EMPIRICAL STUDY

Collocational Processing in L1 and L2: The Effects of Word Frequency, Collocational Frequency, and Association

Doğuş Öksüz [ID],[a,b] Vaclav Brezina,[a] and Patrick Rebuschat [ID][a,c]

[a]Lancaster University, [b]University of Leeds, and [c]University of Tübingen

This study investigated the effects of individual word frequency, collocational frequency, and association on L1 and L2 collocational processing. An acceptability judgment task was administered to L1 and L2 speakers of English. Response times were analyzed using mixed-effects modeling for 3 types of adjective–noun pairs: (a) high-frequency, (b) low-frequency, and (c) baseline items. This study extends previous research by examining whether the effects of individual word and collocation frequency counts differ for L1 and L2 speakers' processing of collocations. This study also compared the extent to which L1 and L2 speakers' response times are affected by mutual information and log Dice scores, which are corpus-derived association measures. Both groups of participants demonstrated sensitivity to individual word and collocation frequency counts. However, there was a reduced effect of individual word frequency counts

However, there was a reduced effect of individual word frequency counts

for processing high-frequency collocations compared to low-frequency collocations. Both groups of participants were similarly sensitive to the association measures used.

Keywords collocation; multiword sequences; collocational processing; advanced learners; association measures; mutual information; log Dice

Introduction

There has been a growing interest in research dedicated to the processing and use of multiword sequences. For language acquisition and processing specifically, the importance of multiword sequences has been highlighted by usage-based approaches to language acquisition that have been gaining prominence (Bannard, Lieven, & Tomasello, 2009; Christiansen & Chater, 2016; Tomasello, 2003). Within usage-based approaches, linguistic productivity is seen as a gradually emerging process of storing and abstracting multiword sequences (e.g., Goldberg, 2006; McCauley & Christiansen, 2017; Tomasello, 2003). Such perspectives view both single words and multiword sequences as essential building blocks for language acquisition and processing (Christiansen & Chater, 2016; Goldberg, 2006; McCelland, 2010). These approaches have received considerable empirical support from corpus studies that have reported that large numbers of multiword sequences are used in both spoken and written language. For example, Jackendoff (1997) carried out a small-scale corpus analysis of utterances used in a TV show and showed that multiword sequences were used as often as single words in daily language. DeCock, Granger, Leech, and McEnery (1998) estimated that multiword sequences constitute up to 50% of both written and spoken native-speaker discourse. In addition to corpus evidence, there is substantial psycholinguistic evidence that both children (Arnon & Clark, 2011; Bannard & Matthews, 2008) and adults (Arnon & Snider, 2010; Jolsvai, McCauley, & Christiansen, 2013; Tremblay, Derwing, Libben, & Westbury, 2011) are sensitive to multiword sequences during comprehension and production tasks. Furthermore, both first language (L1) and second language (L2) speakers appear to process multiword sequences faster than matched control phrases (Siyanova-Chanturia, Conklin, & van Heuven, 2011; Wolter & Yamashita, 2018).

Multiword sequences include collocations, lexical bundles, binomials, and idioms. Despite their obvious similarities, they vary considerably in completeness, structure, length, and transparency of meaning. Many studies have found that idioms (e.g., *kick the bucket*) are processed faster than matched control phrases (e.g., Rommers, Dijkstra, & Bastiaansen, 2013; Siyanova-Chanturia, Conklin, & Schmitt, 2011; Vespignani, Canal, Molinaro, Fonda, & Cacciari,

2010). The same effect holds for lexical bundles (e.g., *in the middle of the*), which have been defined as sequences of three or four words that occur as wholes at least 10 times per million words (Biber et al., 1999). Tremblay et al. (2011) showed a processing advantage for lexical bundles in a self-paced reading experiment. Similar results have been reported for binomials, which are phrases consisting of two content words of the same class with a conjunction between them (e.g., *knife and fork*). Siyanova-Chanturia, Conklin, and van Heuven (2011), using eye-tracking, found that the original form of a binomial is processed faster than its reversed form (e.g., *fork and knife*) by both L1 and L2 speakers. These findings provided empirical evidence that multiword sequences are processed faster than matched novel phrases due to their phrasal frequency and predictability. This is consonant with usage-based approaches to language acquisition (Barlow & Kemmer, 2000; Christiansen & Chater, 2016; Ellis, 2002; Tomasello, 2003). These approaches have underscored that language is rich with various types of distributional information such as frequency, variability, and co-occurrence probability and that the human mind is sensitive to such distributional information (Erickson & Thiessen, 2015). For the processing and acquisition of multiword sequences specifically, two types of statistical information play an important role, namely frequency and association (Gries & Ellis 2015; Yi, 2018).

A prominent type of multiword sequence that has received special attention in psycholinguistics, corpus linguistics, and language education studies is collocations. Different approaches to operationalizing the complex notion of collocations have been put forth (McEnery & Hardie, 2012, pp. 122–123). The two most widely known approaches are the phraseological approach, and the distributional or frequency-based approach. The phraseological approach focuses on the semantic relationship between two or more words and the degree of noncompositionality of their meaning (Howarth, 1998; Nesselhauf, 2005). According to the phraseological approach, collocations are not simply free combinations of semantically transparent words; they follow some selection restrictions (e.g*., slash one's wrist* rather than *cut one's wrist*). The frequency-based approach draws on quantitative evidence of word co-occurrence in corpora (Evert, 2008; Gablasova, Brezina, & McEnery, 2017; McEnery & Hardie, 2012; Paquot & Granger, 2012), from which collocations are extracted using frequency cutoff scores and collocational association measures (see Evert, 2008; Gablasova et al., 2017, for a review of association measures). In this study, we adopted a frequency-based approach because we were primarily concerned with the effects of single word frequency, collocational frequency, and collocational strength on processing collocations by L1 and L2 speakers.

Background Literature
Variables Affecting Collocational Processing

An important question is whether high-frequency collocations are a psychological reality for L1 and L2 speakers. Siyanova and Schmitt (2008) explored collocational processing by L1 and L2 speakers of English using a variation of an acceptability judgment task, finding that participants responded to high-frequency collocations faster than to noncollocations. Durrant and Doherty (2010) conducted lexical decision tasks with L1 speakers to investigate whether high collocation frequency or semantic association between the collocates led to faster processing of adjective–noun collocations. They found a priming effect in the processing of very high-frequency collocations, even if the collocates were not semantically associated. Wolter and Gyllstad (2013) looked at congruency effects on collocational processing for L2 speakers and collocational frequency effects for both L1 and L2 speakers of English. They showed that their participants processed collocations faster in a L2 if the collocations were congruent (i.e., a translation equivalent existed in the participants' L1). Furthermore, both L1 and advanced L2 speakers were sensitive to collocation frequency because they responded faster to more frequent collocations than to less frequent collocations. Wolter and Yamashita (2015) developed this idea by examining whether collocations that exist in participants' L1 (Japanese) but not in the L2 (English) were still facilitated when participants processed collocations translated into the L2. They found no facilitation effect for such translations.

More recently, Wolter and Yamashita (2018) investigated the congruency effect for L2 speakers. In addition, they examined single-word and collocational frequency effects for L1 and L2 speakers' processing of adjective–noun collocations. Replicating previous findings, they found a processing advantage for congruent collocations for L2 speakers and no facilitation effect for the L1-only collocations translated into the L2. They suggested that the learners' age or the order in which something is learned affect how deeply a collocation becomes entrenched in the language system, helping to explain the discrepancy between processing congruent and incongruent collocations. As learners gain L2 experience, the transferred congruent collocations from L1 to L2 become more entrenched through repeated exposure and the nontransferable incongruent collocations become less entrenched due to lack of reinforcement. They also found that processing by both L1 English and advanced L2 groups was affected by word-level frequency and collocational frequency simultaneously, showing that the processing of L2 learners with advanced proficiency

and of L1 speakers was affected by frequency information at multiple levels of representation.

Eye-movement studies have also looked at L1 and L2 collocational processing. For example, Sonbul (2015) conducted a study that included L1 and L2 English speakers' on-line (eye-movement) and off-line (rating) measures of collocational processing. She developed three types of adjective–noun pairs: high-frequency collocations (e.g., *fatal mistake*), low-frequency (e.g., *awful mistake*), and nonattested synonymous pairs (e.g., *extreme mistake*). She examined how collocational frequency affects processing, finding that both L1 and L2 speakers are sensitive to collocation frequency in early measures of eye-movements but not late measures. Thus she suggested that collocations are not entirely fixed phrases. When reading an unexpected word pair, her readers initially needed longer time to process the pair, but once they had incorporated it into a more general adjective–noun schema, they were able to process nonattested phrases comparably fast. Vilkaité (2016) looked at adult L1 speakers' eye-movements to test whether nonadjacent collocations (e.g., *provide some of the information*) facilitated processing in the same way that adjacent collocations do (e.g., *provide information*). She found that her L1 participants were sensitive to both; adjacent and nonadjacent collocations showed similar processing advantages for entire-phrase reading times. However, the final-word reading measures showed a processing advantage only for adjacent collocations.

Overall, studies on collocational processing have confirmed that there is a processing advantage for collocations due to their high-frequency. However, only a few studies have looked at the effects of probabilistic relationships of collocations, known as strength of association (see Evert, 2008; Gablasova et al., 2017) and also defined as word-to-word contingency statistics (Yi, 2018) or transition probabilities (see McCauley & Christiansen, 2017; McDonald & Shillcock, 2003). In one example, McDonald and Shillcock (2003) analyzed L1 English speakers' eye-movements to identify how strength of verb–noun collocations measured by transitional probabilities affected their processing. They found that initial-fixation duration was significantly shorter for verb–noun collocations with high transitional probability (e.g., *avoid confusion*) than for pairs with low transitional probability (e.g., *avoid discovery*). However, Frisson, Rayner, and Pickering (2005) found that transitional probabilities had no significant effect on collocational processing if contextual predictability was controlled. Nevertheless, they argued that contextual predictability (measured by cloze tests) involves some aspects of transitional probabilities, so one cannot entirely dismiss their effects on language processing. Ellis,

Simpson-Vlach, and Maynard (2008) investigated the psychological reality of multiword sequences in academic contexts (e.g., *a wide variety of*) using a series of comprehension and production tasks. They found that L1 speakers' processing of multiword sequences was affected by mutual information scores—mutual information is a corpus-based association measure highlighting the rare exclusivity of word combinations (see Gablasova et al., 2017). However, advanced L2 speakers' processing of multiword sequences appears to be affected by their phrasal frequency. These findings were interesting, but, because of the small sample size and lack of control over confounding variables (e.g., single word frequency and collocation frequency), the findings were limited.

