001***	−0.95	0.16	−6.00	<.001***
Phrase Zipf	0.17	0.09	1.91	.08	0.33	0.16	2.12	.054
Forward Association	1.59	0.78	2.05	.06	–	–	–	–
Random effects:	**Variance**	**SD**			**Variance**	**SD**		
Subject	0.04	0.20			0.04	0.21		
Subject|Category	0.00	0.01			0.00	0.02		
Item	0.00	0.04			0.02	0.12		
Residual	0.35	0.59			0.15	0.38		
Word 1								
(Intercept)	10.16	0.89	11.42	<.001***	13.18	1.15	11.44	<.001***
Category (4–repetition)	−0.05	0.03	−1.52	.13	0.00	0.06	0.00	1.00
Iteration (last)	−0.09	0.03	−3.02	.003**	−0.08	0.03	−2.46	.01*
Iteration (reversed)	−0.13	0.03	−4.12	<.001***	−0.16	0.03	−4.75	<.001***
Category (4–repetition) × Iteration (last)	0.10	0.04	2.34	.02*	0.02	0.05	0.36	.72
Category (4–repetition) × Iteration (reversed)	0.06	0.04	1.41	.16	0.01	0.05	0.16	.88
W1 Length	0.02	0.01	2.20	.04*	0.00	0.03	0.11	.91
W3 Length	–	–	–	–	0.03	0.03	1.08	.30
W1 Zipf	−0.01	0.03	−0.32	.76	−0.12	0.08	−1.52	.15
W3 Zipf	–	–	–	–	−0.06	0.07	−0.83	.42
Vocabulary Test Score (log)	−0.52	0.10	−5.17	<.001***	−0.75	0.12	−6.23	<.001***
Forward Association	0.73	0.37	1.96	.07	–	–	–	–

(*Continued*)

Table 3. (*Continued*)

	First pass RT				Total RT			
	Intercept	*SE*	*t*	*p*	*Intercept*	*SE*	*t*	*p*
Random effects:	**Variance**	**SD**			**Variance**	**SD**		
Subject	0.02	0.13			0.03	0.17		
Subject|Category	0.00	0.05			0.01	0.09		
Item	0.00	0.02			0.01	0.10		
Residual	0.17	0.41			0.21	0.45		
Word 3								
(Intercept)	9.11	1.05	8.70	<.001***	10.87	1.16	9.39	<.001***
Category (4–repetition)	−0.04	0.04	−0.96	.35	−0.08	0.06	−1.18	.25
Iteration (last)	−0.10	0.03	−3.67	<.001***	−0.15	0.03	−4.74	<.001***
Iteration (reversed)	−0.17	0.03	−5.78	<.001***	−0.25	0.03	−7.57	<.001***
Category (4–repetition) × Iteration (last)	0.00	0.04	0.11	.91	−0.03	0.05	−0.71	.48
Category (4–repetition) × Iteration (reversed)	0.01	0.04	0.19	.85	−0.00	0.05	−0.08	.93
W1 Length	0.03	0.02	1.45	.17	0.03	0.03	0.96	.36
W3 Length	0.00	0.02	0.13	.90	0.01	0.03	0.33	.75
W1 Zipf	−0.12	0.06	−1.92	.08	−0.18	0.10	−1.79	.10
W3 Zipf	0.00	0.05	0.09	.93	0.05	0.08	0.61	.55
Vocabulary Test Score (log)	−0.36	0.11	−3.21	.003**	0.32	0.15	2.15	.051
Phrase Zipf	0.19	0.10	1.93	.08	−0.56	0.12	−4.77	<.001***
Random effects:	**Variance**	**SD**			**Variance**	**SD**		
Subject	0.02	0.13			0.02	0.14		
Subject|Category	0.00	0.00			0.00	0.01		
Item	0.00	0.07			0.01	0.11		
Residual	0.15	0.39			0.20	0.44		

Note: p–values are estimated using the *lmerTest* package in R (Kuznetsova et al. 2017).
*p <0.05; **p <0.01; ***p <.001.

the findings reported in Conklin and Carrol (2021) for native speakers, suggesting that just like natives, nonnatives seem to quickly develop a link between the two words in memory, exhibiting an advantage in processing, even when they appear in a different order than previous encounters.

