Figurativeness Matters in the Second Language Processing of Collocations: Evidence From a Self-Paced Reading Experiment

Jinfang Shi[a], Gang Peng[b], and Dechao Li[c]

a *East China Jiaotong University*

b *Research Centre for Language, Cognition, and Neuroscience, Department of Chinese and Bilingual Studies, The Hong Kong Polytechnic University*

c *Department of Chinese and Bilingual Studies, The Hong Kong Polytechnic University*

Abstract: This study reports on a self-paced reading experiment exploring whether the figurativeness of collocations affects L2 processing of collocations. The participants were 40 English native speakers and 44 Chinese-speaking English foreign language learners (including doctoral, postgraduate, and undergraduate students). To ensure that the effect emerged from the figurativeness of collocations rather than other item-related confounds, this study added a literal–literal comparison (e.g., *choose a career* vs. *choose a house*) as a control to the experimental figurative–literal comparison (e.g., *build a career* vs. *build a house*). Results showed that L2 speakers processed figurative collocations more slowly than literal collocation controls but native speakers did not. Importantly, this processing cost for figurative collocations in L2 speakers varied by L2 proficiency but not phrase familiarity. We discuss the results in terms of the dual-route model of formulaic and novel language processing and also incorporate them into the literal salience model of bilingual figurative processing.

Keywords figurativeness effect; L2 collocational processing; self-paced reading; English foreign language learners

---

We are grateful to the anonymous reviewers, Associate Editor Theres Grüter, and Editor Emma Marsden for their insightful comments and suggestions on earlier versions of this article. We also thank Dr. Politzer-Ahles for his valuable suggestions and Prof. Shiyu Wu for sharing his skills in R in data analysis. This research was part of the project "The Processing and Acquisition of Metaphoric Constructions by Chinese EFL Learners" (21YY09) supported by the Jiangxi Social Science Foundation, P.R.C.

Correspondence concerning this article should be addressed to Jinfang Shi, School of Foreign Languages, East China Jiaotong University, No. 808, East Shuanggang Avenue, Nanchang, Jiangxi Province, China. E-mail: jfshi821@163.com

The handling editor for this article was Theres Grüter.

Introduction

Collocations (e.g., *pay a visit, cut the cost*), a subcategory of formulaic language (i.e., idioms, binominals, lexical bundles), are a feature common across languages. The significance of collocation competence has been widely recognized in the field of second language (L2) acquisition and teaching. Corpus-based studies have found that collocations present great challenges to L2 speakers, even for those with an advanced level of proficiency (Granger & Bestgen, 2014; Laufer & Waldman, 2011; Nesselhauf, 2005). Psycholinguistic work has provided distinct insights into the processing and acquisition of L2 collocations, and this field of inquiry has increased considerably in the past decade (Gyllstad & Wolter, 2016; Pulido & Dussias, 2020; Vilkaitė & Schmitt, 2019; Wolter & Gyllstad, 2011, 2013; Wolter & Yamashita, 2015, 2018; Yamashita & Jiang, 2010).

Figurativeness, a semantic variable, is pervasive in collocations. Macis and Schmitt (2017a) analyzed the semantic properties of collocations derived from the Corpus of Contemporary American English (COCA, Davies, 2008) and discovered that a substantial number of collocations have figurative meanings, yet Macis and Schmitt (2017b) also found that even tertiary-level L2 learners had relatively poor knowledge of the figurative meanings of collocations. These findings suggested that figurative collocations pose a great challenge to L2 acquisition. Even so, there is still something of a research gap about how these particular types of collocations are processed and represented in the L2 mental lexicon. Using a cross-modal semantic priming paradigm, Werkmann Horvat, Bolognesi, and Kohl (2021) investigated L2 processing of conventional metaphorical expressions and found that figurative meanings of even very conventional metaphors could cause difficulty for advanced-level L2 speakers. Much of the literature on L2 processing of figurative expressions has focused on idioms (Cieślicka, 2017; Werkmann Horvat, Bolognesi, & Lahiri, 2021). A recurring finding for idioms is that learners usually take longer to process the figurative meaning of idioms (either L2 idioms or translated L1 idioms) than they do for the literal meanings of idioms (Carrol & Conklin, 2014, 2017; Siyanova-Chanturia, Conklin, & Schmitt, 2011). But then could the processing cost for figurativeness extend beyond idioms to more transparent and more decomposable phrases such as collocations; that is, how are figurative collocations processed in the L2 mental lexicon? To answer these questions, our study focused on figurativeness in collocations to explore whether and how this semantic property might affect L2 processing of collocations.

Background Literature

Definition of Collocations

What has been referred to as collocations has varied greatly across studies. Some scholars have defined collocations based purely on phrasal frequency (e.g., Öksüz, Brezina, & Rebuschat, 2021; Vilkaitė, 2016; Vilkaitė & Schmitt, 2019; Wolter & Gyllstad, 2013), while other scholars have based their defini- tion on the semantic relationships between individual words in a collocation (e.g., Gyllstad & Wolter, 2016; Nesselhauf, 2005). These two ways of defining collocations generally correspond to two distinct approaches that researchers have used to operationalize collocations: a corpus-based statistical approach and a semantic-based phraseological approach. The corpus-based statistical approach owes much to corpus linguistics and relies on statistical data, namely, frequency of co-occurrence, to determine whether a word combination is a col- location or not (Sinclair, 1991). By contrast, the semantic-based phraseological approach emphasizes either semantic properties (i.e., whether a phrase or its discrete components have a figurative meaning) or syntactic restrictions (i.e., whether and to what degree the constituent words of a formulaic sequence are restricted) of word combinations (Howarth, 1998; Nesselhauf, 2005). For ex- ample, Howarth (1996, 1998) proposed a collocational continuum model in which four categories of word combinations are distinguished: free combina- tions (open/free collocations), restricted collocations, figurative idioms, and pure idioms. At one end of the continuum lie free combinations, which have the most transparent meaning and in which each lexical unit can be replaced by others following certain grammatical rules (e.g., *pay a bill*). At the other end of the continuum are pure idioms, the most fixed type (e.g., *pay the piper*), whose meanings are often very opaque and cannot be directly derived from the mean- ings of the words composing the idioms. The other two categories, restricted collocations and figurative idioms, fall somewhere in the middle of the cline, having differing degrees of figurativeness. Restricted collocations such as *pay a visit* have one component *pay* used in a figurative, delexicalized, or techni- cal sense and the other *visit* in its literal sense, whereas figurative idioms such as *pay the price* can have either a "figurative meaning in terms of the whole or have a current literal interpretation" (Howarth, 1998, p. 29; examples taken from Gyllstad & Wolter, 2016).

Both approaches have advantages and limitations. The statistical approach allows researchers to operationalize collocational frequency in an objective way but is limited in distinguishing collocations from other types of formulaic sequences like idioms (Howarth, 1998). Because our study was designed to address the effect of figurativeness at the level of collocations, we felt it neces- sary to follow the phraseological tradition and differentiate collocations from idioms and free combinations according to Howarth's model (Howarth, 1996, 1998). Thus, figurative collocations in this study were word combinations in which one of the constituents was used in a figurative sense, whereas literal col- locations were word combinations in which each element was used in its con- crete sense. The literal category was composed of free combinations, whereas the figurative category was a subtype of restricted collocations in Howarth's framework. (Other types of restricted collocations, namely those with a delex- icalized or technical constituent such as in *make progress*, were excluded from this study.) For example, *build a house* is a literal collocation because no figura- tive meaning is involved, but *build a career* is a figurative collocation because the verb *build* is used in a figurative sense (i.e., careers cannot be constructed in the same material way as houses). Meanwhile, figurative collocations are dis- tinct from figurative idioms in Howarth's model because the collocate version (e.g., *pay a visit*) deploys one constituent figuratively, but the idiomatic ver- sion (e.g., *pay* *the price*) engages both figurative (to suffer the consequences of one's actions or misdeeds) and literal meanings (i.e., to pay money for goods or services). Furthermore, figurative collocations are also distinct from pure id- ioms because their meanings are more transparent (e.g., *build a career, bright students*) than those of pure idioms (e.g., *kick the bucket, spill the beans*), of which the meaning cannot be derived from their component words because pure idioms tend to be less transparent or noncompositional. Despite these dif- ferences, figurative collocations are similar to idioms in other respects. For ex- ample, both are formulaic sequences that are widely used by native speakers in everyday language, and their meanings are usually conventionalized to the ex- tent that they are not even considered figurative by native speakers (Werkmann Horvat, Bolognesi, & Lahiri, 2021). Given the special relationship between id- ioms and figurative collocations, we first review studies on idiom processing and then on collocational processing in the first language (L1) and the L2.

Idiom Processing in the First Language and the Second Language

Much of the research on L1 processing of idioms has shown that native speakers tend to process idioms more quickly and accurately than they do matched novel phrases (Cacciari & Tabossi, 1988; Carrol & Conklin, 2014, 2017; Carrol, Conklin, & Gyllstad, 2016; Conklin & Schmitt, 2008; Siyanova-Chanturia et al., 2011). The dual-route model (Van Lancker Sidtis, 2012) provides a reasonable account for the processing advantage of idioms over novel controls. According to this model, speakers access formulaic language through two pathways: direct retrieval and the computational route. Direct retrieval is the default route for frequent, familiar, and formulaic sequences because these have already been stored in the mental lexicon and therefore can be accessed directly. The computational route, however, usually handles the processing of novel phrases because these phrases have no separate representation in long-term memory and can be accessed only through computing and integrating the individual components.

It is noteworthy, however, that direct retrieval cannot be taken as an indicator of holistic storage of formulaic sequences (Siyanova-Chanturia & Martinez, 2015). The processing advantages of idioms over novel phrases could stem from the "simultaneous activation of components of a phrase or the priming of multiple combinations via the base components" (Carrol & Conklin, 2014, p.784). Therefore, computation also takes place in the direct retrieval route when known phrases such as in *flog a dead horse* are encountered, but computation terminates when the language comprehender reaches the recognition point, for instance *dead* in *flog a dead horse*, which is the key of the idiom in Cacciari and Tabossi's (1988) configuration hypothesis. Thus, direct retrieval is not available until the final component of the configuration *horse* is automatically activated by the idiomatic key. So, if the preactivated final word *horse* appears as expected, it will be processed more quickly than a nonactivated and unexpected final word such as *sheep* in the hypothetical phrase *flog a dead sheep*. Furthermore, the two routes in the dual-route model relate more to the speed of processing rather than to an either/or choice, and whether a direct retrieval comes into play is often a matter of the language comprehender's subjective familiarity with the idiom (Carrol & Conklin, 2014).