Some recent experimental and computational modeling studies have also looked at the effects of collocational strength on processing. Yi (2018) examined L1 and advanced L2 learners' sensitivity to frequency and association of adjective–noun collocations, revealing that both groups were sensitive to both measures using mutual information scores. Furthermore, advanced L2 speakers' sensitivity to collocational frequency and association statistics was considerably stronger than that of L1 speakers. McCauley and Christiansen (2017) compared L1 and L2 learners' use of multiword sequences, employing a large-scale corpus-based computational model. They found that L2 learners are significantly more sensitive to the phrasal frequency of multiword sequences than to their associations, measured by mutual information scores. Due to these contrasting findings, it has remained unclear whether L2 speakers are sensitive to collocational strength, or whether the corpus-based association measures used (e.g., mutual information) directly affect the findings for speakers' sensitivity to collocational strength.

Operationalizing Collocational Strength: Reviewing Corpus-Based Association Measures

The corpus-based association measures used in psycholinguistic studies are likely to directly and significantly affect the findings and, consequently, the insights into language learning and processing (Gablasova et al., 2017) derived through them. Although various studies with a corpus linguistic focus have made efforts to standardize the conflicting terminology (e.g., Ebeling & Hasselgård, 2015; Evert, 2008; Gablasova et al., 2017), the rationales behind the selection of the association measures in psycholinguistic studies have not always been fully transparent and systematic. Despite the availability of many association measures (see comprehensive overviews by Evert, 2005; Pecina, 2009; Wiechmann, 2008), so far, the mutual information measure has been predominantly used in psycholinguistic research either to extract collocations

(e.g., Vilkaité, 2016) or to investigate language users' sensitivity to collocational strength (e.g., Yi, 2018). Therefore, we first review those studies that chose the mutual information measure, their justifications for using it, and the mathematical reasoning behind the mutual information measure. We then review the log Dice measure that we used in this study as an alternative to the mutual information measure. Finally, other possible measures are briefly discussed.

The mutual information score is a field-standard measure for calculating collocational strength in psycholinguistic research (e.g., Ellis et al., 2008; McCauley & Christiansen, 2017; Vilkaité, 2016; Wolter & Yamashita, 2015; Yi, 2018). It has variously been described as a measure of appropriateness (Siyanova & Schmitt, 2008), coherence (Ellis et al., 2008), and significant co-occurrence (Wolter & Yamashita, 2015). It operates on a binary logarithmic scale expressing the ratio between the collocation frequency and the frequency of the random co-occurrence of the two words in a collocation (Church & Hanks, 1990). Random co-occurrence would occur if, for example, a corpus were a box in which all words were written on small pieces of paper and the box were shaken thoroughly (Gablasova et al., 2017). The reliability of this random co-occurrence model as a baseline is questionable because it assumes no structural properties of language, which is, by definition, not accurate. It favors low-frequency word pairs whose components are likely to be low-frequency themselves (Garner, Crossley, & Kyle, 2018, 2019; Schmitt, 2012). The measure also tends to assign inflated scores to low-frequency combinations (see Appendix S1 in the Supporting Information online for the mathematical equations of the mutual information and log Dice measures). Thus the value does not only indicate the exclusivity of collocations but also how infrequently they occur in corpora (see also Evert, 2008; Gablasova et al., 2017). Researchers must therefore be careful not to automatically interpret larger mutual information scores as indicators of more coherent word combinations, because the mutual information score is not constructed to highlight coherence or semantic unity of word combinations. Another disadvantage is that it operates on a scale that does not have theoretical (corpus-independent) minimum and maximum values, preventing easy interpretation of mutual information scores for collocations extracted from different corpora.

As an alternative measure, Gablasova et al. (2017) introduced the log Dice measure, which has not yet been used in psycholinguistic and corpus-based language learning research. The log Dice score uses the harmonic mean of two proportions that express the tendency of two words to co-occur relative to the frequency of these words in the corpus (Evert, 2008; Smadja, McKeown, &

Hatzivassiloglou, 1996). Therefore, the log Dice score is an index of exclusive, but not necessarily rare combinations, and does not rely on the shake-the-box, random distribution model of language because it does not include the expected frequency in its equation. As a standardized measure on a scale with a fixed maximum value of 14, a log Dice score is easier to interpret than a mutual information score. It is therefore possible to see how far the value of a particular combination is from the theoretical maximum value (Gablasova et al., 2017). Word pairs with a high log Dice score (over 13) include *vice versa* and *zig zag* in the British National Corpus (BNC XML edition; 2007). In sum, the log Dice measure is preferable to the mutual information measure if researchers aim to look at the exclusivity of collocations without a low-frequency bias (Gablasova et al., 2017).

In practical terms, the mutual information and log Dice measures capture slightly different aspects of the collocational relationships. The mutual information measure highlights rare exclusivity because it is negatively associated with frequency. In other words, it rewards lower frequency combinations for which there is less evidence in the corpus (see also Evert, 2008; Gablasova et al., 2017). For instance, the combination *ceteris paribus* receives a lower mutual information score (raw frequency = 46, mutual information = 21) than *jampa ndogrup* (raw frequency = 10, mutual information = 23.2), according to the BNC XML edition. Although both combinations are exclusively associated, the former combination is considerably more frequent than the latter one. Log Dice is thus an ideal measure because it highlights exclusivity between words in a collocation without favoring low-frequency combinations (Gablasova et al., 2017). Furthermore, log Dice scores are reliable across corpora and sub-corpora because the scores are not affected by corpus size. Even though the current study focused on L1 and L2 speakers' sensitivity to the mutual information and log Dice measures that highlight the exclusivity of collocations, alternative association measures can capture other dimensions of collocational association. For example, delta P, arising out of associative learning theory, highlights directionality of collocational strength (Gries, 2013). It identifies whether the first word better predicts the second one, or vice versa (see Garner et al., 2018, 2019 for applications of the delta P measure in learner corpus research). Dispersion is another dimension of collocational association; it considers the distribution of the node and collocates in the corpus (Gries, 2008). Cohen's *d*, the commonly used measure of effect size (Cohen, 1988), can be used as an association measure to explore the distribution of collocates in different texts or subcorpora (Brezina, McEnery, & Wattam, 2015). Among other association measures are *t*-score, MI2, MI3, *z*-score. Due to space

constraints, we have not discussed these measures here (see Brezina, 2018, pp. 66–75; Evert 2005; Gries, 2008; 2013, for a detailed review of association measures).

The Current Study

In the present study, we first examined the prominence of single-word and collocation frequency information for processing high- and low-frequency collocations. More specifically, we aimed to examine whether L1 and L2 speakers' sensitivity to collocation and single-word frequency counts differ when processing high- and low-frequency collocations. Second, we wanted to examine whether there is a difference between L1 and L2 English speakers' sensitivity to association of collocations in relation to the specific measures used. A few studies have previously investigated L1 and L2 speakers' sensitivity to collocational association, but they produced contrasting findings. The literature has yet to reach a consensus on the effect of collocational association on L1 and L2 speakers' processing (e.g., Ellis et al., 2008; McDonald & Shillcock, 2003; Yi, 2018). Furthermore, the possible effect of specific association measures used on speakers' sensitivity remains underexplored. Therefore, in the present study, we aimed to test whether speakers' sensitivity to collocational association depends on how associations are operationalized. In this way it may be possible to assess the extent to which the specific association measure used in previous studies affected their findings. We explored the following research questions for this study:

1. To what extent is there a difference between L1 and advanced level L2 speakers' sensitivity to both word-level frequency information and collocation frequency information when processing collocations?
2. To what extent is there a difference between L1 and advanced level L2 speakers' sensitivity to word-level frequency information when processing high- and low-frequency collocations?
3. To what extent is there a difference between L1 and advanced level L2 speakers' sensitivity to the strength of collocations as measured using mutual information and log Dice scores?

In the light of our review of theoretical positions and empirical studies, we predicted that both L1 and L2 speakers would be sensitive to both single-word and collocation frequency information simultaneously (see Wolter & Yamashita, 2018). However, we also expected that the frequency of the collocations would cause a difference in the prominence of word-level and collocation-level frequency information for both groups of participants.

Specifically, we expected the effect of individual word frequency information to be weaker for processing high-frequency collocations than for low-frequency collocations because, with increasing frequency, the whole would gain prominence relative to the part (Arnon & Cohen Priva, 2014). Finally, we predicted that the specific collocational association measures used would affect L2 speakers' sensitivity to collocational associations.

Method
Participants
A group of L1 English speakers ($n = 30$) and a group of advanced level L2 learners of English (L1 Turkish, $n = 32$) participated in the study. The L1 English group consisted of 24 undergraduate and six postgraduate students, all from a university in the United Kingdom. The L2 English group consisted of 22 undergraduate and 10 postgraduate students, all from two universities in Turkey. We administered the LexTALE,[1] a test of vocabulary knowledge for advanced learners of English (Lemhöfer & Broersma, 2012), to the L2 English learners to assess their vocabulary knowledge as a proxy for general English proficiency. LexTALE scores were found to be substantially and significantly correlated with the Oxford Quick Placement Test, which is used to group learners into seven levels linked to the proficiency levels of the Common European Framework of Reference for Languages, ranging from beginner to upper advanced. Following the LexTALE norms reported by Lemhöfer and Broersma (2012), we used a LexTALE score of 80.5% (corresponding to an Oxford Placement Test score of 80%) as a cut-off score to recruit advanced level L2 users of English. The L1 English group ($M = 90.82$) had significantly larger vocabulary size than did the L2 group ($M = 84.85$), $t(56.47) = 5.12$, $p < .001$. From the L2 group, 21 participants had lived in an English-speaking country for longer than 1 month (full biographical data for the participants are provided in Table 1).

Materials
To address our research questions, we used an acceptability judgment task (Wolter & Gyllstad, 2013). A key assumption underlying the task is that low-frequency collocations should elicit slower response times for both the L1 and L2 groups in comparison to their response times to high-frequency collocations. With this assumption in mind, we extracted a total of 120 English adjective–noun combinations from the BNC XML edition. We preferred adjective–noun combinations, following the methodological choice of Wolter and Gyllstad (2013), because variability in determiners in verb–noun

Table 1 Participant background variables

Variable	L1		L2	
	M (SD)	95% CI	M (SD)	95% CI
Age (years)	20.58 (2.16)	[19.8, 21.35]	24.43 (4.01)	[23, 25.81]
Age started to learn English	–	–	10.96 (4.04)	[9.56, 12.35]
Self-rated English proficiency[a]:				
Speaking	–	–	5.28 (0.44)	[5.12, 5.43]
Listening	–	–	5.62 (0.48)	[5.45, 5.78]
Reading	–	–	5.65 (0.47)	[5.48, 5.81]
Writing	–	–	5.53 (0.49)	[5.36, 5.69]
LexTALE[b] scores	90.82 (3.71)	[89.49, 92.14]	84.85 (5.22)	[83, 86.65]
	Frequency			
Gender (m/f)[c]	12/17		16/16	
Dexterity (r/l/both)	28/2/0		28/3/1	

[a]Based on a scale from 1 (*beginner*) to 6 (*advanced*).
[b]Lexical Test for Advanced Learners of English.
[c]One L1 English speaker did not indicate gender.

combinations (e.g., *make a mistake* vs. *make progress*) introduces another confounding variable, whereas adjective–noun combinations allow for more control over item consistency by not including determiners. The items fell into one of the three critical conditions: (a) high-frequency collocations ($k = 30$), (b) low-frequency collocations ($k = 30$), (c) noncollocational (baseline) items $k = 60$. We used the noncollocational items for establishing threshold response times and for measuring the relative response times for the items in conditions (a) and (b). We obtained single word frequency counts for the adjectives and nouns, collocation frequency counts, log Dice score, and mutual information scores of the items from the BNC XML edition. For this study, we preferred to use nonlemmatized frequency counts at both the single-word and collocation level. Although arguments have been put forth favoring either the use of lemmatized frequencies or the use of nonlemmatized frequencies, Durrant (2014) found no clear differences between the two forms for predicting L2 learners' knowledge of collocations (p. 465).