As a final step in the analysis, we also tested for the contribution of the Type × Direction × Vocabulary Score interaction to the model fit.[9] This was intended to reveal any modulating effect of L2 proficiency, that is, to examine if nonnatives with higher vocabulary scores read existing and novel binomials in the forward and reverse direction differently. This three-way interaction did not significantly contribute to any of the models, suggesting similar patterns regardless of L2 proficiency.

One final notable finding in Table 2 is that number of repetitions was not significant in all models, suggesting no difference between items that were seen twice and those that were seen four times. The next research question aimed to examine the possibility that number of encounters might modulate this effect in more detail.

RQ2: What is the effect of frequency of exposure on nonnatives' sensitivity to novel binomials in a real-world context?

The effect of frequency of exposure on RTs of novel binomials (word 1, word 3 and whole phrase) was explored in Analysis 2. This analysis included Category (2-repetition vs. 4-repetition) and Iteration (first, last, reversed) as fixed effects and by-subject random slopes for Category. All other fixed factors and covariates were the same as those included in Analysis 1. Additionally, we examined the three-way interaction between Category, Iteration, and Vocabulary Score, but it did not significantly improve any of the models.

The resulting models are presented in Table 3. Overall, there appears to be a main effect for Vocabulary Score (but see the total RT measure for Word 3) and a main effect for Iteration, but not Category (but see first-pass RT for the whole phrase). Pairwise comparisons across the three Iteration levels, regardless of repetition, are presented in Appendix S5 (see Online Supplementary Materials). These are computed using the *difflsmeans* function in the *lmerTest* package.

The general pattern seems to suggest significantly different RTs across the three iterations with a small to medium effect (first > last > reversed). Thus, it appears that with more exposure to binomials, nonnatives developed a sensitivity to the co-occurrence of the content words, spending less time reading them each time they appeared together. As for the reversed form, which was always included after all occurrences of the forward form, the results suggest that nonnatives dealt with it as another exposure to the binomial, ignoring the configuration mismatch.

Discussion

Research examining how nonnative speakers process novel lexical patterns in context is fairly limited. The present study aims to fill this gap by recording the eye movement patterns of nonnative speakers of English (L1 Arabic) as they read stories seeded with novel binomials to address two research questions. First, we examined whether the nonnatives developed a sensitivity to the canonical order of novel binomials after

[9]This model also included all possible two-way interactions to control for their effect: Direction × Vocabulary Score and Type × Vocabulary Score. This was also the case in Analysis 2 which included Category × Vocabulary Score and Iteration × Vocabulary Score.

https://doi.org/10.1017/S0272263122000237 Published online by Cambridge University Press

exposure and compared their processing to existing binomials (Research Question 1). Second, we looked at the effect of frequency (two vs. four exposures) on the development of sensitivity to the novel binomials (Research Question 2).

In response to the first research question, results showed no processing advantage for existing, common, binomials (*time and money*) over their less frequent reversed forms (*money and time*). Thus, unlike natives in Conklin and Carrol (2021), nonnatives in the present study were generally not sensitive to binomials' canonical word order. This result seems to support Siyanova-Chanturia et al.'s (2011) finding for a limited nonnative sensitivity to word order preferences that emerged only as proficiency increased. In the present study, however, proficiency did not seem to modulate sensitivity to binomials' configuration. While participants in Siyanova-Chanturia et al. (2011) came from a variety of L1 backgrounds, we targeted a homogenous nonnative population (L1 Arabic–L2 English). Previous research on collocations and idioms often report congruency as an important factor in the nonnative processing of lexical patterns (e.g., Carrol et al., 2016; Sonbul & El-Dakhs, 2020). A follow-up analysis that was conducted on a subset of binomials that matched in the two languages showed the same pattern of results, namely, no sensitivity to the canonical configuration (see "Materials" for details). The fact that the L1-L2 matched existing binomials in the present study comprised only six items might explain the lack of effect. Future research on nonnative binomial processing should address the congruency effect more directly with a larger set of congruent and incongruent items. Another factor to consider in future research is binomial familiarity. The norming study that we conducted with a group of L1 Arabic–L2 English speakers comparable to the main participant pool (see "Materials") showed that familiarity with novel binomials elicited ratings toward the mid-point of the scale and was only slightly (though significantly) lower than the ratings for existing binomials. More research is needed in this area to tease apart off-line familiarity ratings and online real time performance.