Although the idiom superiority effect has been found to be clear for native speakers, the evidence has been mixed for nonnative speakers (Carrol et al., 2016; Siyanova-Chanturia et al., 2011). At the same time, much of the research on the processing of ambiguous idioms has indicated that L2 speakers tend to show a processing cost for figurative meanings compared to literal meanings (Conklin & Schmitt, 2008; Siyanova-Chanturia et al., 2011), and in one study, this cost was present even when known L1 idioms were presented in a translated L2 form (Carrol & Conklin, 2017). These findings have lent support to the literal salience model (Cies´licka, 2006), which holds that L2 learners rely mainly on literal analysis to compute the noncompositional figurative meaning of L2 idioms. Thus, when they encounter an idiom like *flog a dead horse* in the sentence *We've all moved on from that problem, so there's no use flogging a dead horse,* nonnative speakers might first think of the literal meaning of each word in the idiom and interpret the phrase as meaning that a dead horse had been beaten, leading to a failure to understand the sentence. Therefore, the figurative computation happens only when the literal interpretation fails or is determined to be inappropriate (Cies´licka, 2017; Clark & Lucy, 1975). The literal-first obligation observed in the L2 processing of ambiguous idioms has provided suggestive evidence that computational analysis is more of a default route for L2 speakers than direct retrieval when they are accessing figurative phrases like idioms.

Collocational Processing in the First Language and the Second Language

Collocations have received increasing attention in psycholinguistic studies because they have "a psychological association between words" that enables the words to prime each other mentally (Hoey, 2005, p. 3). Durrant and Doherty (2010) confirmed this hypothesis with a lexical decision task with which they showed that collocational priming emerged between frequent collocates, regardless of whether the semantic association between collocates was strong (e.g., *eat* and *food*) or weak (e.g., *eat* and *words*). Automatic priming (i.e., priming based on automatic, nonconscious processes, which involves a much shorter stimulus-onset asynchrony than does collocation priming) was restricted to frequent collocates with a strong semantic association (e.g., *eat* and *food*). Eye-tracking researchers also revealed that the probability of one collocate preceding or following another collocate (transitional probabilities) has a significant influence on the fixation durations for a collocation: the higher the transitional probability between collocates, the faster the upcoming collo- cate is accessed (Frisson, Rayner, & Pickering, 2005; McDonald & Shillcock, 2003). Wray (2012) interpreted this collocational priming as a reflection of faster mapping of components onto meanings. She proposed that there is a fundamental difference between L1 and L2 processing of collocations. For native speakers, collocations could be accessed faster through a simultaneous activation of components and a faster mapping between orthographical or phonological forms and emergent phrasal meanings. For nonnative speakers, however, a simultaneous activation of components is unlikely to happen because they tend to decompose collocations into individual components and then infer what the phrases might entail after the individual components have been accessed (Wray, 2002). For example, when the collocation *flat rate* is encountered, native speakers tend to link it to a single unit of meaning—a price that is the same for everyone and in all situations, but nonnative speakers are more likely to think about what *flat* may refer to first and then analyze the meaning of the entire phrase via words-and-rules computation.

There has been a wealth of empirical evidence showing that that native speakers have a processing advantage for collocations over novel sequences, but for nonnative speakers the processing advantage is conditional, modulated by variables such as L1–L2 congruency (Wolter & Gyllstad, 2011, 2013; Wolter & Yamashita, 2015, 2018; Yamashita & Jiang, 2010), frequency (Ellis, Frey & Jalkanen, 2009; Öksüz et al., 2021; Siyanova & Schmitt, 2008; Sonbul, 2015; Vilkaitė & Schmitt, 2019; Wolter & Gyllstad, 2013; Wolter & Yamashita, 2018), semantic transparency of collocations (Gyllstad & Wolter, 2016; Yamashita, 2018), and L2 proficiency (Sonbul, 2015; Wolter & Gyllstad, 2013; Wolter & Yamashita, 2015, 2018; Yamashita & Jiang, 2010). Among the findings, the processing advantages for congruent collocations with a word-for-word translation equivalent in the L1 over incongruent collocations have been found only in the L2 processing of collocations. By contrast, frequency, whether word-level or collocation-level, has been found to affect both L1 and L2 processing of collocations (Öksüz et al., 2021), even though L2 speakers may rely more heavily on word-level frequency than L1 speakers do and may demonstrate a progression toward nativelike processing with gains in L2 proficiency, with a shift in reliance from more reliance on word-level frequency to more reliance on the collocation-level frequency (Wolter & Yamashita, 2018). The findings lend support to Wray's (2002) general account of L2 acquisition of formulaic language: L2 learners without sufficient exposure or with limited L2 proficiency attend more to lexical information and rely more on an analytic approach to understanding collocations, whereas L1 speakers depend more on phrasal representation to access collocational semantics.

But a very important point to note is that most researchers have adopted the statistical approach to defining collocations. There may thus be addi- tional semantic involvement beyond congruency in collocational processing (Yamashita, 2018). For example, Gyllstad and Wolter (2016), operationalizing collocations according to the phraseological approach, found that both native and nonnative speakers processed semitransparent L1–L2 congruent collocations more slowly than they did free combinations. Jankowiak, Rataj, and Naskręcki (2017), in their event-related potential research, found that late bilinguals' sensitivity to figurative meaning varied with the degree of conventionality of the metaphors. Amplitudes of the late N400 component—an indication of enhanced cognitive effort—were most strongly evoked by novel metaphors (e.g., *to harvest courage*), followed by conventional metaphoric phrases (e.g., *to gather courage*) and literal phrases (e.g., *to experience courage*). These findings seem to indicate the same conclusion: beyond-congruency semantic properties are as important as frequency and so should not be ignored in accounts of L2 processing of collocations.

Figurativeness is an important semantic phenomenon that is pervasive in idioms, collocations, and phrasal verbs. Yet, different forms of figurativeness are potentially processed differently (Werkmann Horvat, Bolognesi, & Lahiri, 2021). Compared with idioms, phrasal verbs have been found to pose little challenge for comprehension by learners (Paulmann, Ghareeb-Ali, & Felser, 2015). The figurativeness of collocations has not been given due attention so far because much of the research on L2 figurative processing focuses on the L2 processing of idioms, followed by phrasal verbs (Cies´licka, 2017). It is therefore still unclear whether the processing cost for figurativeness of idioms can extend to more transparent, decomposable collocations. Thus, the issue of how L2 figurative collocations are processed remains an empirical question.

The Current Study

Our study aimed to explore how figurative collocations are processed in the mental lexicon of L1 and L2 speakers. Specifically, we set out to answer three research questions:

1. Do figurative collocations require more processing effort than literal control collocations for both L1 and L2 speakers?
2. Is the processing cost for figurative collocations (i.e., the figurativeness effect) modulated by subjective familiarity?
3. Is the figurativeness effect modulated by L2 proficiency?

The first question that we explored was whether figurativeness affects L1 and L2 processing of collocations. We assumed that the two groups would process figurative collocations in different ways. For L1 speakers, figurative collocations might be familiar to them and might not stand out as having figurative status (Werkmann Horvat, Bolognesi, & Lahiri, 2021). Therefore, we predicted that L1 speakers would present no difference in their process- ing of figurative and literal collocations. In contrast, for L2 speakers who had learned English in a foreign language setting, figurative collocations might be relatively novel, and they might invoke figurative computation after the literal analysis had failed for them (Cies´licka, 2006; Clark & Lucy, 1975). Therefore, we predicted that L2 speakers would present a processing cost for figurative collocations.

The second question was to investigate whether the figurativeness effect, if it appeared, would be modulated by familiarity (i.e., whether participants reported knowing the meaning of the collocation). We checked the influence of familiarity because it has been demonstrated that overall familiarity is a good predictor of idiom processing (Libben & Titone, 2008), is the main driver of the processing advantage for formulaic sequences (Tabossi, Fanari & Wolf, 2009) and has a facilitating effect on both L1 and L2 figurative comprehension (Carrol et al., 2016). Therefore, we predicted that subjective familiarity would shape the figurativeness effect.

The third question was to test whether the figurativeness effect, if it emerged in the L2 speakers, would be modulated by L2 proficiency. Proficiency has been revealed as a significant facilitator in developing nativelike representations of collocations (Wolter & Yamashita, 2018; Yamashita & Jiang, 2010) even though even advanced L2 speakers have been found to have difficulties in processing conventional metaphors (Werkmann Horvat, Bolognesi & Kohl, 2021). Therefore, we predicted that L2 proficiency would modulate the influence of figurativeness in L2 processing of collocations.

Methods

Participants

We recruited 44 nonnative English speakers (Chinese learners of English: seven males and 37 females) and 40 native English speakers (16 males and 24 females) to participate in this experiment. The nonnative group included eight doctoral candidates studying at a university in Hong Kong, China, as well as nine postgraduates and 27 undergraduates from a university in mainland China (M_{age} = 23.16 years, *SD* = 5.26, 95% CI [21.56, 24.76]). All the nonnative participants had learned English in mainland China, and only one doctoral student had experienced one year of study in an English- speaking country. The native English speakers were students or exchange stu- dents at a university in Hong Kong, China, and most of them came from the United States, Britain, and Canada (M_{age} = 22.78 years, *SD* = 5.38, 95% CI [21.14, 24.42]). All the participants had normal or corrected-to-normal vision and were offered payment for their participation. To study the influ- ence of L2 proficiency on the processing of figurative collocations, we ad- ministered a 50-item C-test[1] (see Appendix S1 in the online Supporting In- formation; the test is also available in the IRIS Database, Shi, Peng, & Li, 2022a, and at https://osf.io/946vt/) to the group of Chinese participants to measure their L2 proficiency. The participants' scores in the C-test demon- strated high internal consistency (Cronbach's α = .904). The mean and the standard deviation of the C-test score were 83.77 (out of 100) and 11.31, re- spectively. Table 1 presents the information regarding the Chinese participants' English-learning background obtained with a questionnaire administered immediately before or after the experiment. The Pearson product-moment cor- relation test indicated that the nonnative C-test scores and self-rating scores did not correlate highly (all *r*s < .23) nor significantly with alpha set at .05.