To be able to extract the items for the three critical conditions, we explored the scales of raw frequencies for adjective–noun collocations and log Dice scores in the BNC XML edition. In order to determine the frequency and log Dice cut-off scores for high- and low-frequency collocations, we selected 10 noun node words from various raw frequency counts with a high-frequency count of 121,591 (e.g., *people*) and a low-frequency count of 8,961 (e.g., *officer*). Using the selected noun nodes, we extracted a total of 4,718 two-word adjective–noun combinations from the CQPWEB tool (Hardie, 2012). To determine the cut-off frequency counts and log Dice scores for high-frequency and low-frequency collocations, we looked closely at the distributions of the raw frequency counts of collocations and the range of log Dice scores for the adjective–noun pairs with various raw frequency scores ($\leq$100, 101$-$200, 201$-$300, 301$-$400, >400) in the BNC XML edition. We were not surprised that the frequency counts of the adjective–noun combinations followed Zipf-like skewed distributions, with a small number of high-frequency collocations and a very large number of low-frequency adjective–noun combinations (see Appendix S2 in the Supporting Information online for a table of the noun nodes and visual illustrations of collocations' frequency information). To measure collocations' strength of associations in each frequency band, we used the log Dice measure because it is not negatively linked to frequency (Gablasova et al., 2017). We defined high-frequency collocations as adjective–noun collocations with raw frequency counts greater than or equal to 300 and log Dice scores greater than or equal to 7. We defined as low-frequency collocations adjective–noun collocations with raw frequency counts between 10 and 150

and log Dice scores between 2 and 4 within the span of three words to the left and three words to the right of the collocations.

To select high-frequency collocations, we checked the nouns in the BNC XML edition word frequency list for whether they collocated with an adjective in a way that met the cut-off for raw frequency and had log Dice scores high-frequency collocations. An initial list of 36 collocations satisfied the selection criteria for high-frequency collocations. We discarded four of the collocations in the list because they were incongruent with Turkish (e.g., *supreme court*, *British library*) given the empirical evidence that lexical congruency affects collocational processing in L2 (Wolter & Gyllstad, 2011, 2013). Because the main goal of this study was to investigate whether there is a difference in L1 and L2 speakers' sensitivity to single-word and collocation frequency information, including incongruent collocations would have introduced a confounding variable. To identify congruent items, we used the following procedure. Initially, the first author (a native speaker of Turkish with a high command of English) translated English items into Turkish. Then we checked the translations against the Turkish National Corpus (Aksan et al., 2012), a large, balanced, and representative corpus for modern Turkish consisting of 50 million words. We identified the translated items that occurred frequently in the Turkish National Corpus as congruent collocations, and we considered the items that were not found in the Turkish National Corpus as incongruent. Cognates were a concern because they may elicit faster response times (Lemhöfer et al., 2008). We therefore discarded collocations whose component words were Turkish. Nonetheless, we could not fully eliminate all potential cognates. The number of remaining potential cognates corresponded to 8.3% of all items. A list of 30 high-frequency English collocations remained. The mean log Dice score for all high-frequency collocations was 7.80, with a low score of 7.0 (for the items *dark hair* and *left hand*) and with a high score of 10.95 (for the item *prime minister*).

To select low-frequency collocations, we checked whether the unused nouns in the BNC XML edition high-frequency collocations word list collocated with an adjective in a way that met the cut-off for raw frequency and log Dice scores that allowed them to be categorized as low-frequency collocations. The selected low-frequency collocations had raw frequency counts between 10 and 150 and log Dice scores between 2 and 4 within the span of three words to the left and three words to the right of the collocations. As with the high-frequency collocations, the low-frequency collocations were also congruent with Turkish. We selected none of the nouns and adjectives used for the items in the high-frequency collocations for the items in the low-frequency

collocations. However, we closely matched single words (both adjectives and nouns) in both types of items for frequency and for item length, operationalized as number of letters. We extracted a list of 30 low-frequency collocations. The mean log Dice score for all the low-frequency collocations was 3.24, with a low score of 2.54 (for the item *away game*) and with a high score of 3.91 (for the item *vital information*). We checked concordance lines to ensure that for each of the high-frequency and low-frequency collocations, adjectives modified the nouns. We log transformed all single word and collocational frequency counts using SUBTLEX Zipf scale (Van Heuven, Mandera, Keuleers, & Brysbaert, 2014).

The baseline items consisted of random combinations of the nouns used for the high-frequency or low-frequency collocations with adjectives that had not been used for the high-frequency or low-frequency collocations. On the one hand, repeating the same nouns in different conditions was an ideal way of ensuring that we had perfectly matched the single word length and frequency counts of the nouns in the collocational and baseline conditions. On the other hand, this meant that each noun appeared in the task twice, and this inevitably introduced another potential confounder in that participants saw nouns twice under different conditions, potentially lowering the activation thresholds. To address this, we presented all items to the participants in an individually randomized order. Thus, any advantage that the participants gained from seeing word for a second time was evened out both within an individual participant's test and across all of the participants; that is to say, we used each noun once in a collocation and once in a baseline item. We closely matched adjectives in the collocational and baseline conditions for frequency and length. We checked all combined nouns and adjectives that we had used to construct the baseline items against the BNC XML edition to make sure that there were no co-occurrences. When we found any co-occurrences in the BNC XML edition, we checked the log Dice scores to make sure that they were negative values. If the combinations produced positive log Dice scores, we repeated the process. We eventually obtained a list of 60 baseline items. However, given the very large size of the BNC XML edition, it was not possible to fully eliminate all positive log Dice scores. We therefore decided to retain two items with positive but very low log Dice scores. The mean log Dice score for all baseline items was −0.93, with a low score of −3.22 (for the item *dirty time*) and with a high score of 0.45 (for the item *clear trade*). The baseline items had raw frequency counts less than or equal to 10. We checked the concordance lines to make sure that they were idiosyncratic rather than meaningful co-occurrences (Item characteristics are summarized in Table 2). See Appendix S3 in the Supporting Information

Table 2 Test item characteristics with Wilcoxon Tests

Item characteristic	Collocation frequency							
	High		Low		Non			
	M (SD)	95% CI	M (SD)	95% CI	M (SD)	95% CI	W	p
Item length	10.86 (2.97)	[9.79, 11.92]	11.10 (2.30)	[10.27, 11.92]	11.10 (2.52)	[10.19, 12]	401.0	.46
Adjective frequency	5.17 (0.31)	[5.05, 5,28]	5.17 (0.42)	[5.01, 5.32]	5.15 (0.24)	[5.06, 5.23]	467.5	.79
Noun frequency	5.36 (0.29)	[5.25, 5.46]	5.36 (0.21)	[5.28, 5.43]	5.36 (0.25)	[5.27, 5.44]	415.5	.60
Collocational frequency	4.03 (0.34)	[3.9, 4.15]	2.70 (0.30)	[2.59, 2.8]	1.18 (0.52)	[0.99, 1.36]	891.0	<.001
Log Dice scores	7.8 (0.82)	[7.5, 8.09]	3.24 (0.39)	[3.1, 3.37]	−0.93 (0.85)	[−1.23, −0.62]	900.0	<.001

online for the full list of items. Following open science practices, the items are also available from htpps://osf.io/dxvak/.

Procedure

We collected the response times for the items in the three critical conditions by means of acceptability judgments. We administered this task using the PsychoPy software (Peirce et al., 2019). The judgment task required participants to indicate whether or not they thought the items are acceptable. Such tasks have most frequently been used to elicit grammatical acceptability ratings, for which judgments are more straightforward. However, the vast majority of our adjective–noun combinations were mostly grammatical, though the word combinations could indicate something that is highly unlikely and so perhaps could be deemed semantically unacceptable (e.g., *old child*). Therefore, adjective–noun combinations can almost always be perceived as "acceptable" if the respondent is flexible in interpreting them. To avoid this obstacle, we followed the phrasing for our instructions to participants that was used by Wolter and Gyllstad (2013) and asked participants to indicate whether or not the word combinations were commonly used in English. The exact instructions were:

> In this experiment, you will be presented with 120 word combinations. Your task is to decide, as quickly and accurately as possible, whether the word combinations are commonly used in English or not. For instance, the word combination *harsh words* is a commonly used word combination in English, but *complex force* is not a commonly used word combination in English. Please press the "YES" button on the game pad if the word combination is commonly used and the "NO" button if it is not commonly used in English.

Figure 1 illustrates the presentation sequence. First, the eye fixation (#########) was presented for 250 milliseconds and was followed by a blank screen. After the blank screen, the task item was presented in lowercase characters in the Times News Roman 12-point font. The item remained on the screen either until the participants had indicated their responses (via pressing a button) or after a 4,000-milliseconds timeout. The participants answered *yes* by pressing the button corresponding to the forefinger of the dominant hand and *no* by pressing the button corresponding to the forefinger of the nondominant hand (in line with Ferrand et al., 2010; Robert & Rico Duarte, 2016; Sato & Athanasopoulos, 2018; Shatzman & Schiller, 2004). The acceptability judgment task began with a practice session to familiarize the participants with the task. We

senior officer

3. Item displayed (4000ms)

##########

2. Blank screen (50 ms)

1. Eye fixation (250 ms)
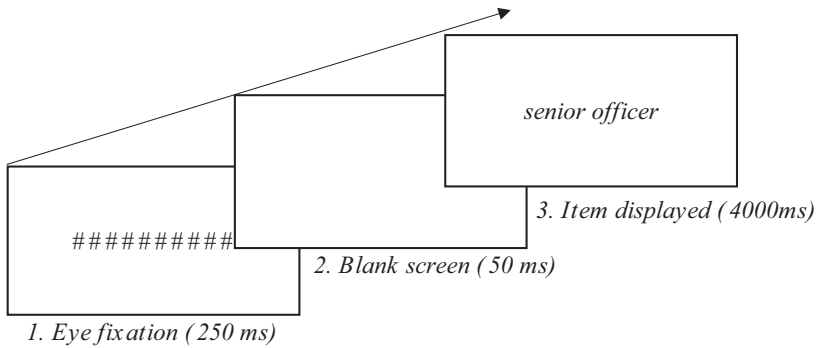
Figure 1 Presentation sequence for items in the acceptability judgment task.

allowed the participants a short break after the practice session. Most participants completed the judgement task in 5 to 6 minutes. Afterward, we administered to both groups of participants the LexTALE test. In addition, we also asked the L2 group to complete a questionnaire to self-rate their perceptions of their English proficiency in the four skills: speaking, listening, reading, and writing (see Table 1 for the L2 groups' mean self-rating proficiency scores).