For novel binomials, the results of Analysis 1 (initial forward exposure vs. reversed forms) showed a robust significant advantage (with a small/medium effect) for the reversed form over the forward form for all eye-movement measures (both early and late). As indicated in the preceding text, this result complements Conklin and Carrol's (2021) finding for natives who initially exhibited sensitivity to the combination of single words ("wires" and "pipes") regardless of direction. Thus, like natives in Conklin and Carrol's (2021) study, nonnatives in the present study seem to keep a record of all occurrences of lexical patterns in the input; but unlike native speakers, they might not initially build sensitivity to the preferred word order (*wires and pipes* vs. *pipes and wires*). This seems to support Durrant and Schmitt's (2010) finding that nonnatives are able to extract co-occurrence restrictions from input, refuting traditional claims (e.g., Wray, 2000), and backing up usage-based models of language processing (Bybee, 2006; Ellis, 2002). However, as noted earlier, binomials may be different from other forms of lexical patterns in that they involve co-occurrence and configuration restrictions. In a study on the processing of lexical bundles, Ellis et al. (2008) found that nonnative speakers are sensitive just to the frequency in the language, but native speakers seem to extract the nuance of the bundle's association strength (i.e., how often two words tend to co-occur above chance). Similarly, one might claim that the nonnative participants in the present study were able to exhibit sensitivity to mere frequency but were not sensitive to higher-level restrictions on word order, at least not after one encounter as Analysis 1 seems to suggest.

The second research question (Analysis 2) examined the possible modulating effect of frequency of encounters (two vs. four) on the development of a sensitivity to the

https://doi.org/10.1017/S0272263122000237 Published online by Cambridge University Press

canonical order of binomials. Conklin and Carrol (2021) found that their native speakers processed the forward forms of novel binomials faster than their reversed forms, similar to existing binomials, after four to five exposures.[10] Crucially, exposure to the subsequent reversed form led to a cost (a marked rise in processing time) compared to the most recent encounter, despite this being faster than the first exposure. However, results of Analysis 2 failed to report similar effects for our nonnative participants even after four exposures. In line with the findings of Analysis 1, nonnatives in the present study processed the last reversed form (*pipes and wires*) significantly more quickly (with a small/medium effect) than the experimentally defined forward form (*wires and pipes*) regardless of how many times it was encountered. This finding is further supported by the raw RTs in Table 1, showing similar processing times for the third encounter in the 2-repetition category (backward) and the third encounter in the 4-repetition category (forward). Thus, it seems to be the case that whilst nonnatives did register co-occurrence restrictions in terms of which words go together, they did not register the configuration/order of the words. This lack of a configuration effect stands in contrast with findings of Alotaibi et al. (2022) who found that nonnative Arab learners of English were able to develop sensitivity to the preferred order of binomials (similar to existing phrases) after six exposures. It should be noted, however, that unlike the present study, Alotaibi et al. (2022) included higher frequency levels (up to six occurrences) than the present study (with a maximum of four encounters) and employed declarative post-treatment measures. Sonbul and Schmitt (2013) found a dissociation between gains in declarative (paper-and-pencil) tests and those reflecting online performance. In the present study, we did not employ any post-treatment declarative measure of sensitivity to target binomials' word order. Future research can benefit from combining both online (eye-movement) measures and off-line (paper-and-pencil) tasks to compare findings at both processing levels. Moreover, more encounters can be included to allow participants to develop sensitivity to configuration restrictions. A relevant point, relating to Arabic speakers of English, may be that Arabic seems to be less fixed than English regarding the order of binomials' components (see "Materials"). Thus, it can be speculated that, given the flexibility in their L1, Arabic speakers of L2 English might not develop sensitivity to binomial restrictions in context. Because research on the structure of binomials in Arabic is extremely limited (but see Kaye, 2009), this possibility can only be viewed as a hypothesis that needs to be explored by future research comparing L2 English speakers from a variety of L1 backgrounds. Another related issue is that, in contrast to English, the Arabic script is read from right to left. As the focus of the present study is on binomials' word order in L2 English, Arabic native speakers might be disadvantaged (in comparison to L2 English speakers whose native language is read from left to right). This would be an interesting question to explore in future research.

