Individual word and phrase frequency effects in collocational processing: Evidence from typologically different languages, English and Turkish

Doğuş Öksüz [(1)], Vaclav Brezina[(1)], Padraic Monaghan [(1)] and Patrick Rebuschat [(1,2)]

[(1)] Lancaster University, [(2)] University of Tübingen

Author Note: Correspondence concerning this article should be addressed to: Doğuş Öksüz, d.c.oksuz1@lancaster.ac.uk, the Department of Linguistics and English Language, Lancaster University. The data analysis scripts and stimuli for both corpus analysis and reaction time studies are available on the Open Science Framework platform, https://osf.io/muwjz/.

Abstract

Collocations are understood to be integral building blocks of language processing, alongside individual words, but thus far evidence for the psychological reality of collocations has tended to be confined to English. In contrast to English, Turkish is an agglutinating language, utilising productive morphology to convey complex meanings using a single word. Given this, we expected Turkish speakers to be less sensitive to phrasal frequencies than English speakers. In Study 1, we conducted a corpus analysis of translation-equivalent adjective-noun collocations (e.g. *front door*), and found differences between the two languages in frequency counts. In Study 2, we conducted a reaction time experiment to determine the sensitivity of native speakers of English and Turkish to the frequency of adjectives, nouns and whole collocations. Turkish speakers were less sensitive to whole-phrase frequencies, as predicted, indicating that collocations are processed less holistically in Turkish than English. Both groups demonstrated that processing collocations involves combining information about individual words and phrases. Taken together, we show that speakers are sensitive to frequency information at multiple grain sizes that are attuned to the typology of different languages.

Keywords: Collocations, Multiword units, Word frequency, Phrase frequency, Typology.

Introduction

Despite its enormous creative potential, much of the language that speakers produce on a daily basis is largely 'formulaic'. Corpus analyses have indicated that recurring multiword units, variably sized continuous and discontinuous phrases,[1] constitute a high proportion (up to 50%) of the language produced by English native-speakers in both written and spoken discourses (e.g. DeCock et al., 1998). That is, words are very likely to co-occur in specific linguistic configurations, known as multi-word units. Recent theoretical advances in the language sciences, including emergentist, usage-based, exemplar-based models of language, have therefore viewed multiword units, like individual words, as integral building blocks of language processing (Arnon et al., 2017; Christiansen & Chater, 2016; Goldberg, 2006). In these approaches, language is not viewed as a system of abstract symbolic rules - like the words-and-rules approach (e.g., Chomsky, 2000) - but as a statistical accumulation of experiences that changes every time linguistic input (a word or phrase) is encountered (e.g., Arnon, 2021). These approaches emphasize the role of experience in how people use language, including the ability to track statistical properties of language itself (Saffran & Kirkham, 2018; Rebuschat et al., 2021).

There is mounting psycholinguistic evidence that children and adults are sensitive to the statistical properties of multiword units both in their native and additional languages (e.g., Arnon & Clark, 2011; Siyanova-Chanturia et al., 2011). For example, developmental studies have shown that 1-year olds are able to distinguish between frequent and infrequent multiword units (*clap your hand* versus *take your hands*) before they begin producing them (Skarabela et al., 2021). Four-year olds produce irregular plurals more accurately when they occur inside

[1] They include collocations (e.g., *front door*), binomials (e.g., *bread and butter*), idioms (e.g., *kick the bucket*), lexical bundles (e.g. *in the middle of*), and speech routines (e.g. *How's it going*).

frequent multiword units (such as *brush your teeth*; Arnon & Clark, 2011). Adult speakers also show sensitivity to the frequency with which multiword units occur across different experimental paradigms including phrasal decision (*don't have to worry* versus *don't have to wait*; Arnon & Snider, 2010), self-paced reading (e.g., Reali & Christiansen, 2007), priming (Cangir et al., 2017), eye-tracking (e.g., Siyanova-Chanturia et al., 2011; Vilkaite, 2016), and event-related potentials (Pulido, 2021; Siyanova-Chanturia et al., 2017). Adults are sensitive to the frequency of multiword units when they produce language (e.g., Janssen & Barber, 2012), and there is growing evidence that adult speakers are sensitive to the frequency of words, like internal unigram and bigram level frequency, making up multiword units (Arnon & Cohen Priva, 2014; Öksüz et al., 2021). In addition to frequency effects, adult speakers are sensitive to the meaningfulness or compositionality of phrases since the more meaningful a multiword unit was judged to be, the quicker it was processed (Jolsvai et al., 2020). Finally, similar to Age-of-Acquisition effects observed for individual words (e.g., Ellis & Morrison, 1998), multiword units acquired early in childhood are processed faster in adulthood (Arnon et al., 2017).

Taken together, these findings demonstrate that language users represent phrases alongside individual *words* in their language processing and production system (see also Arnon et al., 2017; Arnon & Christiansen, 2017 for similar arguments). However, there is uncertainty as to the extent to which such representation is generalizable across typologically diverse languages. Importantly, so far, the vast majority of experimental work has focused on a narrow range of primarily Indo-European languages, especially English, to examine the processing of multiword units. This leaves open whether the findings of these experiments are generalizable across languages that represent linguistic diversity. Christiansen et al. (2022) have argued that lack of typological diversity is problematic for research on language learning and processing but also for psychology and cognitive science more generally, and systematic comparisons

between different languages would play a major role in yielding a complete understanding of how languages are learned. Regarding multiword units, it is necessary to examine their processing in typologically different languages employing a comparative perspective, with respect to the effects of individual words and whole phrase frequency counts. In this article, we investigate whether the typological characteristics of languages impact the processing of multiword units.

There is insufficient evidence to conclude that multiword units are realized in similar ways across all languages (Arnon, 2021). For example, languages with more complex morphology and flexible word order have fewer recurring multi-word units than other languages. This is because having different morphological markings requires that the same utterance will necessarily have different forms, depending on the gender, or number of participants. Russian is a good example of this, with fewer frequently occurring multiword units than English or German (Stoll, Abbot-Smith & Lieven, 2009). Moreover, depending on the typological characteristics of the language, multiword units may play a more facilitative role in learning grammatical relations in some languages than in others. In languages like English, grammatical relations typically take place between words, and this translates into multiword units (Arnon, 2021; Arnon & Christiansen, 2017; Siegelman & Arnon, 2015). For example, English marks definiteness in its article-system; children acquiring English as a native-language rely on article-noun pairings to master the use of definite and indefinite articles. In polysynthetic or agglutinating languages like Turkish, however, children rely on the information within a single "multimorphemic" word to learn the grammatical relations since many grammatical relations take place within a single multimorphemic word (e.g., Allen, 2017; Courtney & Saville-Troike, 2002).

A prominent type of multiword units that has been particularly well-researched in psycholinguistics is collocations. These can be defined as an "association between lexical

words so that the words co-occur more frequently than expected by chance" (Biber et al., 1999, p. 998). As this definition implies, collocations are usually extracted from large corpora based on quantitative evidence of word co-occurrence. Collocations are fairly heterogenous constructions, including semantically noncompositional word combinations (e.g., *run a risk*, *draw a conclusion*) as well as combinations of lexical items (e.g., *young man*, *dark night*) (Gablasova et al., 2017). In this paper, we investigate collocational processing as a useful test-case to investigate the predictive relations between adjectives and nouns in two typologically distinct languages, English and Turkish.

Individual word and phrase frequency effects in comprehension and production of collocations

There is compelling evidence that both individual word and phrasal frequencies, the frequency of co-occurrence of words, affect the comprehension and production of collocations, together with other multiword units. For comprehension of English adjective-noun collocations, Jacobs et al. (2016) showed that recognition memory for them was impacted, albeit in different ways, by the frequencies of whole phrases and individual words. Examining processing of adjective-noun collocations, Öksüz et al. (2021) found that speakers' response times were affected by the frequencies of nouns and whole collocations. Speakers' sensitivity to noun frequencies depended on the frequency of collocations; as the frequency of the collocations increased, the impact of the noun frequencies on response times became weaker (see also Arnon & Cohen Priva, 2014 for similar findings). Such findings highlight that frequently co-occurring multiword units lead to the growing prominence of the larger unit relative to the parts, yet information related to the parts is still accessible. Similarly, Dutch speakers' production of adjective-noun and determiner-adjective-noun phrases was affected by both phrasal and individual word frequencies (Shao et al., 2019). For adjective-noun pairs, the speech onset

latencies were affected by the frequencies of whole phrases and adjectives but not by the frequencies of nouns. However, for determiner-adjective-noun phrases, the frequencies of nouns and whole phrases impacted the speech onset latencies. Shao et al. (2019) argued that for adjective-noun pairs, frequency-sensitive retrieval processes for nouns were taking place too late to impact phrase onset latencies. However, for determiner-adjective-noun phrases, as the form of Dutch definite articles depend on the grammatical gender of the noun, information regarding nouns was retrieved before speech onset and thus noun frequencies were obtained.

There is also strong evidence that phrasal frequency appears to impact processing in morphologically complex languages. Cangir et al. (2017) demonstrated priming effects for Turkish verb-noun and adjective-noun collocations. The priming effect for verb-noun collocations was stronger than for adjective-noun collocations, though the study did not include morphological forms of collocations. Recently, Vilkaite-Lozdiene (2022) provided eye-tracking evidence that different morphological forms of Lithuanian, a morphologically rich language, verb-noun collocations, including past-tense 3rd person, infinitive and participle forms, were read more quickly than novel word combinations. This suggests that collocations can be used in different morphological forms without losing their processing advantage. Vilkaite-Lozdiene also found that adult Lithuanian native-speakers were sensitive to both form collocation frequency (i.e. the frequency of the verbatim phrase used in the study) and inflectional family frequencies of collocations (i.e. the sum of frequency counts of all the potential inflectional forms of the collocations).[2] Thus, speakers' overall experience with different morphological forms of the same collocation seem to contribute to processing, and the morphological forms are somehow related rather than existing as entirely different collocations on their own.

[2] In this study we also operationalised noun and collocation frequencies as form-frequency (also known as surface frequency), the frequency of the verbatim phrase occurring in the corpus, and inflectional family frequency, the sum of frequency counts of all the potential inflectional forms of the collocations (see also Nagy, Anderson, Schommer, Scott & Stallman, 1989).