Results
Preliminary Analyses
Following open science practices, we have made available all participant data including the LexTALE scores and response times at htpps://osf.io/dxvak/. The main concern of the present study was how the L1 and L2 participants processed the high- and low-frequency collocations that they perceived as commonly used compared to the baseline items that they perceived as not commonly used. Therefore, we analyzed the response times to the high- and low-frequency collocations that received a *yes* response and compared them to the baseline items that received a *no* response.

This approach could potentially have been problematic in two ways. First, if the majority of the high- and low-frequency collocations received a *no* response or a majority of the baseline items received a *yes* response. Fortunately neither was the case for our two groups of participants. The L1 group judged 98% of the high-frequency collocations and 78.11% of the low-frequency collocations to be commonly used in English, and they decided that 78.77% of the baseline items are not commonly used in English. The L2 group judged 97.5% of the high-frequency collocations and 76.56% of the low-frequency

collocations to be commonly used in English. They decided that 71.19% of the baseline items are not commonly used in English.

The second reason this approach could have been problematic is that the corpus data that we used did not fully represent the individual experiences of the participants (see also e.g., Durrant, 2013; González Fernández & Schmitt, 2015); that is to say, the participants' different individual language experiences might have led them to judge certain items based on their own language experiences with English that were different from the corpus-based evidence. However, considering the findings that both the L1 and L2 speaker groups judged the vast majority of high- and low-frequency collocations to be commonly used and baseline items to be not commonly used, this problem did not materialize in the present study. To begin the statistical analyses, we calculated mean response times in milliseconds for each item type, that is, the high- and low-frequency collocations that received *yes* responses and baseline items that received *no* responses. Table 3 shows the mean response times in the three conditions for both groups (see Appendix S4 in the Supporting Information online for a visual illustration of the same data).

Model Development

We used the lme4 package (Bates, Maechler, Bolker, & Walker, 2015) for the R statistical platform (R Core Team, 2012) to construct mixed-effects models comparing response times[2] and calculated the *p* values using the lmerTest package (see Kuznetsova, Brockhoff, & Christensen, 2017). Before constructing the models, we prepared the data for analysis. The first step in this process was to prepare the response time data. Following the minimal data trimming choice suggested by Gyllstad and Wolter (2016) and Wolter and Yamashita (2018), we excluded only responses that were faster than 450 milliseconds and responses that timed out at 4,000 milliseconds.[3] We carefully examined the histograms of log transformed residuals and the residuals of the raw response time models. Because the distribution of the model residuals was not normal, we log transformed the remaining response times (see also Baayen & Milin, 2010). The second step was to prepare the continuous predictors. We centered and standardized all the continuous predictors and treated the first versus second occurrence variables as categorical (i.e., whether a participant was seeing a particular noun for the first or second time). In our third step, we recoded the categorical variable group (L1 vs. L2) using contrast coding. This provided some interpretational advantages for analyzing the interactions. The recoding included converting the group variables into numeric variables (L1 = 0.5, L2 = −0.5). We coded the other categorical variable item type using the

Table 3 Group response times in milliseconds

Item type	L1 (n = 30)		L2 (n = 32)	
	M (SD)	95% CI	M (SD)	95% CI
Collocation frequency				
High	892 (338)	[771.05, 1,012.95]	943 (383)	[810.30, 1,075.70]
Low	1,075 (431)	[920.77, 1,229.23]	1,146 (477)	[980.73, 1,311.27]
Baseline items	1,303 (527)	[1,111.14, 1,491.58]	1,326 (559)	[1,132.32, 1,519.68]

treatment coding in which we defined baseline items as the reference level and compared high-frequency and low-frequency collocations to the baseline items (baseline = 0, high-frequency = 1, low-frequency = 1). Finally, we calculated the variance information factor (VIF) scores[4] using the car package in R (Fox & Weisberg, 2019) to check whether there were any multicollinearity problems among the predictor variables. Finally, we computed effect sizes of the models using the MuMIn package in R[5] (Barton, 2019).

We constructed the first model to investigate whether there were significant group differences of overall mean response times for any of the item types. It included participant and item as crossed random effects. We also had a by-subject (participant) random intercept for subject, a by-subject random slope for item type, a by-item random intercept for item, and also a by-item random slope for group. We included the following variables as fixed effects in the first model: group (L1 or L2), item type (high-frequency, low-frequency, or baseline), LexTALE scores, and the interaction between group and item type. Furthermore, we added item length, participants' age, gender, and the first versus second occurrence of the nouns. The VIF scores of item type, group, gender, age, LexTALE scores, and item length did not indicate any problems with multicollinearity (VIF scores < 2.00).

Table 4 presents the results for the first mixed-effects model comparing the L1 and L2 participants' response times for high-frequency, low-frequency, and baseline items. Considering the R^2 values (R^2 marginal = .20, R^2 = conditional = .42), this model seems to explain an acceptable amount of variance. These values are very close to the R^2 values reported by previous studies investigating collocational processing (e.g., Gyllstad & Wolter, 2016; Wolter & Yamashita, 2018). The results revealed no significant differences between L1 and L2 groups either in terms of overall mean response times, or with respect to group by item type interactions. We ran a series of pairwise comparisons tests to decompose the interactions between group and item type using the emmeans package in R with Tukey adjustments for multiple comparisons (Lenth, 2018). The results showed no significant differences between L1 and L2 groups' overall mean response times for high-frequency, Estimate = 0.01, SE = 0.04, z = 0.22, p = .99, and low-frequency collocations, Estimate = 0.008, SE = 0.05, z = 0.15, p = 1.00 (the term "estimate" indicates the estimate of mean difference). We also compared this model including the interactions (group by item type) and without the interactions using a log-likelihood ratio test to determine whether the inclusion of these interactions produced a better-fitting model. There was not a significant difference between the two models according to the log-likelihood ratio test, $\chi^2(2) = 0.46$, $p = .79$, and this finding

Table 4 Mixed-effects Model 1 comparing L1 and L2 participants' response times for high-frequency, low-frequency, and baseline items

Fixed effects	b	SE	95% CI	t	df	p
(Intercept)	050	0.38	[−0.23, 1.23]	1.32	62.2	.18
L1 vs. L2	003	0.06	[−0.08, 0.14]	0.53	78.1	.59
Male vs. Female[a]	010	0.04	[0.02, 0.16]	2.65	61.7	.01
Gender not stated vs. Female	003	0.14	[−0.25, 0.30]	0.18	61.1	.85
LexTALE[b] scores	000	0.00	[−0.11, 0.00]	−1.07	61.7	.28
Age	000	0.00	[−0.00, 0.01]	0.51	61.8	.60
High-frequency vs. Baseline	−036	0.02	[−0.41, −0.31]	−14.87	118.4	<.001
Low-frequency vs. Baseline	−018	0.01	[−0.22, −0.13]	−7.74	118.4	<.001
Item length	003	0.01	[0.02, 0.04]	4.67	115.7	<.001
Order of occurrence	001	0.01	[−0.02, 0.00]	−0.69	5880	.48
L1 vs. L2 × High-frequency (vs. Baseline)	−002	0.04	[−0.09, 0.05]	−0.56	65.5	.57
L1 vs. L2 × Low-frequency (vs. Baseline)	−002	0.04	[−0.09, 0.04]	−0.67	64.2	.50

Note. R^2 marginal = .20; R^2 conditional = .42.
[a]One L1 English speaker did not indicate gender so we included three levels (male, female, not stated).
[b]Lexical Test for Advanced Learners of English.

provided further support for the conclusion that L1 and L2 participants performed very similarly with respect to their response times for all item types.

As we had expected, the participants responded to both the high-frequency, and the low-frequency collocations, faster than to the baseline items. Furthermore, releveling the model to directly compare high-frequency collocations with low-frequency collocations revealed that the participants responded to high-frequency collocations faster than to the low-frequency collocations, $b = -0.18$, $SE = 0.02$, 95% CI $[-0.22, -0.13]$, $t(103.8) = -8.27$, $p < .001$. Male participants had significantly slower response times on average than female participants, and the participants' ages did not seem to affect their response times. The effect of the LexTALE scores was not significant. It was not surprising that items with more letters elicited slower response times. The effect of the order of occurrence of nouns was not significant.

We constructed the second model to investigate the possible differences between L1 and L2 speakers' sensitivity to word-level frequency counts for adjectives and nouns and collocation frequency counts. For this model, we first eliminated the baseline items because nearly all of the baseline items had collocation frequency counts of 0. Furthermore, baseline items required a *no* response and high- and low-frequency collocations required a *yes* response in the acceptability judgment task. Because different mechanisms might affect the processing of collocations and baseline items, it seemed useful to analyze them separately. Due to the multicollinearity problem between collocation frequency and item type, we needed to discard item type from this model (VIF = 10.55, 9.78, respectively). As with the first model, this model also included participant and item as crossed random effects. Additionally, we included a by-subject (participant) random intercept for subject, a by-item random intercept for item, and a by-item random slope for group. For fixed effects, we added the following variables to the second model: group (L1 or L2), single word frequency counts for adjectives and nouns, collocation frequency, and item length. Furthermore, we added group by adjective frequency, group by noun frequency, and group by collocation frequency counts as interactions to the second model.

Table 5 shows that we found no significant differences of overall mean response times between the L1 and L2 groups in our second model. We used the emtrends function to decompose the interactions between the L1 and L2 groups and different frequency measures. The results also yielded nonsignificant interaction effects between group and adjective frequency, Estimate = 0.01, $SE = 0.01$, $z = 0.91$, $p = .35$, group and noun frequency, Estimate = -0.0005, $SE = 0.01$, $z = -0.042$, $p = .96$, group and collocation frequency counts,

Table 5 Mixed-effects Model 2 investigating L1 and L2 participants' sensitivity to adjective, noun, and collocation frequency counts

Fixed effects	b	SE	95% CI	t	df	p
(Intercept)	−0.06	0.02	[−0.09, −0.01]	−2.90	78	.004
Group (L1 vs. L2)	−0.05	0.04	[−0.12, 0.03]	−1.23	63.1	.22
Adjective frequency	0.02	0.01	[0.00, 0.04]	2.42	54.3	.01
Noun frequency	0.00	0.01	[−0.02, 0.02]	0.00	50.1	.99
Collocation frequency	−0.10	0.01	[−0.12, −0.08]	−10.31	51.4	<.001
Item length	0.04	0.01	[0.02, 0.06]	4.23	49.3	<.001
L1 vs. L2 × Adjective Frequency	0.01	0.01	[−0.01, 0.03]	0.91	69.	.36
L1 vs. L2 × Noun Frequency	0.00	0.01	[−0.02, 0.02]	−0.04	57.6	.96
L1 vs. L2 × Collocation Frequency	0.00	0.01	[−0.02, 0.02]	−0.08	62	.93

Note. R^2 marginal = .10; R^2 conditional = .30.