The fact that nonnatives in the present study developed a sensitivity to one aspect of binomials (i.e., co-occurrence restrictions) but not another (i.e., configuration restrictions) is in line with eye-tracking evidence for the incidental acquisition of single words from context. As indicated earlier, the limited available research in this area (Pellicer-Sánchez, 2016) seems to suggest that nonnative learners tend to develop different word knowledge aspects at different rates: While the form is acquired quickly, knowledge of

[10]While Conklin and Carrol (2021) found a processing cost for reversing a novel binomial after four to five exposures, it is the only study to investigate the development of a sensitivity to word order for novel binomials in native speakers from natural reading. More research is needed to firmly establish the emergence of such a sensitivity.

https://doi.org/10.1017/S0272263122000237 Published online by Cambridge University Press

meaning fails to develop even after repeated exposure. Along the same lines, it can be claimed that not all aspects of binomial knowledge are learned at the same pace. Online sensitivity to the co-occurrence restrictions of binomials can develop quickly from exposure, but sensitivity to the canonical order does not develop even after several exposures. Further eye-tracking research on the processing of novel binomials can include more encounters to arrive at an estimated frequency after which nonnatives develop a sensitivity to binomials' preferred order.

Another parallel in eye-movement patterns between individual words and binomials is the effect of multiple encounters on processing. In their study on individual words, Godfroid et al. (2018) found a tendency for RTs of novel words to initially decrease but then increase around the seventh encounter, reflecting "increased cognitive effort and attempts on their [participants'] part to integrate the words into the sentence contexts and make form-meaning connections" (p. 575). Similarly, our findings showed a slight increasing trend at the fourth encounter for the 4-repetition novel binomials (see Table 1) which disappeared with the presentation of the reversed form. Thus, although we did not intend to examine the gradual effect of exposure on the processing of novel binomials in the present study, a tendency seems to emerge at the fourth encounter, similar to the effect reported by Godfroid et al. (2018) for single words. Building on predictions of the type of processing resource allocation (TOPRA) Model (Barcroft, 2002), one might argue that in the first few encounters with a given binomial, the cognitive demands are high as language users are encoding co-occurrence restrictions (i.e., learning that 'wires' and 'pipes' occur together). Then, around the fourth encounter, more cognitive resources are freed and can thus be devoted to the specific configuration and the overall meaning that is denoted by the phrase. With more exposure, the forward form of novel binomials might show further decrease in RTs. However, given the limited number of encounters in the present study (maximum four), this interpretation is speculative. Future eye-tracking research on the online processing of novel binomials would do well to include more encounters with novel binomials, in line with Godfroid et al.'s (2018) design, to fully explore the possible S-shaped processing of novel lexical patterns in context.