Table 1 English learning background of Chinese participants in the experiment

Learning background	*M*	*SD*	95% CI
Age started learning English	10.02	1.89	[9.45, 10.59]
Years studying English	11.70	2.18	[11.04, 12.36]
General English proficiency	6.55	1.08	[6.22, 6.88]
English listening	6.14	1.73	[5.61, 6.67]
English speaking	6.18	1.61	[5.69, 6.67]
English reading	6.75	1.58	[6.27, 7.23]
English writing	5.91	1.29	[5.52, 6.30]

Note. General proficiency, listening, speaking, reading, and writing were self-ratings scored on a 10-point scale: 1 (*minimal proficiency*); 10 (*near-native proficiency*).

Design

To examine differences in processing figurative and literal collocations, we adopted a self-paced reading task because it allowed for a direct comparison of the reaction time (RT) for terminal words in verb + noun (VN) colloca- tions in the figurative and literal conditions. We used VN collocations because they are "the standard, first-choice way of expressing certain concepts" (Hill, Lewis, & Lewis, 2000, p. 99) and the most frequent form of figurative language (Cameron, 2003; Deignan, 2005; Steen, Dorst, Herrmann, Kaal, & Krennmayr, 2010). Therefore, in the self-paced experiment, the second word (the noun of the VN collocation) always varied across the items. The underlying assumption of the task was that, if a figurative collocation required a greater processing cost than did a literal collocation, it would be revealed through a longer RT to the noun of the figurative collocation. This, however, presupposes that the lexical and sentential characteristics are well controlled; otherwise, it might be hard to justify the claim that a longer RT for a figurative collocation was the result of its figurativeness. For example, Cardillo, Schmidt, Kranjec, and Chatterjee (2010) contended that it might be stimulus-related variables such as lexical differences in word length, frequency, or sentential differences rather than figurativeness that account for differences in the processing of metaphoric and literal meaning. Therefore, to ensure that a figurative effect emerged from the figurativeness of collocations rather than some item-related confound, we added a control, literal versus literal comparison to the experimental comparison of figurative and literal collocations[2] in this study (see Table 2).

Table 2 The four conditions of the experiment with examples

Comparison	Figurativeness	Condition	Examples
Experiment	Figurative Literal	Experiment–figurative Experiment–literal	He ***built a career (vs. built a house)*** in the city after graduating from university.
Control	Literal	Control–*figurative[a] Control–literal	He ***chose a career (vs. chose a house)*** in the city after graduating from university.

Note. [a]Control–*figurative indicates that the phrase in this condition was not figurative but literal. However, the same noun difference used for the experiment comparison was maintained in the control comparison.

It is important to note that the phrase in the control–*figurative condition (e.g., *choose a career*) is not figurative. We have used this nomenclature only as a convenient shorthand to describe that the same noun also appeared as a fig- urative item in the experiment condition. Thus, the control pair contained two literal VN collocations but had the same noun difference as the experimen- tal pair. We did this to control for variables unrelated to figurativeness such as word length, frequency, and concreteness without excluding figurativeness. For example, the experimental comparison between the figurative and literal collocations (e.g., *build a career* vs. *build a house*) might help to reveal that figurative collocations are processed more slowly than their literal counter- parts, but differences in word length (e.g., *house* is shorter than *career*), word frequency (e.g., *house* is more frequent than *career*), or in concreteness (e.g., *house* is a more concrete concept than *career*) are also possible variables caus- ing literal phrases such as *build a house* to be processed faster than figurative phrases such as *build a career*. The added control, literal versus literal pair *choose a career* versus *choose a house* had the same lexical differences of *ca- reer* versus *house* as the experimental pair, but did not have the same figurative versus literal difference. Therefore, if a slowdown for *career* relative to *house* identified in the experimental pair (i.e., in the context of "build a __") were greater than for that in the control pair (i.e., in the context of "choose a ___"), the difference could be linked only to the figurativeness in the experiment– figurative condition. Through this process, RT differences could be explained as figurativeness-related slowdown (i.e., the figurativeness effect). The figura- tiveness effect could therefore be quantified as the RT difference between the experimental figurative-literal difference *build a career* versus *build a house* and the control literal-literal difference *choose a career* versus *choose a house*. Accordingly, we adopted a two-by-two factorial design in this study in which we could identify a figurativeness effect via an interaction of the two variables: figurativeness with two levels (i.e., figurative, literal) and compari- son with two levels (i.e., experiment, control). If the difference in RTs between figurative and literal in the experiment comparison was significantly different from the difference in the control comparison, a significant Figurativeness × Comparison interaction would be present and would confirm a figurativeness effect.

Materials

There were 40 items of verb + noun collocations in the experiment and each item had four conditions (see Table 2). The collocations in the experiment–figurative condition were figurative but in the other three conditions (i.e., experiment–literal, control–*figurative, control–literal), the collocations were literal. We constructed the items using the following procedures. First, we consulted VN figurative collocations in the reference *Collins Cobuild Guides 7: Metaphor* (Deignan, 1995). Metaphors in this resource are all examples of contemporary English derived from the COBUILD database, a corpus comprising over 4.5 billion words. Some metaphors were adjacent collocations, so we could directly extract them in their VN form from their syntactic contexts (e.g., *Government grants have enabled a number of the top names in British sport to* ***build a*** *successful* ***career***); others were nonadjacent collocations (e.g., *Haig would have found this* ***advice*** *very hard to* ***swallow***), and we had to reconfigure them into adjacent VN phrases (e.g., ***swallow the advice***) before we used them as stimuli. Second, we consulted COCA (Davies, 2008) to find noun collocates in a window of ± 4 words for the lemmatized verb forms (e.g., for the lemma *BUILD*: *build/builds/built/building*) of the figurative collocations (e.g., *build a career*), whereby we checked concordance lines to identify literal VN collocation candidates (e.g., *build a house*). Third, we consulted COCA again to find lemmatized verb collocates in a window of ± 4 words for the nouns for both the identified figurative and literal phrases (e.g., *career/house*). Any verb collocate shared by the two nouns was selected for the two literal phrases in the control pair (e.g., *choose a career/house*).

The selected phrases had to fit the following criteria: (a) All lexical items constituting the collocations could be found in the vocabulary list of high school English in China to ensure that they could be understood by tertiary level Chinese participants; (b) figurative collocations should have a verb used in a metaphoric sense that would be transparent to Chinese participants (e.g., *build a career*) so that the meaning of each collocation could be inferred when its component words were known[3]; (c) literal collocations could be translated word for word into the participants' Chinese L1 with no loss of meaning (e.g., *build a house*) and figurative collocations would have a congruent mapping from the vehicle to the tenor between L1 to L2. For example, for the collocation *chew the words*, the mapping from the vehicle *words* to the tenor *food* works in both Chinese and English. This was done to minimize the possible confounding effects of L1–L2 congruency found in L2 processing of collocations (Wolter & Gyllstad, 2011, 2013; Wolter & Yamashita, 2018; Yamashita & Jiang, 2010).

We then embedded the four collocations in each item into a neutral and acceptable sentence after insertion of any necessary function items (i.e., *the, a, an, some, his*) between the verb and noun of the phrases. Thus, we created four versions of the same sentence differing in two (or more) words after the noun of the collocation if necessary to ensure the comparability of RT when we measured spillover effects (see the examples in Table 2). Furthermore, to make certain that the figurative collocation in each item conveyed a figurative meaning but that the other three did not, five university teachers, English L2 speakers living in mainland China, performed a figurativeness norming task. All five teachers were metaphor specialists and did not participate in the main experiment. The teachers judged the degree to which each phrase in a stimulus sentence was metaphoric on a scale from 1 (*least metaphoric*) to 7 (*most metaphoric*). The criteria that we employed in the stimulus selection were: (a) The mean metaphoricity of the supposedly figurative collocation should be at least 4 and (b) the metaphoricity gap between figurative and literal collocations in each item should be 2.6 or above[4] to ensure that there was a considerable figurativeness difference within each item. Following these criteria, we retained 40 items of collocations, 160 sentences in total, as the reading material (see Appendix S2 in the online Supporting Information; the test is also available in the IRIS Database, Shi et al., 2022a, and at https://osf.io/946vt/).

We counterbalanced the final 160 sentences across four presentation lists following a Latin square design. In addition, we added 60 filler sentences without verb + noun collocations to the four lists to reduce the participants' awareness of the experimental manipulation. Therefore, each list consisted of 100 sentences, 40 of which were experimental stimuli (10 figurative and 30 literal phrases) and 60 fillers. We divided the experimental sentences into two blocks, and each block contained the same number of stimuli per condition. We presented the sentences within each block pseudorandomly to the participants.

It should be noted that we did not control experimentally for item-related variables such as collocation frequency, collocation strength (mutual information), collocation length, semantic association strength, the naturalness of sentences, and the predictability of the final noun in the four conditions. This was for two reasons: (a) It was difficult to find a syntactic context suitable for an item matched in all four conditions, and (b) we considered all the item-related variables as covariates in the linear mixed-effects modeling so that their possible confounding effects could be controlled statistically. This type of statistical analysis, then, helps to eliminate the need for experimental control (Sonbul, 2015).

We checked the raw frequency and collocation strength (mutual information) scores of the 160 verbatim phrases as they appeared in the task with data obtained from COCA (Davies, 2008).[5] We checked the semantic associa- tion strengths of all the collocations with data obtained from the *University of South Florida Word Association, Rhyme, and Word Fragment Norms* (Nelson, McEvoy & Schreiber, 1998) to ensure that any processing effect was due to the figurative status of collocations rather than semantic associations. Specifically, we included the forward semantic association[6] as a covariate in the statistical analysis because we presented all collocations in the experiment in a forward direction (i.e., reading verb first), and it was more likely that the verb would prompt the noun of the collocations rather than that the noun would prompt the verb.