Estimate $= -0.0009$, $SE = 0.01$, $z = -0.08$, $p = .93$. We compared this model with the main effects only version of the second model that excluded the interactions using a log-likelihood ratio test to find out whether the inclusion of the interactions produced a better-fitting model. There was no significant difference between the two models according to the log-likelihood ratio test, $\chi^2(3) = 0.86$, $p = .83$, and this finding provided further support for the conclusion that L1 and advanced level L2 participants were very similarly sensitive to the word-level and collocation frequency counts. As main effects, collocation frequency counts led to faster response times, but adjective frequency counts led to slower response times. The effect of noun frequency counts was not significant.

As we noted previously, due to the multicollinearity problem (between collocation frequency counts and item type) it was not possible to investigate the interaction between item type and single word frequency counts for adjectives and nouns in the second model. We therefore constructed a third model to explore possible difference in participants' sensitivity to word-level frequency information when processing high- and low-frequency collocations. We eliminated collocation frequency from this model and added item type. For this model, we coded the categorical variable item type using contrast coding (high-frequency $= .5$, low-frequency $= -.5$). We did not include group either as a main effect or as an interaction between group and item type because their effects were not significant in the previous models. This model included participant and item as crossed random effects. We also included a by-subject (participant) random intercept for subject, a by-subject random slope for item type, and a by-item random intercept for item. For fixed effects, we added the following variables to this third model: item type (high-frequency or low-frequency), word-level frequency counts for adjectives and nouns, item length, and interactions between word-level frequency counts for adjectives and nouns and item type.

As Table 6 shows, the participants responded to high-frequency collocations faster than to low-frequency collocations. Noun frequency counts led to significantly faster response times. To interpret the interactions between noun frequency counts and item type, we first obtained the simple slopes for noun frequency counts by each level of item type (high-frequency vs. low-frequency), using the emtrends function within the emmeans package in R (Lenth, 2018). There was a significant interaction between noun frequency counts and item type, Estimate $= 0.04$, $SE = 0.02$, $z = 2.0$, $p = .04$, indicating that the participants' sensitivity to noun frequency counts varied depending on the frequency of the collocations. Figure 2 illustrates the interaction effect.

Table 6 Mixed-effects Model 3 investigating L1 and L2 participants' sensitivity to single word frequency counts for processing high- and low-frequency collocations

Fixed effects	b	SE	95% CI	T	df	p
(Intercept)	−0.05	0.02	[−0.08, 0.00]	−2.25	80.2	.02
High-frequency vs. Low-frequency	−0.18	0.02	[−0.22, −0.14]	−9.02	57.3	<.001
Adjective frequency	0.01	0.01	[−0.01, 0.029]	0.85	57.8	.39
Noun frequency	−0.04	0.01	[−0.06, −0.02]	−3.49	58.5	<.001
Item length	0.04	0.01	[0.02, 0.06]	3.58	52.9	<.001
Item Type (High freq. vs. Low freq.) × Noun Frequency	0.05	0.02	[0.00, 0.09]	2.00	58.8	.04

Note. R^2 marginal = .10; R^2 conditional = .30.
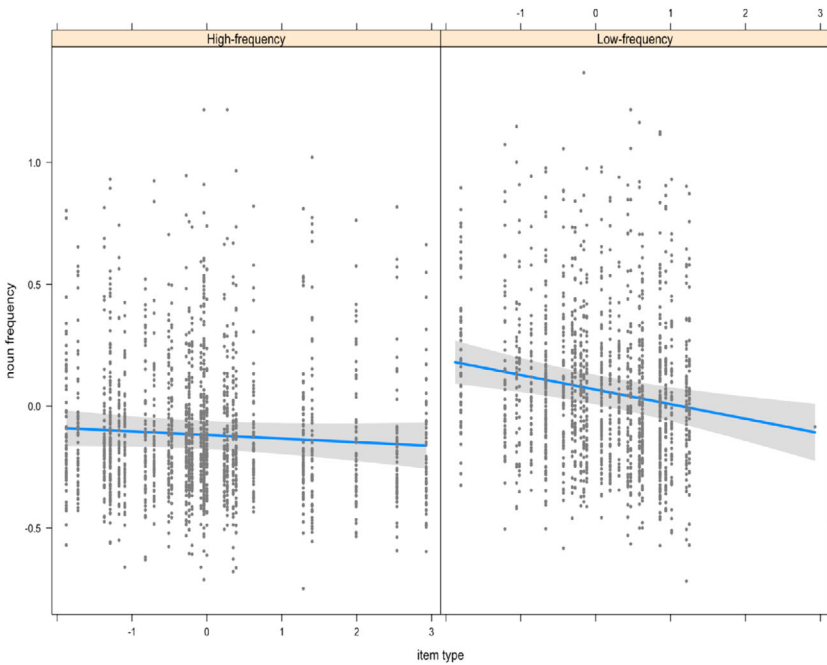
Figure 2 Interaction between item type and noun frequency counts [Color figure can be viewed at wileyonlinelibrary.com]

The effect of noun frequency counts on the participants' response times were in the same direction for both high-frequency and low-frequency collocations, that is to say, as the noun frequency counts increased, participants' response times became faster. However, the effect of noun frequency counts on the participants' response times for low-frequency collocations was stronger than for the high-frequency collocations. More specifically, one unit of increase in noun frequency counts resulted in a -0.059 log response time measure faster for low-frequency collocations, whereas one unit of increase in noun frequency counts resulted in a -0.014 log response time measure faster for high-frequency collocations. The effect of adjective frequency counts was not significant.

Finally, we constructed one more set of models that considered the association statistics of collocations as measured by mutual information and log Dice scores. We aimed to observe whether the way in which collocational association was operationalized would have an effect on L1 and L2 participants' sensitivity to them. The high VIF scores of collocation frequency, log

Dice score, and mutual information score (VIF $= 67.14, 186.75, 198.2$, respectively) indicated a multicollinearity issue. In this case, we could not compare the coefficients of the log Dice and mutual information measures in the same mixed-effects model. Therefore, we decided to observe which association measure produced a better-fitting model of response time by comparing the Akaike Information Criterion (AIC) values of the models.

We constructed two models; one included log Dice scores, and the other included mutual information scores. To observe whether there was a difference between the L1 and L2 participants' sensitivity to collocational strength as measured by mutual information and log Dice scores, we included the interaction between group and the measure of association in the models. Then we compared the two models. The models included participant and item as crossed random effects. We included a by-subject random intercept for subject. We also had a by-item random intercept for item and a by-item random slope for group. According to the AIC values, the model including log Dice scores was a better-fitting model (AIC $= 1,500.2$) than was the model including mutual information scores (AIC $= 1,511.7$). Although the model based on log Dice scores was a better-fitting model than the model based on mutual information scores, the two models were not qualitatively different in their predictions. See Appendix S4 in the Supporting Information online for the table of the model including mutual information scores.

As Table 7 shows, the results for the model including the log Dice scores revealed no significant differences between the L1 and L2 groups for mean response times. As we had expected, log Dice scores were associated with faster response times. The interaction between group and log Dice scores was not significant. Similarly, the results for the model including the mutual information scores indicated no significant differences between the L1 and L2 groups for mean response times, $b = -0.04$, $SE = 0.03$, 95% CI $[-0.12, 0.02]$, $t(63.4) = -1.2$, $p = .23$. Mutual information scores were associated with faster response times, $b = -0.081$, $SE = 0.01$, 95% CI $[-0.1, -0.05]$, $t(61.5) = -6.29$, $p < .001$. The interaction between group and mutual information scores was not significant, $b = -0.004$, $SE = 0.01$, 95% CI $[-0.03, 0.02]$, $t(57.8) = -0.3$, $p = .73$.

Discussion

The results reveal that the adult L1 and advanced L2 participants were sensitive to both word-level and collocation (phrasal) frequency information simultaneously while processing two-word adjective–noun collocations. Both noun frequency and collocation frequency counts led to faster response times for both

Table 7 Mixed-effects Model 3 investigating L1 and L2 participants' sensitivity to collocational strength as measured by log Dice score

Fixed effects	B	SE	95% CI	t	df	p^a
(Intercept)	−0.06	0.02	[−0.10, −0.01]	−2.65	90.4	.01
L1 vs. L2	−0.05	0.04	[−0.12, 0.03]	−1.21	63.4	.22
Log Dice scores	−0.09	0.01	[−0.12, −0.07]	−7.76	59	<.001
Group (L1 vs. L2) × Log Dice	0.00	0.01	[−0.02, 0.03]	0.08	49.7	.93

Note. R^2 marginal = .076. R^2 conditional = .31.

groups. The results further reveal that for both groups of participants, sensitivity to word-level frequency information in relation to nouns differed depending on the frequency of the collocations. As the frequency of the adjective–noun collocations increased, the effect of noun frequency information became weaker for both the L1 and L2 participants. We had expected this finding because increased use of collocations as two-word combinations is likely to make a difference to the prominence of (participants' reliance on) individual word and collocation frequency information. Therefore, we saw reduced effects of noun frequency and increased effects of collocation frequency information for high-frequency collocations. Importantly, the effects of both noun and collocation frequency counts were associated with faster response times. In the case of the effects of adjective frequency counts, however, our findings suggest that they were associated with slower response times for both groups. Finally, the results indicate that there was no difference in sensitivity to collocational strength between the L1 and L2 groups irrespective of how collocational strength was operationalized. We had not expected this finding because the mutual information and log Dice measures underlie different aspects of collocational strength.

In line with our hypothesis, the results of the first and second mixed-effects models show that the L1 and advanced L2 groups' processing was affected by collocation frequency information while the groups processed adjective–noun collocations. Both the L1 and L2 groups responded to the high-frequency collocations faster than to the low-frequency collocations, and they also responded to the low-frequency collocations faster than to the baseline items. This indicates that both the L1 and L2 groups needed a shorter time to process collocations that occur more frequently. In addition, the results of the second mixed-effects model indicate no difference between the L1 and L2 participants' sensitivity to collocation frequency information. Therefore, the results of the present study add to the growing body of empirical evidence that both L1 and L2 speakers' processing is affected by phrasal frequency of multiword sequences (e.g., Siyanova-Chanturia, Conklin, & van Heuven, 2011; Wolter & Yamashita, 2018; Yi, 2018) because both the L1 and L2 groups' response times became faster as collocation frequency increased. This is not to say, however, that the participants' processing was only affected by collocation frequency information or that their response times were not affected by word-level frequency information. The results of the second mixed-effects model show that noun frequency information led to significantly faster response times. Furthermore, the second mixed-effects model indicates no difference between the L1 and L2 participants' sensitivity to noun frequency information. Similar

results have been reported by Wolter and Yamashita (2018), who also used an acceptability judgment task to compare how a L1 group and two groups of L2 speakers with differing proficiency levels processed adjective–noun collocations. They found that all three groups' processing was affected by single word and collocation frequency information simultaneously. In contrast, however they reported that the L2 groups appeared to rely more heavily on word-level frequency information than did the L1 group.