In this paper, we build on the above reported findings and directly address the processing of adjective-noun collocations in two typologically distinct languages, English and Turkish, through a comparative approach. Determining the extent to which this typological variation affects the prominence of individual word and collocation (i.e., phrasal) frequencies in processing will expand our understanding of how processing sequences of words is impacted by language-specific statistical properties.

The Current Study

In the present study, we examined the processing of adjective-noun collocations in English and Turkish, conducting a corpus analysis (Study 1) and processing experiments (Study 2). The corpus analysis, using comparable, large, and balanced reference corpora, investigated the frequencies and associations of adjective-noun collocations in both languages. Doing so provides insight into speakers' cumulative experience with adjective-noun collocations in both languages.

> *Research Question 1:* To what extent do English and Turkish adjective-noun collocations have similar frequencies and association strength?

To answer this first research question, in Study 1, we conducted a corpus analysis to examine the differences between English and Turkish speakers' experience with adjective-noun collocations. Specifically, we compared the frequencies and associations of adjective-noun collocations extracted from the reference corpora of both languages. We predicted that form-frequency of Turkish adjective-noun collocations would have considerably lower frequency counts than English ones, because the surface collocations in English potentially subsume the equivalents of the uninflected as well as the inflected forms in Turkish. Such findings shed light on how collocational relationships are realized in morphologically isolating and agglutinating languages. We then conducted acceptability judgment tasks in Study 2, to

evaluate English and Turkish native-speakers' processing of adjective-noun combinations in their respective languages, to address the following research questions.

Research Question 2: To what extent is there a difference between English and Turkish native-speakers' sensitivity to single word-level and collocation frequency information when processing collocations?

Research Question 3: To what extent is there a difference between English and Turkish native-speakers' sensitivity to single word-level frequency information when processing high- and low-frequency collocations?

Research Question 4: To what extent is there a difference between English and Turkish native-speakers' sensitivity to the strength of adjective-noun collocations when processing collocations?

To answer the second and third research questions, in Study 2, we examined the prominence of single-word and collocation frequency information for processing English and Turkish high- and low-frequency collocations by the native-speakers of both languages. A key hypothesis grounding our approach is that speakers are sensitive to the statistical properties of multiword units at multiple grain sizes. We thus expected that speakers of both languages would be sensitive to individual word and collocation frequencies while processing two-word collocations, consistent with predictions from usage-based models (see also Jacobs et al., 2016; Shao et al., 2019). We also predicted that speakers of English and Turkish (i.e. as typologically and structurally different languages) would differ in their reliance on individual word and collocation frequencies. This is because the statistical frequency of individual words and phrases are different in the two languages and speakers' processing mechanisms adapt to the rich diversity of statistical properties of their respective languages. Crosslinguistic comparison of languages with different distributional properties provides evidence not only that the availability of cues differs between languages but also that certain cues influence learning and

processing more strongly in some languages than others (e.g. Vigliocco, Antón-Méndez, Collina, & Frauenfelder, 2008). Here we predict that Turkish speakers would show less reliance on phrasal frequencies than English speakers, as they are less experienced in processing recurrent phrases than English speakers. For individual word frequency, we predicted that Turkish speakers would show greater reliance than English speakers due to the differences in general processing biases between the two languages. Finally, we predicted that the frequency of the phrases would impact both language speakers' sensitivity to individual word-level frequencies (Arnon & Cohen Priva, 2014; Öksüz et al. 2021).

To answer the fourth research question, in Study 2 we examined English and Turkish speakers' sensitivity to the collocations' association strength, as measured by log Dice, a corpus-derived measure of association. We had no strong predictions regarding research question 4, but investigated this as a test of whether frequency effects that may be observed in addressing research questions 2 and 3 are due to frequency or may be better explained in terms of association strength.

Study 1: Corpus Analysis of Turkish and English Adjective-Noun Collocations

In this study, we present the results of a corpus analysis examining the form and inflectional family frequencies and also collocation strength of adjective-noun collocations in the two languages. In the analysis, we focused on adjective-noun collocations for two reasons. The first reason is that nouns within adjective-noun collocations can be inflected with various types of suffixes in Turkish (including case marking, plural and instrumental), and thus it is possible to observe the influence of the agglutinative morphology on the collocability of adjectives and nouns through a corpus analysis. The second reason is that adjective-noun collocations occur in a certain syntactic order in which adjectives precede the nouns in both Turkish and English,

hence they should be comparable in terms of constructing the meaning of adjective-noun collocations from its parts.

Previous research showed that the nature of multiword sequences in Turkish differs from English. For example, Biber's (2009) corpus study of academic English identified 140 four-word multiword units which occurred with a frequency of at least ten per million words, whereas Durrant (2013), using a comparable Turkish academic corpus identified only 18 four-word phrases meeting the same frequency criteria. This is because meanings which require two-to-three-word phrases in English can be conveyed using a single word in Turkish, due to its agglutinative morphology. In other words, both individual word forms and multiword sequences have, on average, considerably lower frequency counts in Turkish than similar forms in English. Crucially, this raises the question of the extent to which Turkish and English speakers' sensitivity to individual word and collocation frequency information differs during processing in their respective native languages. Turkish, an agglutinating language, builds up complex word forms through an extensive range of suffixes. Inflectional suffixes are productive, and they mark functional relations such as case and number. Thus, adjective-noun[3] phrases might involve nouns inflected with a range of case and number suffixes.

In the examples below, we provide adjective-noun phrases in which the nouns are inflected with a range of case and number suffixes to demonstrate how Turkish agglutinates on the noun stems. The notation used here, adapted from Çöltekin (2013), indicates the function of each morpheme.

(1)[3] modern toplum (*modern society*)
<Adj> <N

(2) modern toplum-lar (*modern societies*)
<Adj> <N> <pl>

(3) modern toplum-lar-ın (*modern societies'*)
<Adj> <N> <pl> <gen>

(4) modern toplum-lar-da *(in modern societies)*
<Adj> <N> <pl> <loc>

(5) modern toplum-a- (*to modern society*)
<Adj> <N> <dat>

Method

Corpus data. We analyzed two widely-used general reference corpora of English and Turkish, namely the British National Corpus (BNC XML edition; 2007) and the Turkish National Corpus (TNC; Aksan et al., 2012). We chose corpora that were designed to be balanced and representative of the input that English and Turkish native-speakers experience on a daily basis. We particularly considered the comparability in the genres they include and their sizes (i.e., number of running words). In the BNC XML edition and the TNC 3.0, the distributions of genres are quite similar, and they are both fairly large datasets. The BNC XML edition is a written and spoken corpus with a size of 98,560,118 tokens in 4.048 text samples (with approximately 10 million tokens in the spoken component), from eight domains and 62 genres. It was tagged automatically by CLAWS part-of-speech tagger (Rayson & Garside, 1998). CQPweb, a web-based corpus analysis system (Hardie, 2012), was used to extract adjective-noun collocations from the BNC XML edition. One drawback of the BNC XML is that it is dated, as it includes the texts between 1960-1993. The TNC is a written and spoken general corpus of Turkish with a size of 50,678,199 tokens (approximately 2 million tokens in the spoken component). Following the framework of the BNC, the TNC includes a collection of 4,438 different text samples, representing nine domains, and thirty-nine different genres, covering a period of 24 years 1990-2013. Ideally, we would have preferred to use a Turkish corpus of a bigger size for comparability purposes. However, the fact that the TNC is both automatically tagged for parts of speech and provides morphological annotation of 90 inflectional morphemes that are frequently used in modern Turkish, makes it an ideal data source for this study.

Corpus analysis. We extracted adjective-noun collocations from both corpora. We first established frequency bands to ensure that we extracted an equal number of high-, mid-, and low-frequency adjective-noun pairs in English and Turkish. We used the BNC and TNC word

frequency lists to establish the frequency bands. Frequency distributions of the BNC and TNC word lists largely follow Zipfian distribution, with a very small number of high-frequency and long tail of low-frequency nouns. In the BNC XML edition, the most frequent noun *time* occurred with relative frequency of 1,842.00 per million words and in the TNC, the most frequent noun *iç* (the inside) occurred with relative frequency of 3,362.58 per million words as lemmas. The infrequent nouns such as *assister*, *leaser*, and *dentin* occurred with relative frequency of 0.010 per million words as lemmas in the BNC. The infrequent nouns such as *firek* (*a kind of tomato*) occurred with the relative frequency of 0.02 per million words as lemmas in the TNC.

To establish frequency bands, we decided on the following cut-off scores: For high-frequency words 400 or above, for mid-frequency words between 150 and 300 and for low-frequency words 100 or below relative frequency (i.e., per million words). We determined these cut-off values to establish high-, mid- and low-frequency bands considering the overall frequencies of nouns in both corpora. There are considerably fewer high-frequency nouns than mid- and low-frequency nouns in both corpora, so we identified a cut-off value for the high-frequency band (400 per million words) that would allow us to extract at least 5 translation-equivalent high-frequency nouns in both corpora. Using this band, we could then extract 50 high-frequency translation-equivalent collocations from both corpora. For the mid-frequency band, we set the upper cut-off value at 300 per million words; we chose a substantially lower value than 400 to ensure that we would not extract collocations with similar frequencies to the ones in the high-frequency band. Applying the same logic, for the low-frequency band, we set the upper cut-off value at 100 per million words, to extract collocations on average with lower frequency counts than those in the mid-frequency band. Using these selected nouns, we aimed to extract 50 adjective-noun collocations from each frequency band. For comparability purposes, only the nouns within the same frequency bands in the BNC and TNC were chosen

(see Table 1). Thus, five nouns were extracted at each frequency level. Using these selected nouns, we aimed to extract 50 adjective-noun collocations from each frequency band.

Collocations were extracted through selecting the most frequent ten adjective collocates of each noun from the BNC. Then the translation equivalents of those collocations were obtained from the TNC. For the low-frequency nouns *Forest, Wood,* and *Cloth,* we were only able to extract 8 translation equivalents adjective-noun collocations from the BNC and TNC. This is because there are a fewer number of adjective collocates available for the nouns in the low-frequency band than nouns in the high- and mid-frequency bands. We only extracted adjacent collocations. To examine the collocability of adjectives and nouns in the two languages, we focused on raw (absolute) frequency, counting the co-occurrence of words. For comparability purposes, we first used relative frequencies per million words, which operate on normalized scales (Gablasova et al., 2017). To examine frequency counts on a standardized scale, we converted the relative frequency counts (frequency per million words) to the Zipf scale frequency measure. It operates on a standardized, logarithmic scale from 1 to 7 (van Heuven et al., 2014). To calculate the Zipf scale frequencies, we took log10 (relative frequency counts) + 3.