Conclusion

The present study was intended to extend Conklin and Carrol's (2021) findings for the native processing of novel binomials to a population of non-native English speakers (L1 Arabic). Results showed that nonnatives had limited sensitivity to the preferred order of existing binomials; they did not develop sensitivity to the experimentally defined configuration of novel binomials even after four encounters. These results seem to suggest that the nonnative participants were simply recording co-occurrence restrictions, disregarding direction preference, which may be a feature of language that is less salient and therefore requires more input to emerge. The study is limited, however, in that it did not include a balanced number of congruent/incongruent binomials to fully examine the congruency effect. Additionally, the study only included two frequency conditions (two vs. four) and did not include a post-exposure measure of declarative knowledge. Future research should explore the effect of increased frequency and L1-L2 congruency on the acquisition process (both online processing and off-line, post-exposure, gains). Despite the limitations, the present study can be viewed as an initial attempt to examine nonnatives' processing of novel binomials in context. This line of research can further our understanding of the conditions that might help nonnatives develop sensitivity to lexical patterns like *bread and butter* (over *butter and bread*), enabling new and broader explanations of input-driven language development.

Acknowledgment. The researchers thank Prince Sultan University for funding this research project through the research lab [Applied Linguistics Research Lab - RL-CH-2019/9/1].

Supplementary Materials. To view supplementary material for this article, please visit http://doi.org/10.1017/S0272263122000237.

Data Availability Statement. The experiment in this article earned Open Materials and Open Materials badges for transparent practices. The materials and data are available at https://osf.io/kymsp/?view_only=ec6b2f7e9ac74f15be7afcd864373f07.

References

Alotaibi, S. (2020). *The effect of input mode, input enhancement and number of exposures on the learning and processing of binomials in the L2* [Unpublished doctoral dissertation]. University of Nottingham.

Alotaibi, S., Pellicer-Sánchez, A., & Conklin, K. (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.

Arcara, G., Lacaita, G., Mattaloni, E., Passarini, L., Mondini, S., Beninca, P., & Semenza, C. (2012). Is "hit and run" a single word? The processing of irreversible binomials in neglect dyslexia. *Frontiers in Psychology*, *3*, 1–11.

Barcroft, J. (2002). Semantic and structural elaboration in L2 lexical acquisition. *Language Learning*, *52*, 323–363.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*, 1–48

Blais, M.-J., & Gonnerman, L. M. (2013). Explicit and implicit semantic processing of verb particle constructions by French–English bilinguals. *Bilingualism: Language and Cognition*, *16*, 829–846.

Boers, F., & Lindstromberg, S. (2012). Experimental and intervention studies on formulaic sequences in a second language. *Annual Review of Applied Linguistics*, *32*, 83–110.

Brysbaert, M., & Stevens, M. (2018). Power analysis and effect size in mixed effects models: A tutorial. *Journal of Cognition*, 1, *1*–20.

Bybee, J. (2006). From usage to grammar: The mind's response to repetition. *Language*, *82*, 711–733.

Carrol, G., & Conklin, K. (2014). Getting your wires crossed: Evidence for fast processing of L1 idioms in an L2. *Bilingualism: Language and Cognition*, *17*, 784–797.

Carrol, G., & Conklin, K. (2020). Is all formulaic language created equal? Unpacking the processing advantage for different types of formulaic sequences. *Language and Speech*, *63*, 95–122.

Carrol, G., Conklin, K., & Gyllstad, H. (2016). Found in translation: The influence of the L1 on the reading of idioms in a L2. *Studies in Second Language Acquisition*, *38*, 403–443.

Choi, S. (2017). Processing and learning of enhanced English collocations: An eye movement study. *Language Teaching Research*, *21*, 403–426.

Christiansen, M.H., & Arnon, I. (2017). More than words: The role of multiword sequences in language learning and use. *Topics in Cognitive Science*, *9*, 542–551.

Conklin, K., & Carrol G. (2021). Words go together like "bread and butter": The rapid, automatic acquisition of lexical patterns. *Applied Linguistics*, *42*, 492–513

Conklin, K., & Schmitt, N. (2008). Formulaic sequences: Are they processed more quickly than nonformulaic language by native and nonnative speakers? *Applied Linguistics*, *29*, 72–89.