Unlike participants in Wolter and Yamashita's (2018) study, our L1 and L2 participants were comparably sensitive to word-level frequency information. The differences in findings are reconcilable, however. One possibility is that we recruited a L2 group with higher proficiency than did Wolter and Yamashita (2018). Both the present study and that of Wolter and Yamashita (2018) found that L1 and L2 speakers' processing is affected by word-level and collocational frequency information simultaneously. These findings conflict with Wray's (2002, 2008) position that natives and nonnatives process multiword sequences in fundamentally different ways; that is to say, L1 speakers rely on their knowledge of meaning assigned to multiword sequences whereas L2 speakers decompose multiword sequences into individual words and rely heavily on the word-level information making up the multiword sequences. On the contrary, the results of psycholinguistic research have indicated that multiword sequences can be processed in similar ways by L1 and proficient L2 speakers. For example, L1-based psycholinguistic and neurolinguistics studies have consistently reported that even if there is a processing advantage for frequent multiword sequences as a whole, word-level frequency information still affects their processing, regardless of whether the phrases are idioms (e.g., Konopka & Bock, 2009; Snider & Arnon, 2012), complex prepositions (Molinaro, Canal, Vespignani, Pesciarelli, & Cacciari, 2013), or lexical bundles (Tremblay et al., 2011). In addition to the findings of the L1-based research, Wolter and Yamashita (2018) reported that both lower and higher proficiency L2 speakers are uniformly sensitive to both word-level and collocation frequency information. The overall trend in the L2 speakers' response times has suggested a progression from less reliance on word-level frequency to more reliance on collocation-level frequency with gains in proficiency. In the present study, with a very high proficiency L2 group, we observed no significant differences between the L1 and L2 speakers' reliance on word-level and collocation frequency information.

The findings that speakers are sensitive to both single word-level and phrasal frequency information also raises questions about how these different frequency measures interact when speakers process collocations on-line and

whether there are differences between L1 and L2 speakers' reliance on word-level and collocational frequency information when they process high- and low-frequency collocations. Therefore, our second research question focused on whether L1 and L2 speakers' reliance on word-level and collocation level frequency information differs depending on the frequency of the collocations. In line with our hypothesis, the results of mixed-effects Model 3 indicate that the effect of noun frequency information on the participants' response times was stronger for the low-frequency collocations than for the high-frequency collocations. In other words, for the high-frequency adjective–noun collocations, the effect of word–level frequency counts of the nouns on the response times decreased, and the effect of collocation frequency increased. However, the results also show that word-level frequency information still played a role in the processing of even high-frequency collocations. However, at this point, the possible reasons for higher adjective frequency counts leading to slower response times need to be considered. Because we failed to reliably establish an interaction effect between item type and adjective frequency counts, we must apply caution when interpreting the findings. Nevertheless, it would be reasonable to suggest that, when participants see collocations that include a very frequent adjective (e.g., *long time*), they will find predicting the upcoming noun to be more difficult. This was also an expected finding from a corpus linguistics perspective because very high-frequency adjectives tend to form collocations with a wide range of nouns, but those collocations are unlikely to be highly exclusive. The exclusivity of collocates refers to the extent to which the two words appear predominantly in each other's company (Gablasova et al., 2017). Exclusivity is strongly linked to predictability of co-occurrence when seeing one part of a collocation brings to mind the other part. One could argue that very high-frequency adjectives such as *long* or *good* are unlikely to facilitate prediction because participants cannot interpret them before they access the meanings of the nouns.

Arnon and Cohen Priva (2014), who focused on L1 English speakers' phonetic duration in spontaneous speech, reported similar patterns of differing effects for single-word and multiword frequencies across the frequency continuum. They found that the effect of multiword frequency information increased with repeated use, but the effect of word-level frequency information decreased when participants produced high-frequency multiword sequences. At this point, it is important to explore the usage-based notion of chunkedness that positions the frequency and probability of input at the core of processing (Bybee & McCelland, 2005; Christiansen, & Chater, 2016; Ellis, 2002; Goldberg, 2006; Siyanova-Chanturia, 2015; Tomasello, 2003).

These usage-based approaches have suggested that frequently used sequences become more accessible and more entrenched. This does not mean that frequently co-occurring multiword sequences are stored and retrieved as unanalyzed holistic units that lack internal analysis as Wray (2002, 2008) has claimed. Instead, usage-based approaches (e.g., Bybee, 2008, Ellis, 2002; Siyanova-Chanturia, 2015) have suggested that frequently co-occurring multiword sequences result in the growing prominence of the sequence relative to the parts, yet information related to the parts is still accessible. The present study (mirroring the findings of Arnon & Cohen Priva, 2014) provides empirical support to usage-based notions of chunkedness in two ways. First, the participants' processing was affected by word-level frequency information for processing collocations, which suggests that collocations are not stored holistically. Second, the effect of the word-level frequency information of nouns differed depending on the frequency of the collocations. Furthermore, usage-based approaches to language acquisition predict that the cumulative experience that speakers have with a target language appears to similarly impact both L1 and L2 speakers (Ellis, 2002). The results of the present study and the study by Wolter and Yamashita (2018) have provided evidence that L1 and L2 speakers' processing is affected by word-level and collocation-frequency information.

Our third research question focused on L1 and L2 speakers' sensitivity to the strength of collocations as determined by the mutual information and log Dice measures. Based on previous literature (e.g., McCauley & Christiansen, 2017), we had predicted that there would be differences between L1 and L2 participants' sensitivity to collocational strength. This prediction was not supported because the participants in both groups were similarly sensitive to both association statistics. It is possible to say that language users are sensitive to the strength of collocations irrespective of their identity as L1 and L2 speakers. Previous studies have produced conflicting results regarding L1 and L2 speakers' sensitivity to association statistics. For example, McCauley and Christiansen (2017) found that L2 learners' chunking scores improved in the raw frequency-based version of their computational model, but L1 child and adult speakers' chunking performance improved in the model based on mutual information scores. They concluded that there may be important differences between the way L1 and L2 speakers chunk and that these differences cannot be explained only on the basis of amount of exposure. Yi (2018) found that L2 speakers were more sensitive to the mutual information scores (that also highlight rare component word frequency) than were L1 speakers. He concluded that language users are sensitive to the statistical regularities regardless

of their identity as L1 and L2. The present study and Yi's (2018) study are comparable because both studies used a similar task with adjective–noun collocations and with a fairly advanced group of L2 speakers. One possible reason for the differences in results between the two studies could be related to the fact that we recruited a group with higher L2 proficiency than Yi did, that is to say, as the level of L2 proficiency increases, L2 speakers' sensitivity to association statistics becomes more and more L1-like.

A further point for discussion is the log Dice score in relation to the mutual information score. According to the AIC values, the model including the log Dice scores was a better-fitting model than the model with mutual information scores. This is not a surprising finding considering the features of the two measures. As Gablasova et al. (2017) observed, the log Dice measure is somewhat similar to the mutual information measure because it is designed to highlight exclusive word pairs. However, unlike the mutual information measure, it does not highlight rare exclusivity. In other words, the log Dice measure does not reward lower-frequency combinations. We can show the inconsistency of the mutual information measure with an example from the high-frequency collocations used in the current study. *Social policy*, one of the high-frequency collocations (raw frequency = 876, mutual information = 3.74), received a considerably lower mutual information score than did *annual report*, another high-frequency collocation (raw frequency = 641, mutual information = 5.78). However, these same two high-frequency collocations obtained fairly similar log Dice scores (7.19 and 7.13, respectively). The nouns *report* and *policy* have fairly similar raw frequency counts; however the adjective *social* (raw frequency = 41,649) occurs more frequently than the adjective *annual* (raw frequency = 8,117) in the BNC XML edition. In this case, we can conclude that the mutual information measure tends to highlight infrequent collocations whose components "may also be infrequent themselves" (Schmitt, 2012, p. 6; see also Garner et al., 2018).

Limitations and Future Directions

Although the present study sheds light on L1 and L2 speakers' sensitivity to frequency and association statistics while processing adjective–noun collocations, there are some limitations that need to be acknowledged. First, the acceptability judgment task used in this study may not be the most ideal one for examining the possible qualitative differences between L1 and L2 speakers' processing of multiword sequences. This task likely requires participants to reflect on adjective–noun pairs, and thus the response times may indicate metalinguistic based processing rather than automatic (subconscious) processing.

Second, most of the two-word adjective–noun collocations used in this study are likely to be considerably more frequent than the three-word sequences that have been used in some previous psycholinguistic studies (e.g., Arnon & Cohen Priva, 2014). Therefore, our findings are limited to two-word collocations and should not be generalized to other types of multiword sequences. It should also be noted that, in this study, we sampled a highly proficient adult L2 population, and we acknowledge that these findings may not apply to L2 populations at other proficiency levels or for other age groups. Another limitation of this study is that some of our items had cognates in Turkish. It would be ideal to fully eliminate them because they might be associated with faster processing (Lemhöfer et al., 2008).

In order to gain a more comprehensive understanding of the processing of multiword sequences, future research needs to focus on L2 populations at different proficiency levels and on individual differences among L1 and L2 speakers, including both personal variables such as length of living in a place where the L2 is spoken and cognitive variables such as declarative memory. Furthermore, future research should also look at the processing of multiword sequences other than collocations, such as three- or four-word lexical bundles, to broaden the scope of the research. This research adopted a frequency-based approach and drew on corpus evidence to identify collocations. However, to reach a more complete picture of collocational processing, future research should focus on semantic relations between words (e.g., Gyllstad & Wolter, 2016) and on L1 and L2 speakers' intuitions of semantic unity of collocations alongside their frequency counts. It is also crucial to acknowledge the importance of previous work at the intersection of experimental and corpus-based approaches to our understanding of the use and processing of multiword sequences. For example, Rebuschat, Meurers, and McEnery (2017) brought together researchers in cognitive psychology, corpus linguistics, and developmental psychology. This type of multimethod approach is particularly useful for research into the processing and learning of multiword sequences. Corpora, as large databases, can provide direct information about language users' word selection and co-selection and reveal regularities in collocational patterns produced by L1 and L2 users, patterns that allow researchers to hypothesize about the variables involved in the acquisition, processing, and representation of collocations (Gablasova et al., 2017). Psycholinguists should explore language users' sensitivity to various aspects of the distributional information including frequency, association, directionality, and dispersion. Corpora can serve to provide researchers with association measures that capture different dimensions of collocational relationships such as directionality (delta P) and dispersion

(Cohen's *d*). In future, researchers should critically evaluate the contribution of these association measures (Gablasova et al., 2017) and investigate language users' collocational processing through the lens of these different measures. One default association measure is unlikely to cover all purposes, no matter how popular it is.