To examine the strength of collocations in the two languages, we used the log Dice measure designed to highlight exclusivity of collocational relations (Gablasova et al., 2017). It takes the harmonic mean of two proportions that express the lexical attraction of two words, relative to the frequencies of those words in the corpus. The mathematical formula for log Dice is provided in Appendix A. Many previous psycholinguistic studies have used mutual information (MI) as a measure of collocational strength (e.g., McCauley & Christiansen, 2017; Vilkaitė-Lozdienė, 2022), but MI has some important limitations. For example, it favours low-frequency word pairs whose component words are likely to be low-frequency themselves. Thus, the value of mutual exclusivity does not only indicate the exclusivity of collocations but

also how infrequently they occur in corpora (Gablasova et al. 2017). Using the log Dice measure we aimed to highlight the exclusivity of collocations, but not only rare combinations.

Table 1.

Nouns used in the three frequency bands

Frequency band	Node words (English)	Node words (Turkish)	Relative frequency counts in the BNC	Relative frequency counts in the TNC
High-frequency (Relative frequency counts) 400<	Time	Zaman	1360.32 (1607.84)	1690 (2771.75)
	Day	Gün	535.57 (820.76)	1091.06 (2505.17)
	Way	Yol	853.69 (981.56)	517.68 (1655.19)
	Man	Adam	524.05 (852.74)	490.8 (1065.9)
	World	Dünya	512.45 (521.3)	605.01 (1625.1)
Mid-frequency (Relative frequency counts) 150-300	Family	Aile	300.55 (373.66)	235.8 (674.12)
	Country	Ülke	279.71 (427.45)	263.31 (1741.42)
	Society	Toplum	209.54 (247.71)	231.68 (828.42)
	Result	Sonuç	195.44 (295.96)	195.63 (1119.89)
	Industry	Sanayi	176.11 (214.63)	170.94 (277.67)
Low-frequency (Relative frequency counts) < 100	Holiday	Tatil	67.99 (94.01)	46.84 (105.09)
	Wood	Tahta	64 (77.47)	44.08 (74)
	Forest	Orman	63.77 (81.18)	69.42 (165.32)
	Distance	Mesafe	58.18 (69.57)	27.78 (73.46)
	Cloth	Bez	16.98 (18.65)	14.4 (31.57)

Frequency counts for inflectional family frequencies are provided in parentheses.

In this study, we examined the frequency counts and collocational strength of adjective-noun collocations using the form and inflectional family frequencies. For calculating the

frequency counts as a measure of the inflectional family, we used the lemma search function of the CQPweb for English. The morphological annotation of the TNC made it possible to obtain the sum of frequencies of all inflected forms of the collocations. That is, to calculate the morphological family frequencies for the nouns and whole collocations, the sum of the uninflected and inflected forms of the nouns were used as the frequency of the noun in the whole corpus. The frequency sums of all the inflected forms of the collocations were taken as the frequency of the collocation[4]. All data and statistical code used in Study 1 are available on the Open Science Framework: https://osf.io/muwjz/

Results

One hundred and forty-four adjective-noun collocations were extracted from the BNC and TNC, 50 from high- and mid-frequency bands, 44 from low-frequency bands (see Appendix B for a complete list of collocations). Unfortunately, we were only able to extract 8 translation-equivalent adjective-noun combinations for three nouns in the low-frequency band. We compared the form and inflectional family frequencies of the selected collocations in each frequency bands, as Zipf scale frequencies in Table 2.

[4] For instance, we calculated the inflectional family frequency of the collocation *young man* by totalling the frequencies of its base form *young man* (Raw frequency = 2787) and the inflected form *young men* (Raw frequency = 1159) so that the inflectional family frequency of the total is 3946, corresponding to the 6.59 Zipf scale frequency.

Table 2.

Average Frequencies for English and Turkish as form and inflectional family in each band

Frequency bands	English Form / Inflectional family	Turkish Form / Inflectional family
High-frequency	5.25 (8.01) / 5.76 (8.92)	5.81 (10.21) / 14.67 (15.83)
Mid-frequency	0.64 (0.62) / 1.37 (1.02)	0.79 (0.83) / 3.31 (3.24)
Low-frequency	0.36 (0.57) / 0.52 (0.98)	0.15 (0.17) / 0.58 (0.68)

Note. Standard deviations are provided in parentheses.

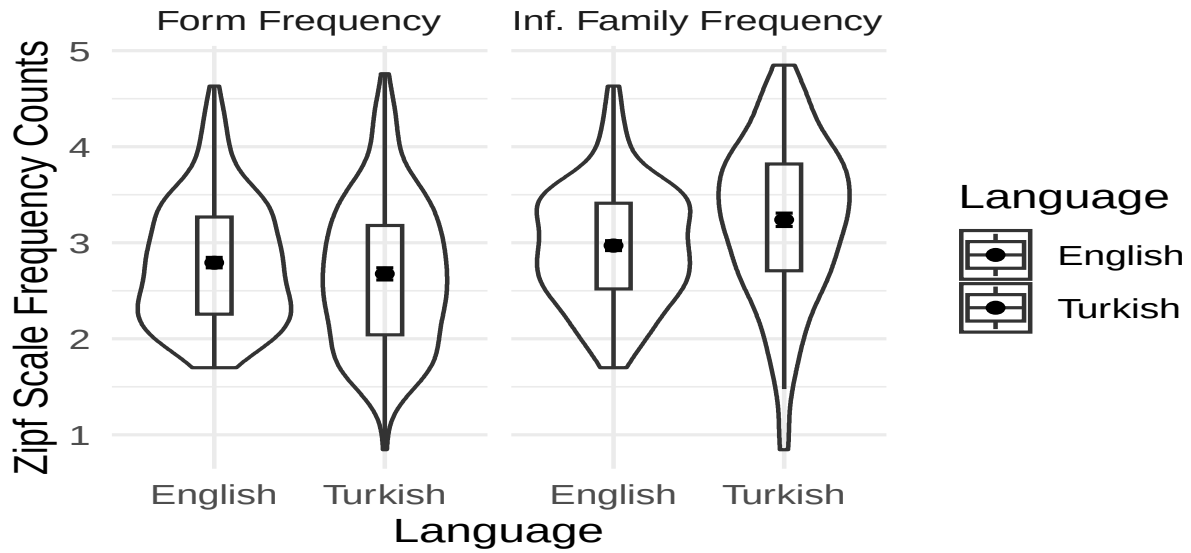
Regarding the frequencies of collocations in English and Turkish the mean form frequencies of English adjective-noun pairs were slightly higher (M = 2.79, SD = 0.64) than the Turkish ones (M = 2.68, SD = 0.76). For the inflectional family frequencies, the Turkish adjective-noun pairs reached a higher mean (M = 3.26, SD = 0.80) frequency counts than the English ones (M = 2.98, SD = 0.61) (see also Figure 1). To model frequency counts of adjective-noun combinations in the two languages inferentially, we used the lme4 package (Bates et al., 2015), constructing a mixed-effects model. We calculated the *p* values using the lmerTest package (Kuznetsova et al., 2017). In Model 1, we included random effect of noun nodes. We added the following variables as fixed effects in Model 1: Language (English vs Turkish), lemmatization status (Form vs Inflectional Family Frequencies), adjective frequencies (i.e., Zipf-scale), and the interaction between language and lemmatization status. Finally, we calculated the variance inflation factor scores using the CAR package in R (Fox & Weisberg, 2019) to check potential multicollinearity issues (VIF scores < 3.05). A table for Model 1 is provided in Appendix C.

There was a main effect of language $b = -0.123$, $SE = 0.056$, 95% CI [−0.234, −0.011], $t(553.95) = -2.16$, $p = .030$, showing that on average, English adjective-noun

collocations had significantly higher frequency counts than Turkish ones. There was also a main effect of lemmatization status b = 0.182, SE = 0.057, 95% CI [0.069, 0.295], t(555.31) = 3.16, p = 0.01. We ran a series of pairwise comparisons tests to decompose the significant interaction effect between language (English vs Turkish) and lemmatisation status (Form vs Inflectional Family Frequencies) b = 0.394, SE = 0.080, 95% CI [0.236, 0.551], t(553.94) = 4.89, $p \leq 001$, using the emmeans package in R with Tukey adjustments for multiple comparisons (Lenth, 2018). The results showed a significant difference between English and Turkish collocations' inflectional family frequencies, Estimate = -0.271, SE = 0.057, df = 554, t = -4.74, p < .001, as Turkish ones had significantly higher frequencies than their English translation equivalents, (see also Appendix C for visual illustrations of this interaction effect).

Figure 1.

Zipf-scale frequency counts of English and Turkish adjective-noun pairs



In addition to frequency, we compared the strength of adjective-noun collocations in the two languages, addressing Research Question 1. We calculated log Dice scores for the extracted adjective-noun collocations using form and inflectional family frequencies

separately. As can be seen in Table 3, log Dice scores based on inflectional family frequencies were higher than those based on form frequencies in both languages. The difference between the two types of log Dice scores in Turkish was larger than that of the English scores (1.03 and .37 respectively), mirroring the results from the frequency comparison. This is because the inflectional family frequencies of the Turkish collocations include frequency counts of base, five different cases (accusative, locative, ablative, dative and genitive), and plural inflected forms, whereas for English they include just the base and plural inflections.

Table 3.