Ding, C., & Reynolds, B. L. (2019). The effects of L1 congruency, L2 proficiency, and the collocate-node relationship on the processing of L2 English collocations by L1-Chinese EFL learners. *Review of Cognitive Linguistics*, *17*, 331–357.

Durrant, P., & Schmitt, N. (2010). Adult learners' retention of collocations from exposure. *Second Language Research*, *26*, 163–188.

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

Ellis, N. C., Simpson-Vlach, R. I. T. A., & Maynard, C. (2008). Formulaic language in native and second language speakers: Psycholinguistics, corpus linguistics, and TESOL. *TESOL Quarterly*, *42*, 375–396.

Erman, B., & Warren, B. (2000). The idiom principle and the open choice principle. *Text and Talk*, *20*, 29–62.

Godfroid, A., Ahn, J., Choi, I., Ballard, L., Cui, Y., Johnston, S., Lee, S., Sarkar, A., & Yoon, H. J. (2018). Incidental vocabulary learning in a natural reading context: An eye-tracking study. *Bilingualism: Language and Cognition*, *21*, 563–584.

Godfroid, A., Boers, F., & Housen, A. (2013). An eye for words: Gauging the role of attention in incidental L2 vocabulary acquisition by means of eye-tracking. *Studies in Second Language Acquisition*, *35*, 483–517.

Goldberg, A. E., & Lee, C. (2021). Accessibility and historical change: An emergent cluster led uncles and aunts to become aunts and uncles. *Frontiers in Psychology*, *12*, 1418.

Gyllstad, H., & Wolter, B. (2016). Collocational processing in light of the phraseological continuum model: Does semantic transparency matter? *Language Learning*, *6*), 296–323.

Irujo, S. (1986). Don't put your leg in your mouth: Transfer in the acquisition of idioms in a second language. *TESOL Quarterly*, *20*, 287–304.

Kaye, A. S. (2009). Cultural ingredients in Arabic lexical pairs (Binomials). *Word*, *60*, 65–78.

Kuznetsova, A., Brockhoff, P., & Christensen, R. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software*, *82*, 1–26.

Libben, M., & Titone, D. (2008). The multidetermined nature of idiom processing. *Memory and Cognition*, *36*, 1103–1121.

Linck, J. A., & Cunnings, I. (2015). The utility and application of mixed-effects models in second language research. *Language Learning*, *65*, 185–207.

Majuddin, E., Siyanova-Chanturia, A., & Boers, F. (2021). Incidental acquisition of multiword expressions through audiovisual materials: The role of repetition and typographic enhancement. *Studies in Second Language Acquisition*, *43*, 985–1008.

Malkiel, Y. (1959). Studies in irreversible binomials. *Lingua*, *8*, 113–160.

Matlock, T., & Heredia, R. R. (2002). Understanding phrasal verbs in monolinguals and bilinguals. In R. R. Herediaand & J. Altarriba (Eds.), *Bilingual sentence processing* (pp. 251–274). Elsevier.

Meara, P., & Jones, G. (1988). Vocabulary size as a placement indicator. In P. Grunwell (Ed.), *Applied linguistics in society* (pp. 80–87). CILT.

Meara, P., & Miralpeix, I. (2017). *Tools for researching vocabulary*. Multilingual Matters.

Miralpeix, I., & Muñoz, C. (2018). Receptive vocabulary size and its relationship to EFL language skills. *International Review of Applied Linguistics in Language Teaching*, *56*, 1–24.

Mollin, S. (2013). Pathways of change in the diachronic development of binomial reversibility in Late Modern American English. *Journal of English Linguistics*, *41*, 168–203.

Mollin, S. (2014). *The (ir) reversibility of English binomials*. John Benjamins.

Morgan, E., & Levy, R. (2016). Abstract knowledge versus direct experience in processing of binomial expressions. *Cognition*, *157*, 384–402.