Conclusion

The present study contributes to the growing body of research showing that both L1 and L2 speakers are sensitive to the frequency distributions of multiword sequences at multiple grain sizes. Indeed, we found that L1 and L2 speakers show sensitivity to both word-level and collocation frequency information simultaneously while processing adjective–noun collocations. Furthermore, the effects of word-level and collocation level frequency information differ for processing low- and high-frequency collocations for both L1 and L2 speakers. As the frequency of the collocations increases, the effect of noun frequency information becomes weaker. It is possible to say that repeated use of multiword sequences leads to growing prominence of the whole sequence, but the information about the parts is still accessible. Finally, there was no difference between the L1 and L2 groups in sensitivity to association statistics irrespective of how the association statistics were operationalized. The findings of the present study are in line with the predictions of the usage-based approaches that the cumulative experience speakers have with a target language appears to similarly impact both L1 and L2 speakers (Ellis, 2002).

Open Research Badges

This article has earned Open Data and Open Materials badges for making publicly 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/dxvak/. All proprietary materials have been precisely identified in the manuscript.

Final revised version accepted 26 May 2020

Notes

1 We used the LexTALE (Lexical Test for Advanced Learners of English) as a proxy for general English proficiency. It enabled us to quickly and reliably identify learners with advanced knowledge of vocabulary. In a large-scale validation study,

LexTALE scores were found to be good predictors of vocabulary knowledge and a fair indicator of general English proficiency (see Lemhöfer & Broersma, 2012).

2 We chose mixed-effects models because they allow including both participant and item as random effects. This enables researchers to account for individual differences (e.g., individuals who have generally slow or fast response times). It eliminated the need for separate analyses with participants and with items (so called F1 and F2 analyses).

3 Overall we had to exclude only 0.39% of the data. We also excluded a total of 29 items from the analysis (across all items and participants) because they did not receive any responses. Only three response times were slower than 450 milliseconds, and 27 items timed out at 4,000 milliseconds.

4 We used VIF scores to detect strongly correlated variables in the mixed-effects models because they tend to have unstable coefficient estimates and large standard errors (Levshina, 2015). Researchers have used different values for VIF cut-off scores, some of them as strict as 5 and others less strict—for example, 10. To avoid any risk of multicollinearity, we used 5 as our cut-off score.

5 The MuMIn package in R computes the effect sizes of linear mixed-effects models. It produces two R^2 values for a fitted mixed-effects model in two forms: marginal and conditional. Marginal R^2 values are an index of the variance explained by the fixed effects, whereas conditional R^2 values reflect the variance explained by both fixed and random effects.

References

Aksan, Y., Aksan, M., Koltuksuz, A., Sezer, T., Mersinli, Ü., Demirhan, U., … Kurtoglu, Ö. (2012). Construction of the Turkish National Corpus (TNC). In N. Calzolari, K. Choukri, T. Declerck, M. U. Doğan, B. Maegaard, J. Mariani, … S. Piperidis (Eds.), *Proceedings of the 8th International Conference on Language Resources and Evaluation* (pp. 3223–3226). Istanbul, Turkey: European Language Resources Association. Retrieved from https://www.tnc.org.tr

Arnon, I., & Clark, E. V. (2011). Why *brush your teeth* is better than *teeth*—Children's word production is facilitated in familiar sentence-frames. *Language Learning and Development*, *7*, 107–129. https://doi.org/10.1080/15475441.2010.505489

Arnon, I., & Cohen Priva, U. (2014). Time and again: The changing effect of word and multiword frequency on phonetic duration for highly frequent sequences. *The Mental Lexicon*, *9*, 377–400. https://doi.org/10.1075/ml.9.3.01arn

Arnon, I., & Snider, N. (2010). More than words: Frequency effects for multi-word phrases. *Journal of Memory and Language*, *62*, 67–82. https://doi.org/10.1016/j.jml.2009.09.005

Baayen, R. H., & Milin, P. (2010). Analysing reaction times. *International Journal of Psychological Research*, *3*(2), 12–28. https://doi.org/10.21500/20112084.807

Bannard, C., & Matthews, D. (2008). Stored word sequences in language learning. *Psychological Science*, *19*, 241–248. https://doi.org/10.1111/j.1467-9280.2008.02075.x

Bannard, C., Lieven, E., & Tomasello, M. (2009). Modelling children's early grammatical knowledge. *Proceedings of the National Academy of Sciences*, *106*, 17284—17289. https://doi.org/10.1073/pnas.0905638106

Barlow, M., & Kemmer, S. (Eds.). (2000). *Usage-based models of language*. Stanford, CA: The Center for the Study of Language and Information Publications.

Barton, K. (2019). MuMIn: Multi-model inference. R package version 1.43.6 [Computer software]. Retrieved from https://cran.r-project.org/web/packages/MuMIn/index.html

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

Biber, D., Johansson, S., Leech, G., Conrad, S., Finegan, E., & Quirk, R. (1999). *Longman grammar of spoken and written English*. Harlow, England: Longman.

Brezina, V. (2018). Collocation graphs and networks: Selected applications. In P. Cantos-Gómez & M. Almela-Sánchez (Eds.), *Lexical collocation analysis* (pp. 59–83). Cham, Switzerland: Springer. https://doi.org/10.1007/978-3-319-92582-0_4

Brezina, V., McEnery, T., & Wattam, S. (2015). Collocations in context: A new perspective on collocation networks. *International Journal of Corpus Linguistics*, *20*, 139–173. https://doi.org/10.1075/ijcl.20.2.01bre

The British National Corpus (Version 3; BNC XML edition). (2007). Distributed by Bodleian Libraries, University of Oxford, on behalf of the BNC Consortium. Retrieved from http://www.natcorp.ox.ac.uk/

Bybee, J. (2008). Usage-based grammar and second-language acquisition. In P. Robinson & N. Ellis (Eds.), *Handbook of cognitive linguistics and second language acquisition* (pp. 216–236). New York, NY: Routledge.

Bybee, J., & McClelland, J. L. (2005). Alternatives to the combinatorial paradigm of linguistic theory based on domain general principles of human cognition. *Linguistic Review*, *22*, 381–410.

Christiansen, M. H., & Chater, N. (2016). The now-or-never bottleneck: A fundamental constraint on language. *Behavioral & Brain Sciences*, *39*, e62. https://doi.org/10.1017/S0140525×1500031X

Church, K. W., & Hanks, P. (1990). Word association norms, mutual information, and lexicography. *Computational Linguistics*, *16*, 22–29. https://doi.org/10.3115/981623.981633

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Hillsdale, NJ: Lawrence Erlbaum.

DeCock, S., Granger, S., Leech, G., & McEnery, T. (1998). An automated approach to the phrasicon of EFL learners. In S. Granger (Ed.), *Learning English on computer* (pp. 67–79). London, UK: Addison, Wesley, Longman.

Durrant, P., & Doherty, A. (2010). Are high-frequency collocations psychologically real? Investigating the thesis of collocational priming. *Corpus Linguistics and Linguistic Theory*, *6*, 125–155. https://doi.org/10.1515/CLLT.2010.06

Durrant, P. (2013). Formulaicity in an agglutinating language: The case of Turkish. *Corpus Linguistics and Linguistic Theory*, *9*, 1–38. https://doi.org/10.1515/cllt-2013-0009

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

Ebeling, S. O., & Hasselgård, H. (2015). Learner corpora and phraseology. In S. Granger, G. Gilquin, & F. Meunier (Eds.), *The Cambridge handbook of learnercorpus research* (pp. 185–206). Cambridge, UK: Cambridge University Press. https://doi.org/10.1017/CBO9781139649414

Ellis, N. C. (2002). Frequency effects in language processing: A review with implications for theories of implicit and explicit language acquisition. *Studies in Second Language Acquisition*, *24*, 143–188. https://doi.org/10.1017/S0272263102002024

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

Erickson, L. C., & Thiessen, E. D. (2015). Statistical learning of language: Theory, validity, and predictions of a statistical learning account of language acquisition. *Developmental Review*, *37*, 66–108. https://doi.org/10.1016/j.dr.2015.05.002

Evert, S. 2005. *The statistics of word co-occurrences: Word pairs and collocations* (Doctoral thesis, University of Stuttgart, Stuttgart, Germany).

Evert, S. (2008). Corpora and collocations. In A. Lüdeling & M. Kytö (Eds.), *Corpus linguistics: An international handbook* (pp. 1212–1248). Berlin, Germany: Mouton de Gruyter.

Ferrand, L., New, B., Brysbaert, M., Keuleers, E., Bonin, P., Meot, A., … Pallier, C. (2010). The French Lexicon Project: Lexical decision data for 38,840 French words and 38,840 pseudowords. *Behavior Research Methods*, *42*, 488–496. https://doi.org/10.3758/BRM.42.2.488

Fox, J., & Weisberg, S. (2019). *An R companion to applied regression* (3rd ed.). Thousands Oaks, CA: Sage.

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

Gablasova, D., Brezina, V., & McEnery, T. (2017). Collocations in corpus-based language learning research: Identifying, comparing and interpreting the evidence. *Language Learning*, *67*(S1), 155–179. https://doi.org/10.1111/lang.12225

Garner, J. R., Crossley, S. A., & Kyle, K. (2018). Beginning and intermediate L2 writer's use of n-grams: An association measures study. *International Review of Applied Linguistics in Language Teaching*, *58*, 51–74. https://doi.org/10.1515/iral-2017-0089

Garner, J., Crossley, S. A., & Kyle, K. (2019). N-gram measures and L2 writing proficiency. *System*, *80*, 176–187. https://doi.org/10.1016/j.system.2018.12.001

Goldberg, A. (2006). *Constructions at work*. Oxford, UK: Oxford University Press.

González Fernández, B., & Schmitt, N. (2015). How much collocation knowledge do L2 learners have?: The effects of frequency and amount of exposure. *International Journal of Applied Linguistics*, *166*, 94–126. https://doi.org/10.1075/itl.166.1.03fer

Gries, S. Th. (2008). Dispersions and adjusted frequencies in corpora. *International Journal of Corpus Linguistics*, *13*, 403–437. https://doi.org/10.1075/ijcl.13.4.02gri

Gries, S. Th. (2013). 50-something years of work on collocations: What is or should be next … *International Journal of Corpus Linguistics*, *18*, 137–166. https://doi.org/10.1075/ijcl.18.1.09gri

Gries, S. Th., & Ellis, N. C. (2015). Statistical measures for usage-based linguistics. *Language Leaning*, *65*, 228–255. https://doi.org/10.1111/lang.12119

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

Hardie, A. (2012). CQPweb–Combining power, flexibility and usability in a corpus analysis tool. *International Journal of Corpus Linguistics*, *17*, 380–409. https://doi.org/10.1075/ijcl.17.3.04har

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

Jackendoff, R. (1997). *The architecture of the language faculty*. Cambridge, MA: MIT Press.