Log Dice Scores

Language	Form frequency	Inflectional family frequency
English	4.30 (2.06), 95% [4.13, 4.47]	4.67 (1.93), 95% [4.51, 4.83]
Turkish	3.74 (2.05), 95% [3.57, 3.91]	4.77 (2.05), 95% [4.60, 4.94]

To model association strength of adjective-noun combinations in the two languages inferentially, we constructed Model 2, a mixed effects model including Log Dice scores as an outcome variable. In Model 2, we included random intercept for noun nodes. As fixed effects, we added the following variables: Language (English vs Turkish), lemmatization status (Form vs Inflectional Family Frequencies), individual word frequencies for adjectives (i.e., zipf-scale frequencies), and the interaction between language and lemmatization status. There was a main effect of language $b = -0.529$, $SE = 0.210$, 95% CI [−0.941, −0.117], $t(558.19) = -2.51$, $p = .012$, showing that on average English adjective-noun collocations had significantly higher Log Dice scores than Turkish ones, and that on average English adjective-noun collocations were more strongly associated than Turkish ones. There was no main effect of lemmatization status $b = 0.383$, $SE = 0.212$, 95% CI [−0.033, 0.800], $t(562.50) = 1.80$, $p = 0.07$. We ran a series of

pairwise comparisons tests to decompose the significant interaction effect between language (English vs Turkish) and lemmatisation status (Form vs Inflectional Family Frequencies) b = 0.689, SE = 0.297, 95% CI [0.106, 1.273], t(558.16) = 2.31, p = 0.20. There was no significant difference between English and Turkish collocations' Log Dice scores when calculated using inflectional family frequencies. Estimate = -0.160, SE = 0.212, df = 562, t = -0.75, p = 0.87. A table for Model 2 is provided in Appendix C.

Discussion

As expected, the corpus analyses indicated that on average English collocations reached higher frequencies than the Turkish collocations. Nevertheless, the Turkish collocations reached higher inflectional family frequencies than the English collocations. Overall, when form frequencies are compared in the two languages, there are fewer high-frequency collocations in Turkish than English (see also Durrant, 2013). Furthermore, inflected forms of the Turkish collocations were relatively rare. To examine whether there is any difference between English and Turkish adjective-noun collocations' strength of association, we compared the English and Turkish collocations' Log Dice scores. English collocations had higher Log Dice scores than their Turkish translation equivalents. This raises the question of whether usage-based factors could explain the difference in processing of collocations in the two languages. The next section presents the acceptability judgment tasks in both languages to investigate similarities and differences in collocational processing in English and Turkish.

Study 2: Acceptability Judgment Experiments

In Study 2, we addressed Research Questions 2-4. Thus, we aimed to determine if there is a difference between English and Turkish speakers' sensitivity to single word, and collocation frequencies (including form and inflectional family) and collocation strength in their respective native languages. Such a finding would support the idea that speakers' processing mechanisms adapt to the rich diversity of statistical properties of their respective languages. We also examined whether speakers' sensitivity to single word and collocation frequencies differs depending on the frequency of the collocations within their native language. If so, this would provide further support that the frequency of the multiword units would cause a difference in the prominence of single word and phrasal level frequencies for processing multiword units.

Method

Participants

Native speakers of English and native speakers of Turkish were recruited to participate in this study. The English native-speaker group (n = 31) were students at Lancaster University (17 women, 13 men, and 1 non-binary; M_{age} = 20.58, SD = 2.16). The Turkish native-speaker group (n = 46) were students at Bogaziçi University in Istanbul (27 women, 19 men, M_{age} = 26.5, SD = 5.51). All participants reported having normal or corrected to normal vision and did not report any language or learning-related disabilities. They provided informed consent, under a protocol that was reviewed and approved by the Institutional Review Board at Lancaster University. They were paid for participation.

Materials

The stimuli in both languages were grouped into the three critical conditions: (a) high-frequency collocations, (b) low-frequency collocations, (c) non-collocational (baseline) items.

We used the non-collocational items for establishing threshold response times and for measuring the relative response times for collocational items in conditions (a) and (b). To be able to closely match individual word frequencies and lengths across conditions, in this study, we decided not to use the wider set of collocations we analyzed in Study 1 as our experimental stimuli. Importantly, we only compared the processing of high- and low-frequency collocations, as we aimed to determine whether speakers' sensitivity to single word and collocation frequencies differed depending on the frequency of the collocations. Examining the processing of high- and low-frequency collocations maximised the possible differences in processing. We extracted single word frequency counts for adjectives, nouns (i.e., form and inflectional family frequencies) and collocation frequencies (i.e., both form and inflectional family). One hundred and twenty English and one hundred and four Turkish adjective-noun collocations served as stimuli. The complete list of stimuli can be found in Appendix D.

The English collocations used in this experiment were taken from the acceptability judgment task of Öksüz et al. (2021). To extract high- and low-frequency collocations, Öksüz et al. (2021) explored the scales of frequencies and log Dice scores using the BNC XML edition (2007). Following the same steps, we created high-frequency, low-frequency and baseline items for Turkish using the TNC V3. That is, we first examined the frequency scales of adjective-noun phrases in the TNC V3. We selected 10 nouns from various frequency counts with a higher-frequency count of 55293 (e.g., *gün, day in English*) and a lower-frequency count of 730 (e.g., *bez, cloth in English*). We used the high-, mid, and low-frequency nouns presented in Table 1. In summary, using those selected nouns, we extracted a pool of 2362 adjective-noun combinations from the TNC V3 (see Appendix D for further details about the materials construction) and compared the frequency distributions of collocations in the two languages. Based on our small-scale corpus analysis, we decided to use the previously used cut-off scores for English to extract high- and low-frequency collocations in Turkish. We defined high-

frequency collocations as adjective-noun pairs with raw frequency counts greater than or equal to 300 (3.45 ≤ Zipf scale) and low-frequency collocations between 10 and 150 (2 < Zipf scale < 3.18).

Applying these criteria to candidate items resulted in 26 Turkish high- and low-frequency collocations. We had 30 English high- and low-frequency collocations. Importantly, the items in the different conditions did not differ in adjective and noun frequency counts, or item length in either of the languages. However, they did differ in collocation frequency counts (item characteristics are summarized in Table 5). The high-frequency collocations had form collocation frequencies between 4.61 – 3.47 in the case of English and between 4.77 – 3.42 in the case of Turkish. The most frequent English collocation was *long time* and the least frequent was *senior officer*; in the case of Turkish the most frequent collocation was *ertesi gün (*next day) and *the* least frequent one *kirmizi sarap* (red wine). The low-frequency collocations had phrasal frequencies between 3.11 – 2.04 in the case of English and between 3.21 – 2.56 in the case of Turkish. The most common English collocation was *difficult life*, and the least common *suitable software*. In Turkish, *fazla üretim* (extra production) was the most common collocation and *yanlış tercih* (wrong choice) the least common one. In collocational conditions, we used each adjective and noun only once. To ensure that adjectives modify the nouns for each of the collocations used, we checked the concordance lines of both corpora. We selected the collocations in a way that their form frequencies did not strongly correlate with the individual word frequencies of the adjectives and nouns, as can be seen in Table 4. We selected items in order to minimize the correlation. Nonetheless, we could not eliminate the positive but weak correlation between English noun and collocation frequencies. This is because we aimed to disentangle the effects of individual word and phrasal frequencies on participants' response times. Identical to the corpus analysis, we only extracted adjacent adjective-noun collocations.

Table 4.

Correlations between individual words and phrasal frequencies

Language	Noun & Phrasal frequencies	Adjective & Phrasal frequencies	Adjective and Noun frequencies
English	$r = 0.18$, $p = 0.043$	$r = 0.08$, $p = 0.369$	$r = -0.02$, $p = 0.794$
Turkish	$r = 0.04$, $p = 0.651$	$r = 0.11$, $p = 0.239$	$r = 0.04$, $p = 0.627$

The baseline items for both languages were created by randomly pairing the nouns used for high-frequency and low-frequency collocations with adjectives that had not been used for either item types. On the one hand, using the same nouns in different conditions was a good way of ensuring that we had perfectly matched the noun frequency counts and word lengths in the collocational and baseline items. On the other hand, using each noun twice in that task inevitably created another potential confound in that it possibly lowered the activation thresholds for the nouns that had been shown to the participants in a different condition. To address this, we presented all the experimental items to the participants in an individually randomised order. Any advantage gained from seeing a word for a second time was evened out both within an individual participants' and across all participants' responses. We also included the nouns' order of occurrence as fixed effects in the mixed-effects models, as detailed in the result section.

As can be seen in Table 5, the length, adjective and noun frequencies, and collocation frequencies in all three conditions were closely matched. For the items in both languages, there is no significant difference in length (i.e., number of letters) and individual word frequencies for adjectives and nouns between the high- and low-frequency collocations. This way we ensured that any processing differences between the two languages and high- and low-frequency collocations would not stem from the frequency differences of individual words. We

checked all paired adjectives and nouns that we had used to construct the baseline items against the BNC XML edition and TNC V3, to make sure that there were no co-occurrences. Given the very large size of the BNC XML edition, it was not possible to fully eliminate all co-occurrences. The English baseline items had raw frequency counts of less than or equal to 10 and 49 out of 52 (94.23%) had frequency counts of 0. All Turkish baseline items had raw frequency counts of 0. We consulted the concordance lines to ensure that any co-occurrences of the remaining 3 baseline items in the corpora were idiosyncratic rather than meaningful co-occurrences.

In addition to the frequencies of collocations, we ensured that the items selected have similar collocational strength in both languages. We employed log Dice as a measure of collocational strength (see also Öksüz et al. 2021; Yi, Man & Maie, 2022, for other experimental studies using log Dice as a measure of collocational association). In both languages, high-frequency collocations had log Dice scores of greater than or equal to 7 and low-frequency collocations had log Dice scores of between 2.5 and 4. The non-collocational (baseline) items in English had a low score of -3.22 (for the item *dirty time*) and a high-score of 0.45 (for the item *clear trade*). Due to the size of the BNC XML edition, it was not possible to fully eliminate all baseline items with positive log Dice scores. We decided to retain two items with positive but very low log Dice scores. The maximum log Dice score included was 0.45 (for the item *clear trade*), and the other positive log Dice score was 0.16 (for the item *public class*). We eliminated all items with positive log Dice scores in Turkish.

Table 5.