We assessed the naturalness of the collocations with a norming test in which we asked 24 English native speakers who did not participate in the main experiment to judge how natural each phrase sounded in the stimulus sentence on a Likert scale from 1 (*very unnatural*) to 7 (*very natural*). We presented the stimulus sentences in four counterbalanced lists (without fillers) and every participant judged the sentences in one of the four lists.

We measured the predictability of the final noun in each collocation with a sentence completion task in which each sentence was presented up to the noun of the collocation (e.g., "Smith would have found it very hard to swallow the __"). We asked 20 English and 20 Chinese native speakers who did not participate in the experiment to fill in the final noun of the collocations with the first word or phrase that came to mind.

Table 3 summarizes the descriptive statistics for the stimulus characteristics that we used as covariates in our study. We performed analyses of variance (ANOVAs) to check whether there were significant differences across the four conditions for each covariate. Table 4 presents the results of the ANOVAs. For the predictability ratings, native and nonnative speakers performed similarly in the ANOVA, and a low proportion of them had guessed the noun correctly in the four conditions.

Table 3 Summary of the descriptive statistics for the stimulus characteristics by the four conditions (40 items; 10 per condition)

	Experiment–figurative		Experiment–literal		Control–*figurative		Control-literal	
Characteristic	*M* (*SD*)	95% CI	*M* (*SD*)	95% CI	*M* (*SD*)	95% CI	*M* (*SD*)	95% CI
Database related and objective characteristics								
Frequency of whole phrase	33.9 (58.4)	[15.2, 52.6]	27.0 (41.9)	[13.6, 40.4]	63.6 (184.0)	[4.8, 122.4]	85.1 (235.0)	[9.9, 160.3]
Collocation strength (MI)	2.53 (1.98)	[1.90, 3.16]	7.35 (32.30)	[–2.90, 17.70]	2.45 (2.15)	[1.76, 3.14]	1.62 (2.02)	[0.97, 2.27]
Semantic association (FSG)	0.00 (0.00)	[0.00, 0.00]	0.063 (0.13)	[0.02, 0.10]	0.002 (0.01)	[–0.001, 0.005]	0.001 (0.005)	[0.00, 0.002]
Phrase length (no. of letters)	15.20 (3.19)	[14.20, 16.20]	13.40 (2.77)	[12.50, 14.30]	15.00 (3.90)	[13.80, 16.20]	13.40 (2.97)	[12.50, 14.3]
Nonparticipants' ratings for norming (7-point Likert scale)								
Naturalness: 24 NSs	4.82 (1.00)	[4.50, 5.14]	5.55 (0.88)	[5.27, 5.83]	5.31 (0.92)	[5.02, 5.60]	4.87 (1.00)	[4.55, 5.19]
Nonparticipants' responses for norming (proportion of nouns correct)								
Predictability: 20 NSs	0.06 (0.11)	[0.02, 0.10]	0.07 (0.12)	[0.03, 0.11]	0.03 (0.06)	[0.01, 0.05]	0.04 (0.13)	[0.00, 0.08]
Predictability: 20 NNSs	0.05 (0.10)	[0.02, 0.08]	0.03 (0.06)	[0.01, 0.05]	0.03 (0.06)	[0.01, 0.05]	0.04 (0.12)	[0.00, 0.08]
Study participants' ratings (7-point Likert scale)								
Familiarity: 40 NSs	6.20 (1.48)	[5.73, 6.67]	6.16 (1.60)	[5.65, 6.67]	6.02 (1.70)	[5.48, 6.56]	5.80 (1.83)	[5.21, 6.39]
Familiarity: 44 NNSs	5.67 (1.78)	[5.13, 6.21]	6.22 (1.43)	[5.79, 6.65]	6.20 (1.47)	[5.75, 6.65]	5.99 (1.65)	[5.49, 6.49]

Note. MI = mutual information; FSG = forward semantic association; NS = native speaker of English; NNS = nonnative speaker of Chinese.

Table 4 One-way ANOVA results comparing the four conditions for the stimuli characteristics

Characteristic	*F(3,156)*	*p*	η^2	95% CI
Database related and objective characteristics				
Frequency of whole phrase	1.24	.30	.02	[.00, .07]
Collocation strength (MI)	1.03	.38	.02	[.00, .07]
Semantic association (FSG)	9.53	**< .001**	.15	[.06, .25]
Phrase length (no. of letters)	3.58	**.02**	.06	[.00, .14]
Nonparticipants' ratings for norming				
Naturalness: 24 NSs	5.40	**.001**	.09	[.02, .18]
Nonparticipants' responses for norming				
Predictability: 20 NSs	1.15	.33	.02	[.00, .07]
Predictability: 20 NNSs	0.59	.62	.01	[.00, .05]
Study participants' ratings				
Familiarity: 40 NSs	4.78	**.00**	.01	[.00, .02]
Familiarity: 44 NNSs	11.16	**< .001**	.02	[.01, .03]

Note. Boldface indicates *F* values significant at alpha = .05. MI = mutual information; FSG = forward semantic association; NS = native speaker of English; NNS = nonnative speaker of Chinese.

Procedures

We administered a self-paced reading task using the DMDX software package (Forster & Forster, 2003) on desktop computers in a quiet language lab either in Hong Kong or in mainland China. DMDX is compatible with Windows10, which enabled us to time the presentation of material and the measurement of RTs to the stimuli with millisecond accuracy—most latencies were around 1 ms. We presented the stimulus sentences in a noncumulative, linear fash- ion using a moving window paradigm with black words (Times New Roman, 24 point) on a white background on a computer screen. All the words in a collocation appeared in the same line and none of the words appeared at a line break. We randomly assigned an equal numbers of participants to read the sentences in one of the four presentation lists based on the order in which they participated in the experiment. We asked them to read the sentences word by word on the computer screen by pressing a button (space bar) as fast as possible to indicate their comprehension. Every button press simultaneously revealed a word in the sentence while veiling the previous word with dashes. All the stimulus sentences and half of the filler sentences were followed by yes/no comprehension questions in which the questions for the stimulus sentences checked whether the participants had understood the VN collocations correctly (e.g., Stimulus: "He built a career in the city after graduating from university." Comprehension question: "Did he start and develop his career in the city after graduating from university?"). We instructed the participants to answer yes by pressing the right-hand shift key or no by pressing the left-hand shift key. After "Correct" or "Error" feedback was displayed on the screen, the participants could begin the next trial by pressing the designated space bar. At the beginning of the task, the participants performed eight practice trials to familiarize themselves with the procedure. When the participants had finished the 50 items in the first block, they could choose to rest before proceeding to the second block. The English native speakers completed the reading task in approximately 15–20 minutes, but Chinese native speakers required about 20–25 minutes.

After the experiment, all the participants completed a familiarity rating questionnaire, measuring the degree to which they were familiar with the collocations underlined in the experimental sentences. We prompted the participants to rate their familiarity in terms of whether they had seen the phrase before the experiment and whether they knew the meaning of the phrase on a Likert scale from 1 (*least familiar*) to 7 (*most familiar*). A summary of the descriptive statistics and the ANOVA results for the familiarity ratings appear in Tables 3 and 4. As mentioned, we then examined the influence of familiarity in the statistical analysis. Finally, we asked the Chinese participants to interpret five to six randomly chosen figurative collocations that they had read in the experiment; the results showed that they all had correctly interpreted these figurative expressions.

We conducted the statistical analysis with linear mixed-effects models using the lme4 package (Bates, Maechler, Bolker, & Walker, 2015) in the R statistical platform (R Core Team, 2020). We set the significance level at an alpha of .05 for the study. Our data and the R code that we used to analyze them are available in the IRIS Database (Shi, Peng, Li, 2022b) and at https://osf.io/946vt/.

Results

Both native and nonnative participant groups scored quite highly in comprehension accuracy (M_{L1} = 93.75%, 95% CI [86.01%, 100%]; M_{L2} = 88.58%, 95%CI [78.91%, 98.25%]) and no participants scored lower than 75%, showing that all the participants had focused on the task with good comprehension outcomes. We excluded those trials with incorrect answers to the comprehension questions from further analysis. A mixed-effects logistic regression showed that there were no significant differences in terms of errors across conditions for either native or nonnative speakers: native speakers, b = 22.71, 95% CI [14.75, 34.96], SE = 1.25, $\chi^2(3) = 3.70$, p = .30; nonnative speakers, b = 13.14, 95% CI [9.51, 18.18], SE = 1.18, $\chi^2(3) = 2.26$, p = .52. There was a very large difference in RTs for critical and postcritical words between native and nonnative speakers across conditions (see Figure 1).

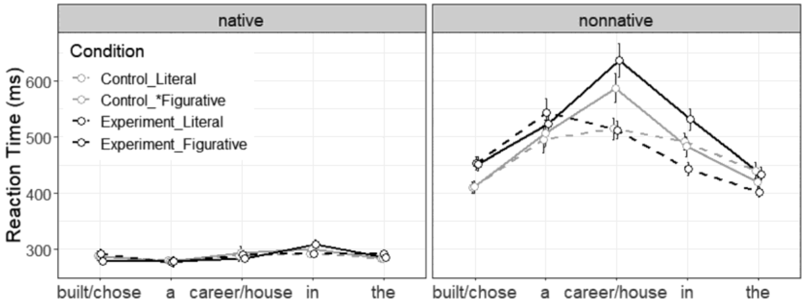

Figure 1 Native and nonnative mean reaction times (milliseconds) for correct trials in each condition. Error bars represent the standard errors calculated over participant means.

In each model, we kept the random-effects structure maximal (Barr, Levy, Scheepers, & Tily, 2013) and used deviation coding for the fixed-effect, categorical variables: figurativeness: literal –0.5 versus figurative 0.5; comparison: control –0.5 versus experiment 0.5; group: native –0.5 versus nonnative 0.5. The covariates that we considered in the model included collocational frequency, collocation strength (mutual information), forward semantic association strength, collocation length, the naturalness of collocations, the predictability of the final noun, RTs for the first and second words preceding the target word,[7] and the random order in which an item was presented. Before the analysis, we log-transformed the response variables—RTs for the critical word, the first and the second postcritical words—to reduce the positive skewness that frequently occurs for RTs (Baayen, Davidson, & Bates, 2008). We centered and standardized the continuous predictors (i.e., familiarity and proficiency) and all the covariates so that model coefficients would be interpretable in the presence of interactions.