Jolsvai, H., McCauley, S. M., & Christiansen, M. H. (2013). Meaning overrides frequency in idiomatic and compositional multiword chunks. In M. Knauff, M. Pauen, N. Sebanz, & I. Wachdmuth (Eds.), *Proceedings of the 35th Annual Conference of the Cognitive Science Society* (pp. 692–697). Austin, TX: Cognitive Science Society.

Konopka, A., & Bock, K. (2009). Lexical or syntactic control of sentence formulation? Structural generalizations from idiom production. *Cognitive Psychology*, *58*, 68–101. https://doi.org/10.1016/j.cogpsych.2008.05.002

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software*, *82*(13), 1–26. Retrieved from https://www.jstatsoft.org/article/view/v082i13

Lemhöfer, K., Dijkstra, T., Schriefers, H., Baayen, R. H., Grainger, J., & Zwitserlood, P. (2008). Native language influences on word recognition in a second language: A megastudy. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*, 12–31. https://doi.org/10.1037/0278-7393.34.1.12

Lemhöfer, K., & Broersma, M. (2012). Introducing LexTALE: A quick and valid lexical test for advanced learners of English. *Behavior Research Methods*, *44*, 325–343. https://doi.org/10.3758/s13428-011-0146-0

Lenth, R. (2018). emmeans: Estimated marginal means, aka least-square means. R package version 1.2.4 [Computer software]. Retrieved from https://cran.r-project.org/package=emmeans

Levshina, N. (2015). *How to do linguistics with R: Data exploration and statistical analysis*. Amsterdam, the Netherlands: John Benjamins.

McCauley, S. M., & Christiansen, M. H. (2017). Computational investigations of multiword chunks in language learning. *Topics in Cognitive Science*, *9*, 637–652. https://doi.org/10.1111/tops.12258

McClelland, J. L. (2010). Emergence in cognitive science. *Topics in Cognitive Science*, *2*, 751–770. https://doi.org/10.1111/tops.12258

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

McEnery, T., & Hardie, A. (2012). *Corpus linguistics: Method, theory and practice*. Cambridge, UK: Cambridge University Press.

Molinaro, N., Canal, P., Vespignani, F., Pesciarelli, F., & Cacciari, C. (2013). Are complex function words processed as semantically empty strings? A reading time and ERP study of collocational complex prepositions. *Language and Cognitive Processes*, *28*, 762–788. https://doi.org/10.1080/01690965.2012.665465

Nesselhauf, N. (2005). Collocations in a learner corpus. Amsterdam, *The Netherlands: John Benjamins, 14*, (pp. 1–10). Amsterdam: John Benjamins. https://doi.org/10.1075/scl.14

Paquot, M., & Granger, S. (2012). Formulaic language in learner corpora. *Annual Review of Applied Linguistics*, *32*, 130–149. https://doi.org/10.1017/S0267190512000098

Pecina, P. (2009). Lexical association measures and collocation extraction. *Language Resources and Evaluation*, *44*, 137–158. https://doi.org/10.1007/s10579-009-9101-4

Peirce, J. W., Gray, J. R., Simpson, S., MacAskill, M. R., Höchenberger, R., Sogo, H., … Lindeløv, J. (2019). PsychoPy2: Experiments in behaviour made easy. *Behaviour Research Methods*, *51*, 195–203. https://doi.org/10.3758/s13428-018-01193-y

R Core Team. (2012). R: A language and environment for statistical computing [Computer software]. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from http://www.R-project.org

Rebuschat, P., Meurers, D., & McEnery, T. (2017). Language learning research at the intersection of experimental, computational and corpus-based approaches: *Language Learning*, *67*(S1), 6–13.

Robert, C., & Rico Duarte, L. (2016). Semantic richness and aging: The effect of number of features in the lexical decision task. *Journal of Psycholinguistic Research*, *45*, 359–365. https://doi.org/10.1007/s10936-015-9352-8

Rommers, J., Dijkstra, T., & Bastiaansen, M. (2013). Context dependent semantic processing in the human brain: Evidence from idiom comprehension. *Journal of Cognitive Neuroscience*, *25*, 762–776. https://doi.org/10.1162/jocn_a_00337

Sato, S., & Athanasopoulos, P. (2018). Grammatical gender affects gender perception: Evidence for the structural feedback hypothesis. *Cognition*, *176*, 220–231. https://doi.org/10.1016/j.cognition.2018.03.014

Schmitt, N. (2012). Formulaic language and collocation. In C. Chapelle (Ed.), *The encyclopedia of applied linguistics* (pp. 1–10). New York, NY: Blackwell.

Shatzman, K. B., & Schiller, N. O. (2004). The word frequency effect in picture naming: Contrasting two hypotheses using homonym pictures. *Brain and Language*, *90*, 160–169. https://doi.org/10.1016/S0093-934X(03)00429-2

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

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/10.1177/0267658310382068

Siyanova-Chanturia, A., Conklin, K., & van Heuven, W. J. (2011). Seeing a phrase "time and again" matters: The role of phrasal frequency in the processing of multiword sequences. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *37*, 776–784. https://doi.org/10.1037/a0022531

Siyanova-Chanturia, A. (2015). On the "holistic" nature of formulaic language. *Corpus Linguistics and Linguistic Theory*, *11*, 285–301. https://doi.org/10.1515/cllt-2014-0016

Smadja, F., McKeown, K. R., & Hatzivassiloglou, V. (1996). Translating collocations for bilingual lexicons: A statistical approach. *Computational linguistics*, *22*, 1–38.

Snider, N., & Arnon, I. (2012). A unified lexicon and grammar? Compositional and non-compositional phrases in the lexicon. In S. Gries & D. Divjak (Eds.), *Frequency effects in language* (pp. 127–163). Berlin, Germany: Mouton de Gruyter.

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

Tomasello, M. (2003). *Constructing a language: A usage-based theory of language acquisition*. Cambridge, MA: Harvard University Press.

Tremblay, A., Derwing, B., Libben, G., & Westbury, C. (2011). Processing advantages of lexical bundles: Evidence from self-paced reading and sentence recall tasks. *Language Learning*, *61*, 569–613. https://doi.org/10.1111/j.1467-9922.2010.00622.x

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

Vespignani, F., Canal, P., Molinaro, N., Fonda, S., & Cacciari, C. (2010). Predictive mechanisms in idiom comprehension. *Journal of Cognitive Neuroscience*, *22*, 1682–1700. https://doi.org/10.1162/jocn.2009.21293

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

Wiechmann, D. (2008). On the computation of collostruction strength: Testing measures of association as expressions of lexical bias. *Corpus Linguistics and Linguistic Theory*, *4*, 253–290. https://doi.org/10.1515/CLLT.2008.011

Wolter, B., & Gyllstad, H. (2011). Collocational links in the L2 mental lexicon and the influence of L1 intralexical knowledge. *Applied Linguistics*, *32*(4), 430–449. https://doi.org/10.1093/applin/amr011

Wolter, B., & Gyllstad, H. (2013). Frequency of input and L2 collocational processing: A comparison of congruent and incongruent collocations. *Studies in Second Language Acquisition*, *35*, 451–482. https://doi.org/10.1017/S0272263113000107

Wolter, B., & Yamashita, J. (2015). Processing collocations in a second language. A case of first language activation? *Applied Psycholinguistics*, *36*, 1193–1221. https://doi.org/10.1017/S0142716414000113

Wolter, B., & Yamashita, J. (2018). Word frequency, collocational frequency, L1 congruency, and proficiency in L2 collocational processing: What accounts for L2 performance? *Studies in Second Language Acquisition*, *40*, 395–416. https://doi.org/10.1017/S0272263117000237

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.

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

Supporting Information

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

Supplementary Material

Appendix S1. Mathematical Equations of the Association Measures, Mutual Information and Log Dice.

Appendix S2. Additional Information for Material Development.

Appendix S3. Complete List of Experimental Items.

Appendix S4. Additional Figure and Tables for Data Analysis.

Appendix S5. R Code for Statistical Analyses.

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

Which Factors Influence First and Second Language Speakers' Processing of Collocations?

What This Research Was About and Why It Is Important

Although it is possible to use language in highly creative ways and convey the same meaning using different words each time, we rarely fulfil language's creative potential. That is, most language we encounter is repetitive and formulaic. These formulaic expressions include collocations (fast food), binomials (knife and fork), and idioms (kick the bucket). Collocations are fundamentally important to language processing due to their versatility—you can match multiple adjectives (good) with multiple nouns (day/boy). Repetitive use enables our brains to process what we hear easily because we expect certain words to follow each other. Some of these are more frequent than others, so we can compare our processing of more common and less common collocations to discover how our experience of language affects how fast we process word sequences. This study aimed to find out which factors affect first (L1) and second (L2) language speakers' processing of collocations.

What the Researchers Did

- Participants were 30 adult native-speakers of English and 32 Turkish-born, highly proficient L2 speakers of English. A vocabulary test was used as a basic measure of their proficiency.
- Using a general corpus of English (BNC), three types of adjective–noun combinations were prepared: high-frequency collocations; low-frequency collocations; and baseline items.
- The baseline items consisted of random combinations of adjectives and nouns.
- Single-word frequency counts of adjectives, nouns, and whole collocations were extracted from the BNC.

- An acceptability judgment task was used to measure participants' processing speeds for high- and low-frequency collocations. In this task participants had to indicate whether collocations were commonly used phrases in English by pressing either a YES or a No button.
- Response times were compared across participant groups and item types. Statistical models were used to identify the effects of (a) individual word frequency, (b) collocation frequency, and (c) the strength of association between the two words (e.g., "old" and "man") on participants' speed of responding in the task.

What the Researchers Found
- Both the frequency of individual words and the frequency of the collocations themselves affected L1 and L2 speakers' processing. This effect was seen for both high- and low-frequency collocations.
- For both L1 speakers and highly proficient L2 speakers, highly frequent collocations were easier to process, and component word frequency had less impact on how easy it was to process these high frequency collocations.
- The strength of association between the words affected processing the collocation for both L1 and L2 speakers.

Things to Consider
- L1 and L2 speakers can process collocations in similar ways. That is to say, the cumulative experience speakers have with a language can impact both L1 and L2 speakers similarly.
- For both groups, repeated use of collocations leads to growing prominence of the collocations as whole phrases, but the information about the single words within a collocation is still accessible.

Materials and Data: Materials and data are publicly available at https://osf.io/dxvak/.

How to cite this summary: Öksüz, D., Brezina, V., & Rebuschat, P. (2020). Which factors influence first and second language speakers' processing of collocations? *OASIS Summary* of Öksüz et al. (2021) in *Language Learning*. https://oasis-database.org

This summary has a *CC BY-NC-SA* license