Item characteristics for both languages with Wilcoxon tests

Item properties	High-frequency collocations English / Turkish	Low-frequency collocations English / Turkish	Non-collocations (Baseline) English / Turkish	Statistical comparisons English / Turkish
Item length	10.86 (2.97) / 10.65 (2.18)	11.1 (2.3) / 10.65 (2.81)	11.1 (2.52) / 10.42 (2.81)	W = 401, p = .467 / W =314.5, p = .672
Adjective frequency	5.17 (0.31) / 5.39 (0.29)	5.17 (0.42) / 5.33 (0.34)	5.15 (0.24) / 5.29 (0.31)	W = 467.5, p = .400 / W = 345, p = .452
Noun frequency	5.36 (0.29) / 5.32 (0.42)	5.36 (0.21) / 5.33 (0.22)	5.36 (0.25) / 5.33 (0.32)	W = 415.5, p = .697 / W =265.5, p = .909
Collocation frequency	4.03 (0.34) / 4.03 (0.4)	2.7 (0.3) / 2.77 (0.16)	1.18 (0.52) / 0.03 (0.21)	W = 891, p <.001 / W = 652, p <.001
Log Dice scores	7.8 (0.82) / 7.92 (0.77)	3.24 (0.39) /3.6 (0.26)	-0.93 (0.85) / 0	

Note. Item length is in number of letters. Frequencies are provided in Zipf scale, and statistical tests are based on Wilcoxon Rank-Sum test, comparing high-frequency and low-frequency collocations. Mean log Dice scores are provided. Since log Dice is an effect size measure, we did not include an additional statistical test.

To examine the processing of collocations in English and Turkish, we used acceptability judgment tasks for both languages (see also Öksüz et al. 2021; Wolter & Yamashita, 2017). A key assumption underlying the task is that low-frequency collocations should elicit slower response times for both English and Turkish speakers than high-frequency collocations. The baseline items were used to measure English and Turkish speakers' relative response times to the high- and low-frequency collocations.

Procedure

Participants completed acceptability judgment tasks in their respective native languages. The task required participants to determine whether the adjective-noun combinations were commonly used in English and Turkish. We asked participants to indicate, as quickly and as accurately as possible, whether the word combinations are commonly used in their respective native languages, English and Turkish. The stimuli were presented using the PsychoPy software (Peirce et al. 2019). Each trial started with the eye fixation (#########) for 250ms and followed by a blank screen. After the blank screen, the item in lowercase characters (e.g., front door) remained on the screen either until the participants indicate their responses or after a 4000ms timeout. 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 on the game pad. The acceptability judgment task started with a practice session and most participants completed it in 5-6 minutes. Each participant saw all the experimental items. All data, experimental materials, and statistical code used in Study 2 are available on the Open Science Framework: https://osf.io/muwjz/

Results

We firstly removed response times of shorter than 450 milliseconds as well as those that timed out at 4000 milliseconds. These cut-off response times are based on two previous studies, Gyllstad and Wolter (2016) and Wolter and Yamashita (2017), who employed a very similar task. Only 4 responses were faster than 450 milliseconds and 151 items timed out at 4000 milliseconds. Overall, we thus only excluded 1.80% of the data.

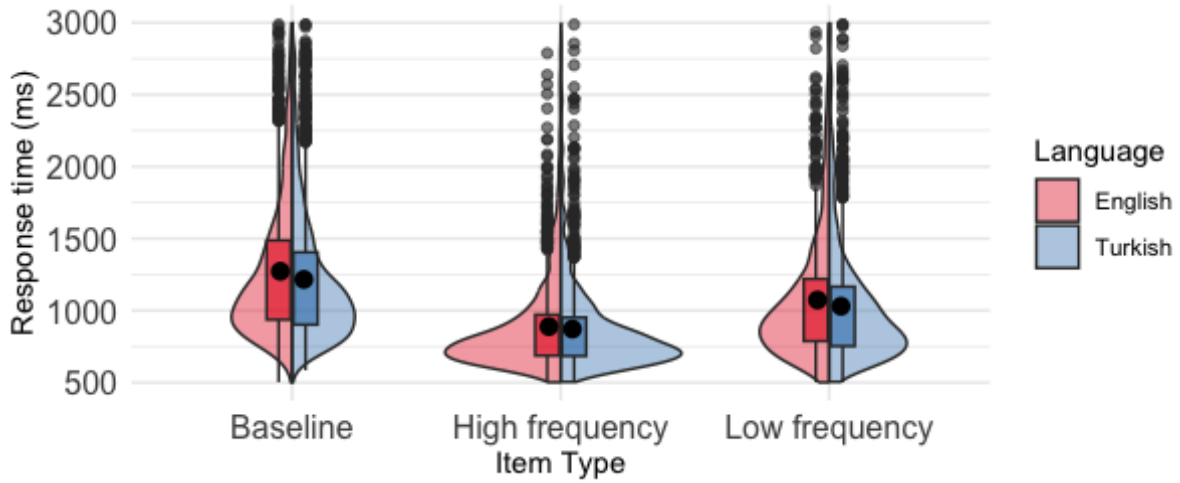
The main concern of the present study was how the English and Turkish native-speaker participants processed the high- and low-frequency collocations that they judged as commonly used compared to the baseline items that they judged as not commonly used. Therefore, we analyzed the high- and low-frequency collocations that received a yes response and compared them to the baseline items that received a no response. The English participants judged 97.75% of the high-frequency and 78.15% of the low-frequency collocations to be commonly used and they decided that 78.75% of the baseline items are not commonly used in English. The Turkish participants judged 98.24% of the high-frequency collocations and 88.12% of the low-frequency collocations to be commonly used and 81.27% of the baseline items to be not commonly used in Turkish[5].

As predicted, response times in both languages were faster for high-frequency collocations (*M* = 893 ms, *SD* = 338 ms) than low-frequency ones (*M* =1076 ms, *SD* = 472 ms) or baseline items (*M* =1304 ms, SD = 528 ms in English), high-frequency collocations (*M* = 875 ms, *SD* = 335 ms), low-frequency collocations (*M* = 1044 ms, *SD* = 457 ms), and baseline items (*M* = 1243 ms, *SD* = 506 ms in Turkish). These distributions are presented in Figure 2.

[5] We also analysed error rate using logistic regression modelling with accuracy treated as a binomial response variable (Correct vs Incorrect). In both languages item type significantly affected error rates. Speakers of both languages responded to the high-frequency collocations more accurately than the baseline items (English High-frequency vs Baseline: Estimate = 2.588, *SE* = 0.244, *z* = 10.576, p <.001, Turkish High-frequency vs Baseline: Estimate = 2.517, *SE* = 0.221, *z* = 11.370, *p* <.001). English speakers did not respond to low-frequency collocations more accurately than the baseline items (English Low-frequency vs Baseline: Estimate = 0.009, *SE* = 0.099, *z* = 0.099, *p* = 0.921. Turkish speakers responded to the low-frequency collocations more accurately than the baseline items Estimate = 0.627, *SE* = 0.106, *z* = 5.917, *p* <.001.

Figure 2.

Distribution of response times (ms) for three item types: baseline, high frequency and low frequency.



In order to test whether the response times in the two languages differed significantly and whether behavior was different between the two language groups, we built three linear mixed-effects models using the lme4 package version 1.1-30.1 (Bates et al., 2015) for the R statistical platform version 4.3.1 (R Core Team, 2022). We calculated the *p* values using the lmerTest package (Kuznetsova et al., 2017). In Model 3, we examined the extent to which English and Turkish speakers are similarly sensitive to collocation frequencies (i.e., as a continuous predictor). In Model 4, we investigated the extent to which English and Turkish speakers are sensitive to single word-level frequency information when processing high- and low-frequency collocations, to address Research Questions 2 and 3. Finally, in Model 5, we analyzed the effect of collocation strength, (i.e., Log Dice scores) on the processing of English and Turkish collocations, addressing Research Question 4.

We used log-transformed response times as our dependent variable because raw response times created convergence issues and may have deviated from a normal distribution. The log-transformed response times had an approximate normal distribution (see Appendix E for visual illustrations of response times). We log transformed all single word and collocational

frequency counts using SUBTLEX Zipf Scale, except baseline items with 0 frequencies (Van Heuven et al. 2014). We set the transformed frequency values for baseline items with zero frequencies as 0. We centered all continuous predictors and treated the first versus second occurrence of nouns as a categorical predictor (i.e., whether the participants were seeing a particular noun for the first or a second time). We recoded the categorical variable language group (English vs Turkish) using contrast coding (English = 0.5, Turkish = -0.5). We coded the item type using the treatment coding in which baseline items were defined as the reference level. Finally, we calculated the variance inflation factor scores using the CAR package in R (Fox & Weisberg, 2019) to check potential multicollinearity issues.

We constructed Model 3 to examine the extent to which English and Turkish speakers are similarly sensitive to collocation frequencies when processing adjective-noun collocations in their respective languages. Due to the multicollinearity problem between collocation frequency and item type, we decided to analyze item type in a separate model. This was an anticipated issue since item type was also based on collocation (i.e., form) frequencies. For Model 3, we defined maximal models as justified by the experimental design (Barr et al., 2013), and we included as random effects subject and item. As random slopes for subject, we included item type, and adjective, noun, and collocation frequencies. Since the language varied between items, these were not included as by-item random slopes. As fixed effects, we had language (English vs Turkish), individual word frequencies for adjectives and nouns (both form and inflectional family), collocation (form and inflectional family) frequencies, length of items (number of letters), and first versus second occurrence of nouns. Moreover, we included all possible interaction terms including the variables of interest: language and collocation frequencies, collocation and individual word frequencies for adjectives and nouns, language and individual word frequencies for adjectives and nouns. Model tables for the maximal models including estimates and test statistics are provided in Appendix E.

Using the above-described maximal model as a starting point, we followed a step-wise model comparison procedure to reduce the number of fixed effects. We identified the best model fit, using the lmerTest package in R and Akaike Information Criterion (AIC) values. The resulting final Model 3 is presented in Table 6. To have a model with non-singular fit, we removed by-subject random slopes for adjective and noun frequencies. The final version of Model 3 included a random intercept for subjects (participants) and items, and by-subject random slopes for the collocation frequency. We found the most complex model consistent with our experimental design (see Barr et al., 2013). In other words, in accordance with Barr et al. (2013) we constructed a linear mixed-effects model with random effects structure that was most complex (in terms of including intercepts and slopes for random effects) such that the model converged and was not singular. As shown in Table 6, there was a main effect of language (English vs. Turkish) on response times, controlling for length, individual word, and collocation frequencies. That is to say, Turkish speakers responded to collocations slightly but significantly faster than English speakers. In addition, for speakers of both languages, collocation (i.e., form) frequencies led to significantly faster response times. One unit of increase in collocation frequency counts resulted, on average, in a -0.165 faster log response time measure, corresponding to 25.512 milliseconds for speakers of both languages. Surprisingly, both adjective and noun frequencies led to slower response times for speakers of both languages, as they took longer to respond to the collocations with more frequent adjectives and noun. Participants needed a longer time to respond to items with more letters.