We conducted a backward stepwise regression analysis, and we began the model selection by exploring the random structure with restricted maximum likelihood estimation. If a model had a boundary (singular) fit or failed to con- verge, we removed the random slope for interactions first and then the random slope for the main effects. We retained the model with the lowest Akaike in- formation criterion values in our model selection. Meanwhile, we calculated the variance inflation factor value of each predictor to ensure that there was no multicollinearity among the predictor variables. We considered regression coefficient estimates to be significant at our alpha of .05 if their *t* values were greater than 1.96 or less than –1.96.

We conducted three sets of mixed-effects models. The first set was omnibus models by which we examined the difference between native and nonnative speakers. We fit the second and third sets of models separately for the native and nonnative data. Because a spillover effect often occurs during self-paced reading, each set of models fit the RT for the critical word, that is, the noun, and the first and the second postcritical words of the collocations separately. We adopted minimal a priori data trimming to delete outliers combined with model criticism (see Baayen & Milin, 2010). First, we excluded the extremely short RTs of below 150 ms for nonnative and 100 ms for native RT data. Then, we performed model criticism after we had fit the omnibus models to remove overly influential outliers. We removed data points that had residuals at a distance of more than 2.5 standard deviations from the fitted line. Overall, the data-trimming procedures resulted in data exclusions from the L1 results at a rate of 3.07% for the critical word, 3.13% for the postcritical word, and 2.60% for the second postcritical word, whereas the L2 data saw exclusion rates at 4.75%, 4.49%, and 3.91%, respectively. Table 5 summarizes the results of the linear mixed-effects models for the critical fixed effects that the study concerned. Appendix S3 in the online Supporting Information provides the structure and complete output for the final omnibus models, the final native, and final nonnative models in Table S3.1, Table S3.2, and Table S3.3, respectively.

The omnibus models (upper panel of Table 5) showed that there was a significant Figurativeness × Comparison interaction for the first postcritical word, indicating that the RT difference between figurative and literal in the experi- ment comparison was significantly different from the difference in the control comparison. This indicated that a figurativeness effect emerged in the process- ing of collocations. More importantly, the significant three-way Figurativeness × Comparison × Group interaction revealed that the figurativeness effect was limited to one of the two groups and that there was thus a significant difference between native and nonnative speakers in the processing of figurative collocations.

Table 5 Summary of the critical fixed effects, marginal R^2, and conditional R^2 for the final omnibus and native and nonnative speaker models

	Critical				Critical+1				Critical+2			
Critical fixed effects	*b*	95% CI	*SE*	*t*	*b*	95% CI	*SE*	*t*	*b*	95% CI	*SE*	*t*
Omnibus models												
Figurativeness × Comparison	–0.02	[–0.07, 0.03]	0.03	–0.62	0.06	[0.02, 0.10]	0.02	**2.67**	0.01	[–0.03, 0.04]	0.02	0.32
Figurativeness × Comparison × Group	0.06	[–0.02, 0.15]	0.04	1.43	0.10	[0.02, 0.17]	0.04	**2.48**	0.06	[–0.01, 0.13]	0.03	1.79
Marginal R^2/Conditional R^2	.421 / .634				.346 / .595				.319 / .565			
Native speaker models												
Figurativeness × Comparison	–0.03	[–0.09, 0.02]	0.03	–1.12	0.01	[–0.04, 0.06]	0.03	0.44	–0.03	[–0.08, 0.02]	0.03	–1.03
Figurativeness × Comparison × Familiarity	–0.02	[–0.08, 0.03]	0.03	–0.94	–0.03	[–0.08, 0.01]	0.02	–1.35	–0.04	[–0.08, 0.01]	0.02	–1.60

(*Continued*)

Table 5 (Continued)

Critical fixed effects	Critical				Critical+1				Critical+2			
	b	95% CI	*SE*	*t*	*b*	95% CI	*SE*	*t*	*b*	95% CI	*SE*	*t*
Marginal R^2/Conditional R^2	.182 / .526				.177 / .573				.206 / .499			
Nonnative speaker models												
Figurativeness × Comparison	–0.01	[–0.09, 0.07]	0.04	–0.15	0.10	[0.04, 0.17]	0.03	**3.16**	0.05	[–0.00, 0.10]	0.03	1.81
Figurativeness × Comparison × Familiarity	0.02	[–0.06, 0.10]	0.04	0.51	0.02	[–0.04, 0.08]	0.03	0.62	–0.02	[–0.08, 0.03]	0.03	–0.92
Figurativeness × Comparison × Proficiency	–0.01	[–0.08, 0.06]	0.04	–0.18	–0.07	[–0.13, –0.01]	0.03	**–2.36**	–0.01	[–0.05, 0.04]	0.02	–0.25
Marginal R^2/Conditional R^2	.179 / .476				.085 / .334				.092 / .383			

Note. Regression coefficients were considered significant at alpha = .05 if their *t* values were > 1.96 or < –1.96. These significant *t* values are presented in boldface.

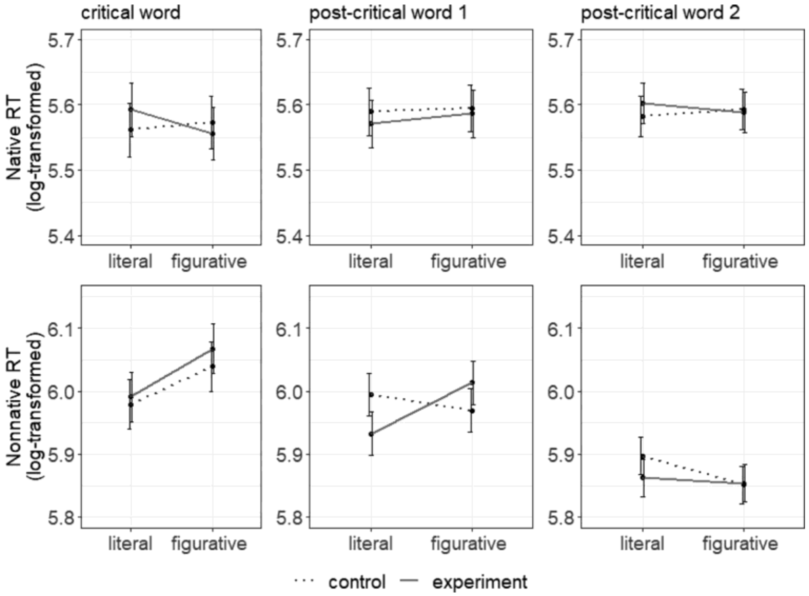

Figure 2 Mean reaction times (RT; log-transformed) as a function of figurativeness and comparison for the critical word and postcritical words in the target sentences. The upper three plots indicate English native speakers' mean RTs (log-transformed) for the critical word, the first and the second postcritical words, respectively. The lower three indicate the same plots for nonnative speakers' mean RTs (log-transformed). The upward-sloping solid lines show longer RTs for the figurative collocations compared to the literal controls in the experiment comparison, while the downward sloping dotted lines indicate shorter RTs for the control–*figurative collocations relative to the control–literal phrases in the control comparison. Error bars represent the standard errors calculated over participant means.

The native speaker models (middle panel of Table 5) further demonstrated that no figurativeness effect was generated in the L1 processing of figurative collocations: there was no significant Figurativeness × Comparison interaction in the processing of the critical word, the first postcritical word, and the second postcritical word. The model further showed that the RT difference between figurative and literal collocations in the experimental pair was even shorter (–8 ms) than the RT difference in the control pair, *b* = –0.03, 95% CI [–0.09, 0.02], *SE* = 0.03, *t* = –1.12, *p* = .132, when the critical word was processed (see Figure 2, top left). This suggested that native speakers seemed to process figurative collocations a little faster than they did literal ones, but the figurative processing advantage was not statistically significant.

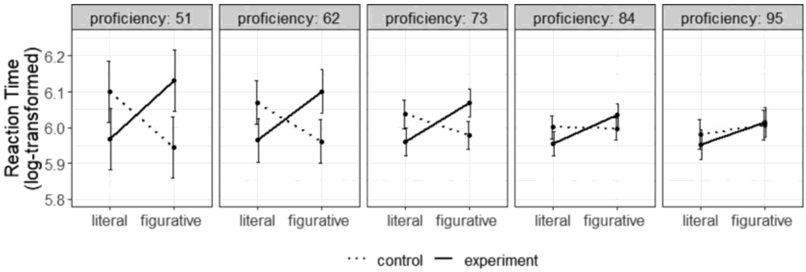

Figure 3 Nonnative mean reaction times (log-transformed) as a function of figurativeness, comparison, and proficiency for the first postcritical word in the target sentences. The five graphs represent the tendency of the way figurativeness effects (indicated by the Figurativeness × Comparison interaction) vary as second language proficiency grows from 51, 62, 73, and 84 to 95 (out of 100) points. The five proficiency points were selected to represent the five levels of second language proficiency: 50–59 ($n = 2$), 60–69 ($n = 3$), 70–79 ($n = 6$), 80–89 ($n = 15$), 90–99 ($n = 18$). Error bars show the standard errors for participant means.

Contrary to the L1 group, the L2 group showed a distinct processing pattern from the native English speakers (see Figure 2). The statistical analysis revealed that the Figurativeness × Comparison interaction was significant for the first postcritical word, but it was not significant in the processing of the critical word and the second postcritical word (lower panel of Table 5). This pattern was consistent with the result of the omnibus model, confirming that the figurativeness effect appeared after the critical word was processed. Fur- ther model statistics showed that the RT difference between figurative and lit- eral was about 42 ms longer in the experimental condition than in the control condition, $b = 0.10$, 95% CI [0.04, 0.17], $SE = 0.03$, $t = 3.16$, $p = .001$.