Nation, I. S. P. (2012). The BNC/COCA word family lists. Document bundled with Range Program with BNC/COCA Lists, 25. https://www.victoria.ac.nz/lals/about/staff/publications/paul-nation/Information-on-the-BNC_COCA-word-family-lists.pdf

Pawley, A., & Syder, F. (1983). Two puzzles for linguistic theory. In J. C. Richards & R. W. Schmidt (Eds.), *Language and communication* (pp. 191–227). Longman.

Pellicer-Sánchez, A. (2016). Incidental L2 vocabulary acquisition from and while reading: An eye-tracking study. *Studies in Second Language Acquisition*, *38*, 97–130.

Pellicer-Sánchez, A. (2017). Learning L2 collocations incidentally from reading. *Language Teaching Research*, *21*, 381–402.

Pellicer-Sánchez, A., & Schmitt, N. (2010). Incidental vocabulary acquisition from an authentic novel: Do things fall apart? *Reading in a Foreign Language*, *22*, 31–55.

Pitts, M., White, H., & Krashen, S. (1989). Acquiring second language vocabulary through reading: A replication of the Clockwork Orange study using second language acquirers. *Reading in a Foreign Language*, *5*, 271–275.

Pritchett, L. K., Vaid, J., & Tosun, S. (2016). Of black sheep and white crows: Extending the bilingual dual coding theory to memory for idioms. *Cogent Psychology*, *3*, 1135512.

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.

R Core Team (2021). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org.

Rommers, J., Dijkstra, T., & Bastiaansen, M. (2013). Context-dependent semantic processing in the human brain: Evidence from idiom comprehension. *Journal of Cognitive Neuroscience*, *25*, 762–776.

Siyanova-Chanturia, A., Conklin, K., & van Heuven, W. J. B. (2011). Seeing a phrase "Timed and Again" matters: The role of phrasal frequency in the processing of multiword sequences. *Journal of Experimental Psychology: Learning Memory and Cognition*, *37*, 776–784.

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

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

Sonbul, S., & Schmitt, N. (2013). Explicit and implicit lexical knowledge: Acquisition of collocations under different input conditions. *Language Learning*, *63*, 121–159.

Titone, D., Columbus, G., Whitford, V., Mercier, J., & Libben, M. (2015). Contrasting bilingual and monolingual idiom processing. In R. R. Heredia & A. B. Cieślicka (Eds.), *Bilingual figurative language processing* (pp. 171–207). Cambridge University Press.

Tiv, M., Gonnerman, L., Whitford, V., Friesen, D., Jared, D., & Titone, D. (2019). Figuring out how verb-particle constructions are understood during L1 and L2 reading. *Frontiers in Psychology*, *10*, 1733.

Uchihara, T., & Clenton, J. (2020). Investigating the role of vocabulary size in second language speaking ability. *Language Teaching Research*, *24*, 540–556.

van Hell, J. G., & Tanner, D. (2012). Second language proficiency and cross-language lexical activation. *Language Learning*, *62*, 148–171.

Waring, R., & Takaki, M. (2003). At what rate do learners learn and retain new vocabulary from reading a graded reader? *Reading in a Foreign Language*, *15*, 130–163.

Webb, S., Newton, J., & Chang, A. C. S. (2013). Incidental learning of collocation. *Language Learning*, *63*, 91–120.

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

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

Wolter, B., & Yamashita, J. (2018). Word frequency, collocational frequency, L1 congruency and proficiency in L2 collocation processing: What accounts for L2 performance? *Studies in Second Language Acquisition*, *40*, 395–416.

Wray, A. (2000). Formulaic sequences in second language teaching: Principle and practice. *Applied Linguistics*, *21*, 463–489.

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.

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

Cite this article: Sonbul, S., El-Dakhs, D. A. S., Conklin, K. and Carrol, G. (2023). "Bread and butter" or "butter and bread"? Nonnatives' processing of novel lexical patterns in context. *Studies in Second Language Acquisition,* **45,** 370–392. https://doi.org/10.1017/S0272263122000237