Table 6.

Mixed effects Model 3, investigating English and Turkish speakers' sensitivity to collocation frequency counts.

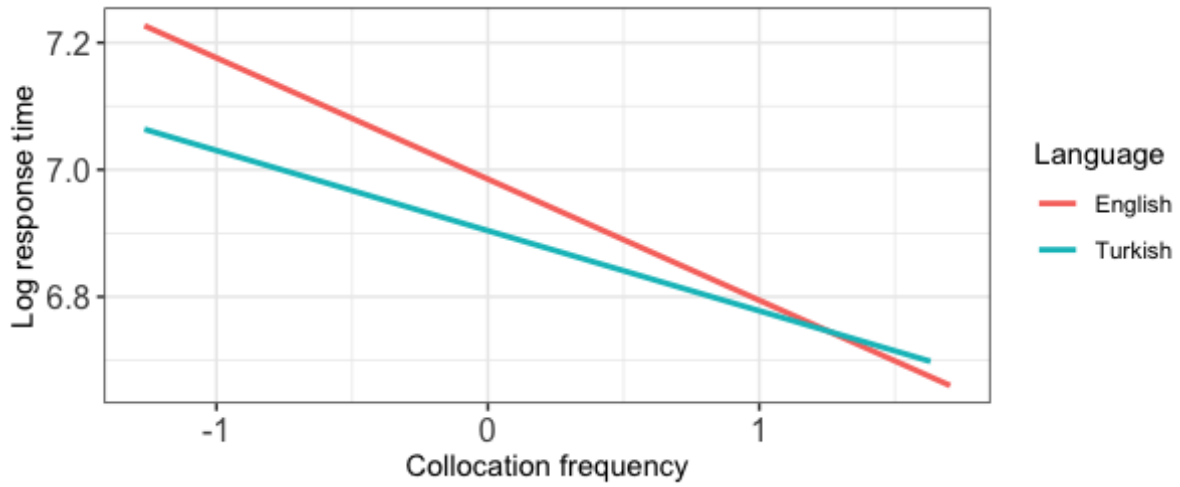
Fixed effects	***b***	***SE***	**95% CI**	***t***	***Df***	***p***
Intercept	6.955	0.019	[6.917, 6.993]	360.893	92.130.	<.001
English vs Turkish	0.083	0.038	[0.007, 0.159]	2.159	93.924	.033
Collocation frequency	-0.165	0.009	[-0.184, -0.146]	-11.019	172.928	<.001
Adjective frequency	0.016	0.009	[0.002, 0.030]	2.372	217.339	.018
Noun frequency	0.014	0.007	[-0.000, 0.028]	1.982	200.973	.004
Length (Number of letters)	0.031	0.006	[0.017, 0.044]	4.576	218.787	<.001
English vs Turkish x Collocation frequency	-0.065	0.019	[-0.103, -0.027]	-2.767	171.047	<.001
Collocation frequency x Noun frequency	-0.003	0.006	[-0.016, 0.009]	-0.011	190.824	0.991

Note. R^2 marginal = .168; R^2 conditional =.410

We conducted simple slopes analysis for collocation frequency counts by each language (English vs Turkish), using the emtrends function within the emmeans package in R (Lenth, 2018), to analyze the extent to which collocation frequencies affect English and Turkish speakers' response times differently. There was a significant interaction between language and collocation frequencies (English: -0.198, Turkish: -0.133), Estimate = -0.065, *SE* = 0.019, *z* = -3.370, $p < .001$, which suggests that, as the collocation frequencies increased, both language speakers' response times became faster. Figure 3 illustrates this interaction effect. The effect of collocation frequencies on English speakers' response times was stronger than that of Turkish speakers. Specifically, one unit of increase in form collocation frequency counts resulted in a -0.198 faster log response time measure for English speakers, corresponding to 31.571 milliseconds, whilst one unit of increase in form collocation frequency counts resulted in a -0.144 faster log response time measure, 22.961 milliseconds for Turkish speakers. The interaction between collocation and noun frequencies was not significant.

Figure 3.

Interaction between language (English vs Turkish) and collocation frequency counts



Additionally, we tested whether there is a significant difference between English and Turkish speakers' sensitivity to inflectional family frequencies. We used the same model including the fixed and random effects of Model 3 but replaced the form with inflectional family collocation frequencies. We observed the same pattern of results as we did in the first model: collocation (inflectional family) frequencies led to significantly faster response times for the speakers of both languages. There was a significant interaction between language and inflectional family collocation frequencies, showing that the effect of collocation frequencies (inflectional family) on English speakers' response times was stronger than that of Turkish speakers[6].

We constructed Model 4 (i.e., addressing Research Question 3) to investigate the extent to which English and Turkish speakers' sensitivity to word-level frequency information differ when processing high- and low-frequency collocations. As with Model 3, Model 4 had subject and item as random effects. We also included by-subject random slopes for adjective, noun frequencies and item type. As fixed effects, Model 4 had language (English vs Turkish), item type (high-frequency, low-frequency, and baseline), individual word frequencies for adjectives and nouns (both form and inflectional family, length of items (number of letters), and nouns' order of occurrence. We also included all possible interaction terms including the variables of interest: language and individual word frequencies for adjectives and nouns. In this model, we treated phrasal frequencies as a categorical variable (high-, low-frequency, baseline). Identical to Model 3, we followed a stepwise model comparison procedure to identify the best fit for Model 4.

[6] Moreover, as an additional analysis, we run the same model excluding the baseline items with 0 frequencies. In the acceptability judgment task, baseline items required a no response and collocational items required a yes response. Because different mechanisms might impact the processing of collocational and baseline items, we examined whether the findings were similar if baseline items are excluded from the analysis. We observed the same pattern of results (see Appendix E for the tables of these additional analysis).

As shown in Table 7, Model 4 included item type, language, adjective and noun frequencies (form), length (number of letters), and the interaction terms 'item type' and 'noun frequencies' as fixed-effects. The final model also included a random intercept for subjects (participants) and items, and also by-subject random slopes for the item type. The fixed-effect nouns' order of occurrence and the interaction terms language and individual word frequencies for adjectives and nouns, item type and adjective frequencies did not enter into the final model.

There was no main effect of language (English vs Turkish) on response times, controlling for length, individual word and phrasal frequencies. Unlike Model 3, we found no significant differences between English and Turkish speaker groups, either in terms of overall mean response times or language by item type interactions. As we predicted, response times were affected by item type; it took a shorter time for the speakers of both languages to respond to the high- and low-frequency collocations than the baseline items. They also responded to the high-frequency collocations faster than the low-frequency items, b = -0.188, SE = 0.025, 95% CI [-0.239, -0.137], t(169.28) = -7.30, $p < .001$, indicating that native-speakers of both languages are sensitive to the collocation frequencies. Unsurprisingly, participants needed longer to respond to items with more letters. Noun frequency counts led to significantly faster response times for only the low-frequency collocations. We obtained the simple slopes for noun frequencies by each level of item type to analyze the interaction between noun frequencies and item type. There was a significant interaction between item type and noun frequencies, indicating that as the noun frequencies increased, participants' responses for high- and low-frequency collocations became faster.

Table 7

Mixed effects Model 4 comparing English and Turkish speakers' response times for high-frequency, low-frequency, and baseline items

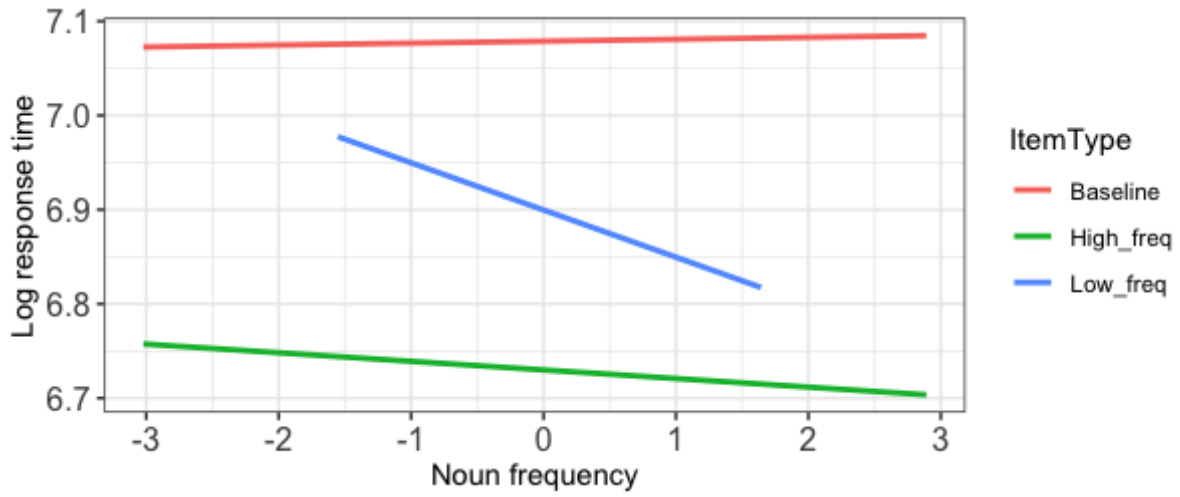
Fixed effects	***b***	***SE***	**95% CI**	***t***	***df***	***p***
Intercept	7.096	0.022	[7.051, 7.141]	308.905	95.305	<.001
English vs. Turkish	0.034	0.036	[-0.037, 0.107]	0.948	96.715	.345
Item Type: High frequency vs. Baseline	-0.365	0.019	[-0.403, -0.326]	-18.550	162.226	<.001
Item Type: Low frequency vs. Baseline	-0.183	0.019	[-0.221, -0.144]	-9.397	163.450	<.001
Adjective Frequency	0.011	0.006	[-0.001, 0.024]	1.737	219.872	.08
Noun frequencies	0.005	0.009	[-0.012, 0.024]	0.634	209.905	.526
Length (Number of letters)	0.031	0.006	[0.018, 0.043]	4.910	223.877	<.001
High frequency vs Baseline x Noun frequency	-0.012	0.014	[-0.040, 0.014]	-0.905	197.354	.366
Low frequency vs Baseline x Noun frequency	-0.06-	0.020	[-0.100, -0.020]	-2.99	219.593	.003

Note. R^2 marginal = .182; R^2 conditional = .412

We conducted simple slopes analyses to interpret the interactions between the effects of noun frequencies and item type, using the emtrends function within the emmeans package in R (Lenth, 2018). There was a significant interaction between noun frequencies and item type (High-frequency: 0.006, Low-frequency: 0.054, Baseline:0.005), Estimate = 0.060, *SE* = 0.020, *z* = -2.992, *p* = .007, indicating that both speaker groups' sensitivity to noun frequencies varied depending on the frequency of the collocations. Figure 4 illustrates this interaction effect. The effect of noun frequencies on the participants' response times for the low-frequency collocations was stronger than for the high-frequency collocations, when compared to the baseline items. That is to say, one unit of increase in noun frequency counts resulted in a -0.054 faster log response time measure for the low-frequency collocations, corresponding to 8.617 milliseconds, whilst one unit of increase in noun frequency counts resulted in a -0.006 faster log response time measure, 0.957 milliseconds, for the high-frequency collocations.