More important, there was a significant Figurativeness × Comparison × Proficiency interaction when the first postcritical word was processed, showing that L2 proficiency could modulate the figurativeness effect in the L2 processing of figurative collocations. Specifically, the negative *b* value and *t* value indicated that there was a negative correlation between the figurativeness effect and L2 proficiency. That is, the figurativeness effect disappeared gradually as L2 proficiency steadily increased (see Figure 3). However, we found no significant Figurativeness × Comparison × Familiarity interaction when the postcritical word was processed, indicating that subjective familiarity did not modulate the figurativeness effect in L2 collocational processing.

Discussion

Taking the native English speakers as the control group, in our study, we investigated whether the semantic property of figurativeness affects the L2 processing of collocations and to what extent the figurativeness effect could be modulated by familiarity and L2 proficiency if it occurred. The results of the self-paced experiment were very clear. Figurativeness, like L1–L2 congruency, has a measurable influence only on L2 collocational processing. More specifically, processing for figurative collocations compared with that for literal controls was slower when the nonnative group processed the first postcritical word, and L2 proficiency, but not subjective familiarity with the collocation, modulated the figurativeness effect. As such, the results confirmed Wray's (2002,

2008) position that there are differences between native and nonnative speakers in access to collocations. The figurative collocations required much more cognitive effort than did the literal collocations for the L2 speakers but posed no difficulty for the L1 speakers.

Figurativeness Effect: First Language Versus Second Language

The difference in the figurativeness effect between native and nonnative English speakers in our study suggests that the nonnative speakers adopted a distinct approach from that of the native speakers in processing collocations. The different approaches were, to a large extent, interpretable within the frameworks of the dual-route model (Van Lancker Sidtis, 2012; Wray, 2012). In this experiment, figurative collocations were basically more familiar to the native speakers than to the L2 speakers, which was also evidenced by the familiarity ratings (see Table 3), so when native speakers processed the figurative collocations, this more likely involved a faster mapping of components, whereas for L2 speakers, it was more of an online computation of novel figurative meanings.

In the course of L2 processing of figurative collocations, L2 speakers, as the literal salience model has described (Cies´licka, 2006), first rely on a literal analysis to access the noncompositional figurative meaning of collocations. The figurative computation happens only when the literal analysis fails or is determined to be inappropriate. That is, upon reading a verb of a VN collocation, L2 speakers first retrieve the verb's literal sense irrespective of its actual use. This suggests that a literal conceptual connection between verb and noun is anticipated in default of a figurative one. For example, *flowers* is more readily activated than *business* when the literal meaning of the verb *grow* has been accessed. The obligatory literal priority in L2 figurative pro- cessing (Cies´licka, 2006, 2017) forces L2 speakers to take an additional step of rejecting or suppressing the activated literal verb–noun conceptual connection *grow flowers* when an unexpected noun such as *business* emerges in the VN collocation *grow businesses*. Only at this point can the figurative computation begin and a metaphoric association between the verb *grow* and the noun *business* be established through comparison, that is, business can grow in the same way as flowers. This indirect figurative computation consumes more cognitive resources than does a direct literal analysis, as was reflected in the average 42 ms increase in RTs in the L2 processing of the first postcritical words of the figurative collocations.

In contrast, in the course of L1 processing, figurative collocations were not novel to the native speakers (see Table 3) and there was no priority of literalness obliging English native speakers to dwell on literal VN interpretations. Accordingly, when the verb *grow* of the VN collocation was encountered, both metaphoric and literal meanings were readily activated (McElree & Nordlie, 1999), and whether a figurative (*businesses*) or a literal (*flowers*) collocate followed made no difference for the native speakers. This is consistent with the dual-route model for formulaic language processing (Van Lancker Sidtis, 2012; Wray, 2012). The figurative computation terminates upon reaching the recognition point, and direct retrieval starts the instant that the final component or components are automatically activated at the recognition point (Cacciari & Tabossi, 1988). However, the temporal advantage of direct retrieval is not so clear for short, familiar idioms such as *kick the bucket*, in which the recognition point *bucket* coincides with the final component *bucket* of the idiom (Carrol & Conklin, 2014, p. 786). In our study, VN collocations such as *grow flowers/business* also had a late recognition point and could not be unequivocally recognized until the final word had been seen, whether that was *flowers* or *business* in this example. Once the final component *business* was encountered, it might have been activated as part of direct retrieval or as part of the computational analysis. Therefore, the late recognition point allowed L1 speakers to do a simultaneous dual-route activation of the figurative collocations, but it was not strong enough to give them a significant processing advantage for figura- tive collocations. This was potentially evidenced by the nonsignificant shorter RTs (–8 ms) for the critical word in the L1 speakers' processing of figurative collocations.

Familiarity

Contrary to the prediction, the statistical analysis showed that familiarity did not have any significant influence on the figurativeness effect generated in the L2 group. This result is not consistent with existing studies on collocations or other types of figurative language for which overall familiarity has been taken as a measure of subjective frequency (Columbus, 2013), or the main driver of formulaic processing (Tabossi et al., 2009; Van Lancker Sidtis, 2012). Given that the participants completed the familiarity rating after the main self-paced experiment and that the participants' subjective familiarity might have been affected by their having seen the expressions just prior to the rating, we repeated the familiarity rating with a separate group of 44 Chinese English-foreign- language learners (24 undergraduates and 20 postgraduates) and reinvestigated the impact of familiarity on the figurativeness effect. The result again indicated that familiarity did not modulate the figurativeness effect in L2 collocational processing (see Appendix S4 in the online Supporting Information).

Two possibilities may account for the absence of a familiarity effect. The first is related to the fact that most of the items in this experiment were quite familiar to the native and nonnative speakers. As Table 3 shows, the famil- iarity ratings in the L2 group were close to or above 6.0 on a 7-point scale. In addition, the greatest familiarity difference was found to be only 0.5 be- tween the most familiar (experiment–literal) collocations and the least familiar (experiment–figurative) ones. These results are a reflection of the fact that we deliberately chose the items in the study to be familiar to the L2 group and did not set out to create a sharp high versus low familiarity contrast. Therefore, more familiar collocations showed no advantage in RTs over the less familiar ones because the familiarity variation was too subtle to affect the RTs. This result mirrors the finding of Carrol and Conklin (2014) regarding familiarity effects on the processing advantage of idioms (either L1 idioms or translated L1-only idioms) over novel control phrases. In sum, it is likely that a famil- iarity effect would be observed if the collocations have a broader familiarity variation.

The second possibility is that familiarity might play a less important role in the L2 processing of congruent collocations. As we noted earlier, all the items in the study were congruent collocations, so the L2 speakers might have relied more on their L1 knowledge rather than on their experience in the L2 for their collocational processing. The absence of familiarity-induced facilitation in our study may lend support to the assertion of Carrol et al. (2016) that L1 knowledge has a greater influence than direct experience in the L2 on how congruent idioms are processed. In their eye-tracking study, Carrol et al. reported that L1 familiarity ratings accounted for a significant facilitation effect in L1–L2 congruent idiom processing, whereas L2 familiarity ratings did not. Furthermore, the meanings of figurative collocations, as used in this study, are more trans- parent than are those of idioms, so L2 speakers could rely on the component words of collocations to deduce their meanings (Wolter & Yamashita, 2018; Wray, 2002). Thus, it is very possible that the easier it is to compute a meaning from component words, the more likely L2 speakers are to rate a collocation as familiar. This possibility was confirmed by the familiarity ratings of literal col- locations by our L2 participants (see Table 3). Therefore, the familiarity effect needs further testing with incongruent collocations so that the influence of L1 knowledge and transparency might be factored out of the familiarity ratings.

Second Language Proficiency

The statistical analyses revealed that L2 proficiency had a negative association with the figurativeness effect, specifically, the figurativeness effect decreased with increased L2 proficiency. This result is consistent with the proficiency effect reported in many studies on collocations (Sonbul, 2015; Wolter & Gyllstad, 2011, 2013; Wolter & Yamashita, 2018; Yamashita & Jiang, 2010) and on figurative language (Siyanova-Chanturia et al., 2011).

The observed proficiency effect suggests the possibility of learners' gradually converging toward nativelike processing mechanism in the L2 processing of collocations, which is in line with the hypothesis of the dual-route model. That is, nonnative speakers of a language are likely to shift gradually from the computation approach to the direct retrieval approach once they encounter the L2 formulaic sequences with enough regularity or reach a certain level of proficiency (Carrol & Conklin, 2014). Specifically, with increased exposure to the L2, more proficient speakers might form stronger associative links be- tween a collocation's components and create configurations for the figurative collocations. Meanwhile, as the revised hierarchical model predicted (Kroll & Stewart, 1994), L2 speakers might become less dependent on the mediation of L1 translation but rely more on a direct link between the L2 lexicon and con- cepts to access the meaning of the target noun of a collocation as proficiency increases. A stronger word-concept link in the L2 could further facilitate faster mapping of the components of a collocation. Thus, the simultaneous dual-route activations displayed in the native group could also take place in the nonnative group, which would diminish the differences between the processing of literal and figurative phrases. It follows that low-proficiency L2 speakers might have less developed associative links between the components of figurative colloca- tions due to limited L2 exposure or knowledge. This leaves them the computa- tional route as the most available option, forcing low-proficiency L2 speakers through the literal and indirect access of figurative language (Cies´licka, 2017; Clark & Lucy, 1975).

Although L2 proficiency plays a role, the significant figurativeness ef- fect that we found indicates that a direct shift from literal computation to simultaneous activation is unlikely to happen for L2-dominant learners, especially for those who learn their L2 in a foreign language learning environment (Cies´licka, 2017). As Figure 3 illustrates, the Figurativeness × Comparison interaction—the figurativeness effect—did not disappear until L2 proficiency reached 95 points out of 100, which is in line with the finding by Werkmann Horvat, Bolognesi, and Kohl (2021) that figurative language, even very conventional, has a special status in the L2 mental lexicon. In this experiment, however, only five participants surpassed this proficiency point, and all were doctoral students specializing in translation and interpreting. Undoubt- edly, these five participants were all equipped with strong bilingual knowledge and competence and may have had a stronger motivation for identifying equiv- alent counterparts for those figurative terms than the average participant be- cause translation and interpreting require advanced bilingual comprehension and production. Over time, they could have developed a relatively strong asso- ciation between components of figurative collocations, in a way comparable to the figurative configuration activation of native speakers.

Limitations and Future Research

There are inevitably some limitations to our study that need to be addressed in the future. First, our study was limited in sample size (44 L2 participants and 40 L1 participants), with just 1,600 observations (400 per condition) for the L1 group and 1,760 observations for the L2 group (440 per condition). Although the post hoc simulation-based power analyses (Kumle, Võ, & Draschkow, 2021) showed that the statistical power for the figurativeness effect was almost 95% in the L2 group, the difference in figurativeness effect between L1 and L2 speakers could be detected with only around 80% probability (see Appendix S5 in the online Supporting Information). Future researchers could double the number of observations to achieve a more reliable result for the difference between L1 and L2 processing of figurative collocations. Moreover, we examined only the figurativeness of L1–L2 congruent collocations and did not reveal a role of familiarity on the figurativeness effect. For future research, it would be interesting to test this result with incongruent collocations of greater familiarity variation to explore how and to what extent phrase familiarity can modulate the figurativeness effect produced in L2 speakers' processing of collocations.

Lastly, we adopted a difference-of-differences approach to quantify the figurativeness effect, and the added control, literal versus literal pair increased the difficulty of finding a sentence context suitable for an item in four conditions.

In many cases, the literal collocations in the control pair differed considerably in covariate properties from the literal collocations in the experiment condi- tion (see Table 3). This may have led to RTs for figurative collocations in the experiment-figurative condition being longer than those for the literal collo- cations in the experiment–literal condition, while not affecting the two literal collocation RTs in the control pair (judging from the error bars in Figure 2, Panel 5). However, the tightly controlled design provided stronger evidence for a figurative influence at the collocational level, which suggests that figu- rativeness matters for L2 speakers irrespective of whether it is in transparent and decomposable collocations or in somewhat opaque and nondecomposable idioms.

Conclusion

Figurativeness poses a similar challenge for L2 processing and acquisition in collocations as it does in idioms. As a result, figurative collocations should re- ceive particular attention in L2 teaching and learning, especially in a foreign language learning setting. For L2 educators, it would be helpful to raise learn- ers' awareness of the figurative thinking behind collocations (Littlemore & Low, 2006) and to help learners build form–meaning connections essential for language use at both the word and collocation levels (Jiang, 2000, 2021). For L2 learners, focusing on entire combinations of figurative collocations rather than on lexical components seems indispensable for strengthening links be- tween collocational forms and figurative concepts while the learners build their L2 experiences and grow more proficient in the target language.

Open Research Badges

This article has earned Open Data and Open Materials badges for making pub- licly available the digitally-shareable data and the components of the research methods needed to reproduce the reported procedure and results. All data and materials that the authors have used and have the right to share are available at https://osf.io/946vt and http://www.iris-database.org. All proprietary materials have been precisely identified in the manuscript.

Notes

1 C-test has been proved to be a highly reliable and valid instrument for measuring general L2 proficiency across the language skills of vocabulary, grammar, reading, and contextual use (Dörnyei & Katona, 1992; Eckes & Grotjahn, 2006). We borrowed the two C-test passages from a 2019 C-test practice platform (https://wuster.uab.es/ctestpracticer/main?x = en) named *UAB Idioms*.

2 People usually control for stimulus-related confounding effects by making well- balanced stimuli. However, it is difficult to find perfectly balanced noun collocates, therefore, the control, literal versus literal pairs were included to control for potential noun-related confounding effects.

3 We asked three English native-speaker doctoral students specializing in linguistics to judge whether or not the figurative expressions were idioms according to Howarth's definitions of figurative idioms and pure idioms. We removed phrases if two raters judged them to be idiomatic. The result showed that no phrase was viewed as an idiom by two raters.

4 Because the rating was based on a 7-point Likert Scale and 2.6 was slightly greater than one-third of 7 (2.4), we set the cut-off point at 2.6 to ensure that there was

more than one-third of a figurativeness difference between the metaphoric and literal combinations for each item.

5 We used the nonlemmatized frequency of the verbatim phrases (total number of occurrences in the corpus) because no clear difference has been found between lemmatized and nonlemmatized frequencies for predicting nonnative speakers' knowledge of collocations (Durrant, 2014, p. 465).

6 Forward semantic association values are the proportion of individuals in a group who produce a particular target in the presence of the cue word.

7 For the critical words *career/house*, the two preceding words are the verbs *built/chose* and the inserted word *a*; for the first postcritical word *in*, it is the inserted word *a* and the critical word *career/house*; for the second postcritical word *the*, it is the critical word *career/house* and the first postcritical word *in*.

References

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of memory and language, 59,* 390–412. https://doi.org/10.1016/j.jml.2007.12.005

Baayen, R. H., & Milin, P. (2010). Analyzing reaction times. *International Journal of Psychological Research, 3,* 12–28. https://doi.org/10.21500/20112084.807

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68,* 255–278. https://doi.org/10.1016/j.jml.2012.11.001

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

Cacciari, C., & Tabossi, P. (1988). The comprehension of idioms. *Journal of Memory and Language, 27*, 668–683. https://doi.org/10.1016/0749-596X(88)90014-9

Cameron, L. (2003). *Metaphor in educational discourse*. London, UK: Continuum.

Cardillo, E. R., Schmidt, G. L., Kranjec, A., & Chatterjee, A. (2010). Stimulus design is an obstacle course: 560 matched literal and metaphorical sentences for testing neural hypotheses about metaphor. *Behavior Research Methods, 42*, 651–664. https://doi.org/10.3758/BRM.42.3.651

Carrol, G., & Conklin, K. (2014). Getting your wires crossed: Evidence for fast processing of L1 idioms in an L2. *Bilingualism: Language and Cognition*, *17,* 784–797. https://doi.org/10.1017/S1366728913000795

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

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

Cies´licka, A. B. (2006). Literal salience in online processing of idiomatic expressions by second language learners. *Second Language Research*, *22,* 115–44. https://doi.org/10.1191/0267658306sr263oa

Cies´licka, A. B. (2017). Bilingual figurative language processing. In A. Ardila, A. B. Cies´licka, R. R. Heredia, & M. Rosselli (Eds.), *Psychology of bilingualism: The cognitive and emotional world of bilinguals* (pp. 75–118). Cham, Switzerland: Springer. https://doi.org/10.1007/978-3-319-64099-0_4

Clark, H. H., & Lucy, P. (1975). Understanding what is meant from what is said: A study in conversationally conveyed requests. *Journal of Verbal Learning and Verbal Behavior*, *14*, 56–72. https://doi.org/10.1016/S0022-5371(75)80006-5

Columbus, G. (2013). In support of multiword unit classifications: Corpus and human rating data validate phraseological classifications of three different multiword unit types. *Yearbook of Phraseology*, *4*, 23–43. https://doi.org/10.1515/phras-2013-0003

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. https://doi.org/10.1093/applin/amm022

Davies, M. (2008). *The Corpus of Contemporary American English*. Retrieved from https://www.english-corpora.org/coca/

Deignan, A. (1995). *Collins Cobuild English guides 7: Metaphor*. London, UK: HarperCollins.

Deignan, A. (2005). *Metaphor and corpus linguistics*. Amsterdam, The Netherlands: John Benjamins.

Dörnyei, Z., & Katona, L. (1992). Validation of the C-test amongst Hungarian EFL learners. *Language Testing*, *9*, 187–206. https://doi.org/10.1177/026553229200900206

Durrant, P. (2014). Corpus frequency and second language learners' knowledge of collocations: A meta-analysis. *International Journal of Corpus Linguistics*, *19*, 443–477. https://doi.org/10.1075/ijcl.19.4.01dur

Durrant, P., & Doherty, A. (2010). Are high-frequency collocations psychologically real? Investigating the thesis of collocational priming. *Corpus Linguistics and Linguistic Theory*, *6*, 125–311. https://doi.org/10.1515/cllt.2010.006

Eckes, T., & Grotjahn, R. (2006). A closer look at the construct validity of C-tests. *Language Testing*, *23*, 290–325. https://doi.org/10.1191/0265532206lt330oa

Ellis, N. C., Frey, E., & Jalkanen, I. (2009). The psycholinguistic reality of collocation and semantic prosody (1): Lexical access. In U. Römer & R. Schulze (Eds.), *Exploring the lexis-grammar interface* (pp. 89–114). Amsterdam, The Netherlands: John Benjamins.

Forster, K. I., & Forster, J. C. (2003). DMDX: A Windows display program with millisecond accuracy. *Behavior Research Methods, Instruments, & Computers*, *35*, 116–124. https://doi.org/10.3758/BF03195503

Frisson, S., Rayner, K., & Pickering, M. J. (2005). Effects of contextual predictability and transitional probability on eye movements during reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*, 862–877. https://doi.org/10.1037/0278-7393.31.5.862

Granger, S., & Bestgen, Y. (2014). The use of collocations by intermediate vs. advanced non-native writers: A bigram-based study. *International Review of Applied Linguistics in Language Teaching*, *52*, 229–252. https://doi.org/10.1515/iral-2014-0011

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

Hill, J., Lewis, M., & Lewis, M. (2000). Classroom strategies, activities and exercises. In M. Lewis (Ed.), *Teaching collocation: Further developments in the lexical approach* (pp. 86–117). Hove, UK: Language Teaching Publications.

Hoey, M. (2005). *Lexical priming: A new theory of words and language*. London, UK: Routledge.

Howarth, P. A. (1996). *Phraseology in English academic writing*. Tübingen, Germany: Max Niemayer Verlag.

Howarth, P. A. (1998). Phraseology and second language proficiency. *Applied Linguistics, 19*, 24–44. https://doi.org/10.1093/applin/19.1.24

Jankowiak, K., Rataj, K., & Naskręcki, R. (2017). To electrify bilingualism: Electrophysiological insights into bilingual metaphor comprehension. *PLOS ONE, 12*(4), e0175578. https://doi.org/10.1371/journal.pone.0175578

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

Jiang, N. (2021). Orthographic friends and lexical strangers in the L2 lexicon. *Journal of Second Language Studies*, *4*, 224–244. https://doi.org/10.1075/jsls.21012.jia

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

Kumle, L., Võ, M. L.-H., & Draschkow, D. (2021). Estimating power in (generalized) linear mixed models: An open introduction and tutorial in R. *Behavior Research Methods*, *53*, 2528–2543. https://doi.org/10.3758/s13428-021-01546-0

Laufer, B., & Waldman, T. (2011). Verb-noun collocations in second language writing: A corpus analysis of learners' English. *Language learning*, *61*, 647–672. https://doi.org/10.1111/j.1467-9922.2010.00621.x

Libben, M. R., & Titone, D. A. (2008). The multidetermined nature of idiom processing. *Memory & Cognition*, *36*, 1103–1121. https://doi.org/10.3758/MC.36.6.1103

Littlemore, J., & Low, G. (2006). *Figurative thinking and foreign language learning*. New York, NY: Palgrave Macmillan.