Figure 4.

Interaction between Item type (High-, low-frequency, baseline) and Noun frequency counts



We constructed Model 5 (i.e., addressing Research Question 4) to investigate potential differences between English and Turkish speakers' sensitivity to collocation strength, as measured by log Dice scores. There was a multicollinearity issue between collocation frequency, item type and log Dice scores so that we decided not to include collocation frequency and item type in this model. We had anticipated this, as log Dice scores highlights exclusive word pairs whose component words are not necessarily infrequent. As with Models 3 and 4, Model 5 had subject and item as random effects. We also included by-subject random slopes for adjective and noun frequencies and log Dice scores. As fixed effects, we had language (English vs Turkish), log Dice scores, and individual word frequencies for adjectives and nouns (both form and inflectional family). We also included all possible interaction terms including the variables of interest: language and log Dice scores, language and individual word frequencies for adjectives and nouns length of items (number of letters), and nouns' order of occurrence. Identical to the Model 3 and Model 4, we followed a stepwise model comparison procedure to identify the best model fit.

Model 5 included (as shown in Table 8) log Dice scores, language (English vs Turkish), adjective and noun frequencies (form), length (number of letters), and the interaction terms between language and noun frequencies (form) and log Dice scores and adjective frequencies as fixed effects. The final model also included a random intercept for subjects (participants) and items, and also by-subject random slopes for the log Dice scores. The fixed-effect nouns' order of occurrence, noun frequencies (inflectional family) and the interaction terms between language and log Dice scores, and language and adjective frequencies did not enter into the final model.

As shown in Table 8, there was no main effect of language (English vs. Turkish) on response times, controlling for length, or individual word frequencies. As we had predicted, response times were affected by collocational strength; it took less time for the speakers of both

languages to respond to the collocations with higher log Dice scores than the ones with lower log Dice scores, indicating that native-speakers of both languages are sensitive to the exclusivity of adjective-noun pairs. However, these collocation strength effects should be interpreted with caution: since we cannot include both collocation frequency counts and log Dice scores in the same model due to the multicollinearity, we cannot completely disentangle the effects of frequency and exclusivity. Noun frequency counts appear to lead to faster response times, but the effect was not significant. Adjective frequencies led to slower response times (see also Model 3).

Table 8.

Mixed effects Model 5 investigating English and Turkish speakers' sensitivity to collocation strength

Fixed effects	***b***	***SE***	**95% CI**	***t***	***df***	***p***
Intercept	7.073	0.021	[7.030, 7.115]	323.43	91.140	<.001
English vs. Turkish	0.017	0.037	[-0.055, 0.091]	0.471	95.345	.638
Collocation strength (Log Dice)	-0.043	0.002	[-0.047, -0.038]	-18.370	159.996	<.001
Adjective Frequency	0.015	0.006	[0.002. 0.028]	2.382	215.460	.018
Noun frequencies	-0.011	0.006	[-0.024, 0.001]	-1.696	202.089	.091
Length (Number of letters)	0.030	0.006	[0.017, 0.043]	4.712	217.936	<.001
English vs Turkish x Noun frequency	-0.024	0.013	[-0.050, 0.001]	-1.843	2101.439	.068

Discussion

In summary, we built three mixed-effects linear regression models to analyze the participants' response times in Study 2. In Models 3 and 4, we treated collocation frequencies as continuous (i.e., based on raw frequency), and categorical predictors (i.e., high-frequency vs. low-frequency) respectively. In Model 5, we analyzed the effect of collocation strength, also known as strength of association (Gablasova et al. 2017). We used the log Dice measure, which combines information about frequency and exclusivity. The results of Model 3 and Model 4 showed that English and Turkish speaker groups' processing was affected by collocation frequencies since both language groups needed a shorter time to process collocations that occur more frequently. This is in line with a number of previous studies that have found a strong effect of phrasal frequency on the processing of multiword sequences (e.g. Arnon & Snider, 2010; Kerz et al., 2019; Vilkaite, 2016). Model 3 showed that English speakers were more sensitive to the collocation frequencies than Turkish speakers. We found no significant differences between English and Turkish speakers' sensitivity to individual word frequencies for adjectives and nouns. For both groups, the higher adjective frequency counts led to slower response times (see Model 3). Unlike adjective frequencies, the higher noun frequency counts led to faster response times for low-frequency collocations (see Model 4 and Figure 4 for a visual illustration of the interaction effect). The results of mixed-effects Model 5 showed that speakers of both languages are similarly sensitive to collocation strength.

General Discussion

Prior research on multiword units has demonstrated compelling frequency effects of multiword sequences affecting language acquisition (e.g., Arnon & Clark, 2011; Skarabela et al., 2021), as well as language processing (e.g. Arnon & Snider, 2010; Kerz et al., 2019) and production (e.g. Janssen & Barber, 2012). Importantly, multiple distributional statistics of multiword units (e.g., individual word level and phrasal frequencies) affect their processing and production latencies for adult speakers (see, e.g., Öksüz et al., 2021 for processing; Shao et al., 2019 for production). In this study, we examined whether these effects are limited to languages with certain typological characteristics, like Indo-European languages, or whether they also apply to morphologically more complex languages. Unlike previous studies which have also focused on the processing of collocations in morphologically complex languages like Turkish (Cangir et al., 2017) and Lithuanian (Vilkaite-Lozdiene, 2022), we directly compared the processing of collocations in a morphologically isolating language like English and in an agglutinating language like Turkish. We examined the effects of individual word frequencies, collocation frequencies, and association strength in both languages. We first conducted corpus analyses to compare frequencies and association strength of adjective-noun collocations in English and Turkish. We then used acceptability judgment tasks to examine processing of the collocations in the two languages.

The results of Study 1, addressing Research Question 1, show that English collocations have higher form frequencies than Turkish ones. Nevertheless, Turkish collocations have significantly higher inflectional family frequencies than their English translation equivalents. This shows that Turkish speakers have significantly more experience with inflected forms of collocations than English speakers. English collocations are significantly more strongly associated than Turkish collocations. The results of the acceptability judgment tasks show that

both language speakers responded to the high-frequency collocations faster than to the low-frequency collocations (as can be seen in both Model 3 and Model 4). Model 4 showed that there was no significant difference in English and Turkish speakers' response times to either high- or low-frequency collocations. Both English and Turkish speakers' processing were significantly affected by the individual word frequencies of adjectives and nouns alongside collocation frequencies. Model 5 showed that speakers of both languages are sensitive to the collocation strength. These findings demonstrate that speakers are sensitive to multiple distributional statistics.

The corpus analysis of Study 1 shows that (a) English speakers have higher overall exposure to adjective-noun collocations than Turkish speakers, considering English collocations' slightly higher form frequencies, and (b) Turkish collocations have slightly higher inflectional family frequencies than English collocations. The acceptability judgment data of Study 2 show that both form and inflectional family collocation frequencies appear to similarly predict the processing latencies for speakers of both languages. This is because speakers' experience with different morphological forms of collocations, as reflected in inflectional family frequencies, contribute towards their overall experience with that collocation (see also Vilkaite-Lozdiene, 2022). This is because different morphological forms of a collocation should be seen as variations of that same collocation rather than collocations in their own right, with their own individual phrasal frequencies and associations. This is in line with Vilkaite-Lozdiene's (2022) recent research on Lithuanian collocations, which found a facilitation effect for collocations in different morphological forms, including very low-frequency ones, suggesting that speakers' overall experience with a collocation plays an important role in their processing.

With regard to the processing of adjective-noun collocations in typologically different languages, high-frequency multiword units show a processing advantage over low-frequency

ones in both languages. In other words, speakers are sensitive to the frequency with which multiword units occur in their respective languages, irrespective of the typology (see also Arnon & Snider, 2010; Siyanova-Chanturia et al., 2011). Crucially, this finding provides evidence that the processing of multiword units in a morphologically complex language like Turkish is consistent with the predictions of usage-based accounts (e.g. Arnon et al., 2017; Christiansen & Chater, 2016; Elman, 2009). Our findings further contribute to the mounting psycholinguistic evidence that speakers are sensitive to the statistical properties of multiword units across languages. Prior research found similar results for English (e.g., Arnon & Snider, 2010; Kerz et al., 2019), Spanish (Pulido, 2021), Dutch (Shao et al., 2019), Lithuanian (Vilkaite-Lozdiene, 2022), Turkish (Cangir et al., 2017) and Chinese (Yi et al., 2017). We provide further evidence that speakers of typologically different languages are similarly sensitivity to the statistical properties of multiword units.

Regarding the effects of individual word and collocation frequencies, we found that speakers of both languages are sensitive to both individual word and collocation frequencies. As in the case of collocation frequency effects (i.e., including form and inflectional family frequencies), they lead to faster response times. This is largely because the more often phrases are encountered, the more strongly they are encoded in the lexicon (Arnon & Snider, 2010; Carrol & Conklin, 2020; Hérnandez et al., 2016; Vilkaite-Lozdiene, 2022). We note that this can be interpreted in different ways, from an entirely "holistic" representation of phrases (especially some types of multiword units like non-decomposable idioms, e.g. *kick the bucket*) (see Swinney & Cutler, 1979) to something more akin to a lexical priming mechanism, whereby links between co-occurring words become stronger through experience (see Baayen, et al., 2013; Cangir et al., 2017; Hoey, 2005). Importantly the adjective-noun collocations used in this study are largely compositional, i.e. both adjectives and nouns contribute to the meaning of the phrase (e.g., *front door*). Nevertheless, the individual contributions of adjectives and

nouns to the fluency of collocational processing are different. Adjective frequencies lead to slower processing of collocations (see Model 3), whereas noun frequencies lead to faster processing of collocations, particularly for the low-frequency collocations (see Model 4). As demonstrated in Figure 4, the effect of noun frequencies on the participants' processing of low-frequency collocations was stronger than that of the high-frequency collocations. These findings highlight that collocational processing calls on knowledge of the component words since they can either facilitate or hinder the processing of the entire collocation (see also Jacobs et al., 2016).