Macis, M., & Schmitt, N. (2017a). The figurative and polysemous nature of collocations and their place in ELT. *ELT Journal*, *71*, 50–59. https://doi.org/10.1093/elt/ccw044

Macis, M., & Schmitt, N. (2017b). Not just "small potatoes": Knowledge of the idiomatic meanings of collocations. *Language Teaching Research*, *21*, 321–340. https://doi.org/10.1177/1362168816645957

McDonald, S. A., & Shillcock, R. C. (2003). Eye movements reveal the on-line computation of lexical probabilities during reading. *Psychological Science*, *14*, 648–652. https://doi.org/10.1046/j.0956-7976.2003.psci_1480.x

McElree, B., & Nordlie, J. (1999). Literal and figurative interpretations are computed in equal time. *Psychonomic Bulletin & Review*, *6*, 486–494. https://doi.org/10.3758/BF03210839

Nelson, D. L., McEvoy, C. L., & Schreiber, T. A. (1998). *The University of South Florida word association, rhyme, and word fragment norms*. Retrieved from http://w3.usf.edu/FreeAssociation/

Nesselhauf, N. (2005). *Collocations in a learner corpus*. Amsterdam, The Netherlands: John Benjamins.

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

Paulmann, S., Ghareeb-Ali, Z., & Felser, C. (2015). Neurophysiological markers of phrasal verb processing: Evidence from L1 and L2 speakers. In R. Heredia & A. Cieślicka (Eds.), *Bilingual figurative language processing* (pp. 245-267). Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9781139342100.013

Pulido, M., & Dussias, P. (2020). Desirable difficulties while learning collocations in a second language: Conditions that induce L1 interference improve learning.*Bilingualism: Language and Cognition*, *23*, 652–667. https://doi.org/10.1017/S1366728919000622

R Core Team. (2020). R: A language and environment for statistical computing (Version3.6.3) [Computer software]. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from http://www.R-project.org

Shi, J., Peng, G., & Li, D. (2022a). *Materials for word test and self-paced reading test. Materials from "Figurativeness matters in the second language processing of collocations: Evidence from a self-paced reading experiment"* [Language test]. IRIS Database, University of York, UK. (https://doi.org/10.48316/1fnq-9r89)

Shi, J., Peng, G., & Li, D. (2022b). *Data and R code. Datasets from "Figurativeness matters in the second language processing of collocations: Evidence from a self-paced reading experiment"* [Collection: Analysis code with data]. IRIS Database, University of York, UK. https://doi.org/10.48316/2fhh-3s66

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

Siyanova-Chanturia, A., Conklin, K., & Schmitt, N. (2011). Adding more fuel to the fire: An eye-tracking study of idiom processing by native and non-native speakers. *Second Language Research*, *27*, 251–272. https://doi.org/10.1037/a0022531

Siyanova-Chanturia, A., & Martinez, R. (2015). The idiom principle revisited. *Applied Linguistics, 36*, 549–569. https://doi.org/10.1093/applin/amt054

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

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. https://doi.org/10.1017/S1366728914000674

Steen, G., Dorst, A., Herrmann, J., Kaal, A. & Krennmayr, T. (2010). Metaphor in usage. *Cognitive Linguistics*, *21*, 765–796. https://doi.org/10.1515/cogl.2010.024

Tabossi, P., Fanari, R., & Wolf, K. (2009). Why are idioms recognized fast? *Memory & Cognition*, *37*, 529–540. https://doi.org/10.3758/MC.37.4.529

Van Lancker Sidtis, D. (2012). Two-track mind: Formulaic and novel language support a dual-process model. In M. Faust (Ed.), *The handbook of the neuropsychology of language* (pp. 342–367). Chichester, UK: Wiley-Blackwell.

Vilkaitė, L. (2016). Are nonadjacent collocations processed faster? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *42*, 1632–1642. https://doi.org/10.1037/xlm0000259

Vilkaitė, L., & Schmitt, N. (2019). Reading collocations in an L2: Do collocation processing benefits extend to non-adjacent collocations? *Applied Linguistics*, *40*, 329–354. https://doi.org/10.1093/applin/amx030

Werkmann Horvat, A., Bolognesi, M., & Kohl, K. (2021). The status of conventional metaphorical meaning in the L2 lexicon. *Intercultural Pragmatics*, *18*, 447–467. https://doi.org/10.1515/ip-2021-4002

Werkmann Horvat, A., Bolognesi, M., & Lahiri, A. (2021). Processing of literal and metaphorical meanings in polysemous verbs: An experiment and its methodological implications. *Journal of Pragmatics*, *171*, 131–146. https://doi.org/10.1016/j.pragma.2020.10.007

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

https://doi.org/10.1093/applin/amr011
Wolter, B., & Gyllstad, H. (2013). Frequency of input and L2 collocational processing: A comparison of congruent and incongruent collocations. *Studies in Second Language Acquisition*, *35*, 451–482. https://doi.org/10.1017/S0272263113000107
Wolter, B., & Yamashita, J. (2015). Processing collocations in a second language: A case of first language activation? *Applied Psycholinguistics*, *36*, 1193–1221. https://doi.org/10.1017/S0142716414000113
Wolter, B., & Yamashita, J. (2018). Word frequency, collocational frequency, L1 congruency, and proficiency in L2 collocation processing: What accounts for L2 performance. *Studies in Second Language Acquisition*, *39*, 1–22. https://doi.org/10.1017/S0272263117000237
Wray, A. (2002). *Formulaic language and the lexicon*. Cambridge, UK: Cambridge University Press.
Wray, A. (2008). *Formulaic language: Pushing the boundaries*. Oxford, UK: Oxford University Press.
Wray, A. (2012). What do we (think we) know about formulaic language? An evaluation of the current state of play. *Annual Review of Applied Linguistics*, *32*, 231–254. https://doi.org/10.1017/S026719051200013X
Yamashita, J. (2018). Possibility of semantic involvement in the L1-L2 congruency effect in the processing of L2 collocations. *Journal of Second Language Studies, 1*, 60–78. https://doi.org/10.1075/jsls.17024.yam
Yamashita, J., & Jiang, N. (2010). L1 influence on the acquisition of L2 collocations: Japanese ESL users and EFL learners acquiring English collocations. *TESOL Quarterly*, *44*, 647–668. https://doi.org/10.5054/tq.2010.235998

Supporting Information

Additional Supporting Information may be found in the online version of this article at the publisher's website:

Appendix S1. C-Test and Answers.
Appendix S2. Materials for the Self-Paced Reading Experiment.
Appendix S3. Structure and Complete Output for the Final Models.
Appendix S4. Familiarity Effects Retested.
Appendix S5. Simulation-Based Power Analysis.

Appendix: Accessible Summary (also publicly available at https://oasis-database.org)

Figurativeness matters in second language processing of collocations

What This Research Was About and Why It Is Important

Figurativeness is an important semantic phenomenon that is pervasive in idioms, collocations, and phrasal verbs. Although much is known about how second language (L2) learners process figurative language, most researchers explored this topic with idioms (e.g., *kick the bucket*) or phrasal verbs (e.g., *run over*). In this study, we addressed the intriguing case of figurativeness in collocations. We explored, using a self-paced reading experiment, whether and how figurativeness might affect the L2 processing of collocations. Results show that figurativeness matters for L2 speakers in that it slows them down compared to literal phrases, irrespective of whether the figurativeness occurs in transparent and decomposable collocations or in somewhat opaque and nondecomposable idioms.

What the Researchers Did

- Participants were 40 English native speakers and 44 Chinese-speaking English foreign language learners (including doctoral, postgraduate, and undergraduate students). A 50-item C-test was used to measure the English foreign language learners' proficiency.
- The participants read (1) sentences with figurative collocations (e.g., *build a career*), (2) sentences with literal collocations (e.g., *build a house*), and (3) control sentences with a literal collocation pair (e.g., *choose a career* vs. *choose a house*). The added literal versus literal control pair had the same lexical differences (i.e., *career* vs. *house*) as the experimental pair, but did not have the same figurative versus literal difference.
- Participants read the sentences word by word in a self-paced manner, using the space bar to proceed through the sentence. Their response times for the critical words—the noun in the verb + noun collocations, and for the first and the second words after that were compared across participant groups and item types.
- The researchers examined the effect of figurativeness on reading times and explored whether the figurativeness effect differed as a result of the participants' subjective familiarity with the collocations and their L2 proficiency.

What the Researchers Found

- English foreign language learners processed figurative collocations more slowly than literal collocation controls. In contrast, English native speakers processed figurative and literal phrases with similar ease.
- More proficient L2 speakers slowed down less when processing figurative collocations than less proficient L2 speakers did.
- The participants' familiarity with the collocation did not influence their reading times.

Things to Consider

- L2 speakers may compute novel figurative meanings in real time during processing when the literal analysis of the collocation fails or is determined to be inappropriate. In comparison, native speakers are more likely to engage in a faster mapping of the different components of a figurative collocation during processing.
- L2 speakers are likely to shift gradually from the computation approach to the direct retrieval approach once they encounter the L2 formulaic sequences with enough regularity or reach a certain level of proficiency.
- Results were limited to L1–L2 congruent collocations with which L2 participants were quite familiar, and the role of familiarity could be reexamined with incongruent items in a future investigation.

Materials and data: Materials and data are publicly available at IRIS (www.iris-database.org) and OSF (https://osf.io/946vt/).

How to cite this summary: Shi, J. (2022). Figurativeness matters in second language processing of collocations. *OASIS Summary* of Shi et al. (2023) in *Language Learning*. https://oasis-database.org