Our results do not fit well with processing models that suggest multiword units are represented entirely holistically (e.g., Janssen & Barber, 2012; Wray, 2008). We propose that processing multiword units involves combining information about individual words and phrases. This is consistent with the large body of psycholinguistic evidence suggesting that language users predict upcoming words (see Baayen et al., 2013; Jacobs et al., 2016; Onnis & Huettig, 2021; Smith & Levy 2013). A key feature of predicting upcoming words is that word-to-word statistical information narrows down the possibilities of the word ahead. That is, information from one word enables predictions about properties of upcoming words. In the present study, we found that higher adjective frequencies lead to slower response times for the speakers of both languages. Specifically, encountering an adjective with a low-predictive value (i.e. a very frequent one like *long* or *good*) makes predicting the upcoming words relatively more difficult. From a corpus perspective, high frequency adjectives tend to form collocations with a large number of nouns, but those collocations are not likely to be exclusive. The exclusivity of collocations refers to the extent to which two words appear in each other's vicinity. From a language processing perspective, exclusivity is strongly linked to the predictability of upcoming words; when seeing one part of a collocation brings to mind the other. Our findings provide further evidence that speakers of both languages are sensitive to

the strength of collocations as measured by log Dice scores, as they responded to the frequent and exclusive collocations faster than the less frequent and less exclusive ones. Carrol and Conklin (2020) observed that the expectations created by seeing the first word are important explanatory variables for collocational processing.

Consistent with the findings of recent studies (e.g., Jacobs et al., 2016; Shao et al., 2019; Öksüz et al., 2021), we found a robust effect of noun frequencies for processing adjective-noun collocations in both languages. Higher noun frequencies led to faster processing of low-frequency collocations. The results of Model 4, addressing Research Question 2 and 3, indicate that the effect of noun frequencies on both speaker groups' processing of low-frequency collocations was stronger than their processing of high-frequency ones. This shows that for the high-frequency collocations, the effect of noun frequencies on the participants' processing decreased. Similarly, Arnon and Cohen Priva, (2014), focusing on native English speakers' phonetic duration in spontaneous speech, found that the effect of phrasal frequencies increased with repeated use, and the effect of individual word-level frequencies on the production of high-frequency multiword units decreased. This does not mean that frequently co-occurring multiword units are stored and retrieved as unanalyzed holistic units lacking internal analysis. Instead, frequently co-occurring multiword units result in the growing prominence of the units relative to the parts, yet information related to the parts is still accessible (Arnon & Cohen Priva, 2014; Öksüz et al., 2021). Thus, language users process phrases by retrieving and combining individual words in a sequential manner. It is plausible that frequent phrases are easier to process than less frequent ones because the selection of specific word combinations and their ordering becomes easier with practice.

We found that English speakers are more sensitive to phrasal frequencies than Turkish speakers when processing adjective-noun collocations in their respective languages. This suggests that language typology at least partly impacts speakers' sensitivity to frequency

information. Possibly Turkish speakers are less sensitive to these phrasal frequencies because the language's agglutinating morphology results in many more individual words, and fewer high-frequency multiword units in general (Durrant 2013). As demonstrated in Figure 3, the effect of collocation frequencies on English speakers' response times was stronger than that of Turkish speakers. So, Turkish speakers may be less experienced than English speakers in processing high-frequency multiword units, including adjective-noun collocations. We note that this does not indicate that Turkish speakers would necessarily be less prone to multiword chunking. Arguably, chunking enables the speakers of agglutinating languages to reduce the computational burden imposed by highly complex morphology and long-distance dependencies – this is a speculation that can be tested in future research.

We predicted that Turkish speakers would be more sensitive to the individual word frequencies than English speakers, because of the differences in general processing biases between the two languages. Therefore, it is surprising that we found no evidence of a difference between English and Turkish speakers' sensitivity to individual word frequencies for either adjectives or nouns. In languages like English, many grammatical relations hold between words, and this translates into multiword units (see also Arnon, 2021). In languages like Turkish however, many of the grammatical relations are contained within a single word and thus relevant larger units are likely to be a single word involving more than one morpheme. Given that the frequencies of phrases and individual words are lower in Turkish than English (Durrant, 2013), our experimental materials, even the low-frequency ones, may be sitting towards the high end of the scale. Thus, the range and the mean frequency of individual words could have been higher in the task than those that Turkish speakers are exposed to in real life texts, indicating a more reduced sensitivity to phrasal frequency than may in fact be the case.

Our findings have important theoretical implications. This is the first study, to our knowledge, that contrastively examines the processing of multiword units in two typologically

different languages. Our results suggest that speakers of both languages are sensitive to the distributional statistics while they process multi-word units. Critically, this is consistent with the predictions of usage-based accounts (Christiansen & Chater, 2016; Ellis, 2002; Tomasello, 2003), providing further support for the role of multiword units as building blocks in both languages (see also Arnon et al., 2017; Jolsvai et al., 2020). Our results further highlight the parallels between words and phrases and challenge the long-held views that words and larger sequences are processed by qualitatively different systems, also known as dual-route processing models (Pinker & Ullman, 2002). Instead, our findings accord well with a single-system view of language (e.g., Baayen et al., 2013; Bybee, 1998; Christiansen & Chater, 2016; Ellis, 2002) where all linguistic experience is processed by similar mechanisms. The current study extends the previous findings on the role of multiword units in processing (Hérnandez, et al., 2016; Jolsvai et al., 2020; Kerz et al., 2019; Öksüz et al., 2021) and highlights the importance of multiword units in comprehension and production models from a typological perspective.

The results also have methodological implications. We address the issue of the lack of linguistic diversity in language acquisition and psycholinguistic research, providing a systematic comparison of the processing of multiword units in two typologically different languages. Importantly, contrasting typologically unrelated languages prompts a re-evaluation of established theoretical positions (Christiansen et al., 2022). Here we tested the predictions of usage-based accounts for processing multiword units in English and Turkish, typologically unrelated languages with regard to their morphological structures. We adopted a multi-method approach, combining corpus and experimental methods. The corpus analysis provided insight into English and Turkish speakers' experience with adjective-noun collocations in their respective languages. Using the acceptability judgment tasks, we evaluated English and Turkish speakers' sensitivity to distributional statistics. In this study, we showed an example

of how the combination of psycholinguistic experiments and corpus analysis offer an ideal toolset in examining language processing.

Limitations and future directions

There are a few limitations in this study. First, the acceptability judgment task used is not the most suitable task for examining the automatic processing of multiword units because this type of judgment task is likely to require participants to consciously reflect on the phrases. Thus, the responses may indicate metalinguistic-based processing rather than automatic (subsconscious) processing (e.g., Rebuschat, 2022; Williams & Rebuschat, 2022). Second, since we specifically examined frequencies of the meaning-equivalent phrases in the two languages, the findings of the corpus analysis only included congruent adjective-noun combinations that have translation-equivalents in both languages. This limited the scope of the analysis to congruent collocations only, and thus frequency comparisons did not provide a full picture of frequencies of adjective-noun collocations in the two languages. Third, it is questionable whether the BNC and TNC are fully comparable, as the size of the BNC is considerably larger than that of the TNC. The BNC was previously used in many studies (e.g. Jacobs et al., 2016; Siyanova-Chanturia et al., 2011) with similar scopes for extracting individual word and phrasal frequencies; however, this is the first study using the TNC for that purpose. Given that it is designed to be a balanced, large-scale corpus of modern Turkish and it follows the framework of the BNC, we decided to use it. Fourth, even though the corpus study showed a difference between the two languages in inflectional family frequencies rather than form frequencies, we selected our experimental materials based on their form frequencies. We cannot be sure whether we would have had a similar pattern of findings if we had selected our items based on inflectional family frequencies. However, we ran Model 3 with inflectional family frequencies to test whether speakers were similarly sensitive to both measures of

frequency in their respective native languages. Importantly both models showed the same pattern of findings: that the effect of collocation frequencies (inflectional family) on English speakers' response times was stronger than that of Turkish speakers.

Another limitation is that our experimental materials, particularly those in the low-frequency band, might be considered as commonly used phrases. That is, we might have missed a certain level of variation in collocation and individual word frequencies for the items we selected for the acceptability judgment task, particularly Turkish ones, due to the smaller size of the TNC. Inevitably, this might have affected speakers' sensitivity to the individual word and collocation frequencies. Finally, we included collocation strength in our analysis, but relying on the findings of Model 3, we cannot clearly disentangle the effects of collocation frequency from exclusivity. This is because we did not select another group of collocations controlling the range of their frequencies, and future studies could do this to further disentangle these influences.

There are various future directions for future research to gain a more complete picture of the comprehension and production of multiword units. First, future research should collect more evidence from comparisons between typologically different languages regarding comprehension and production of multiword units. In the case of agglutinating languages, including phrases and component words of phrases from a wider range of frequencies than we did in this study would play a key role in identifying the potential interaction between language users' sensitivity to individual word-level and phrasal frequencies. Second, although we provided evidence of differences between the speaker groups, there is variance of performance within both language groups. Although this is common for processing tasks since individual differences are a pervasive feature of language acquisition and processing (Kidd, Donnelly & Christiansen, 2018), future research should identify the specific cognitive individual differences impacting processing of multiword units. In this study, we solely analyzed the

impacts of individual word-level and phrasal frequencies on the processing of multiword units. However, there are other linguistic properties which could affect their processing such as semantic relations between component words of multiword units. In the future, researchers should evaluate this critically.

Regarding processing tasks to examine collocational processing, it is important to use tasks that make it possible to examine the degree of automaticity in processing of multiword units. One alternative to acceptability judgement tasks is to use a priming task in which collocating words are used to prime each other. For example, Cangir et al. (2017) used a priming task, with 100 milliseconds of stimulus-onset-asynchrony. They showed that there is a priming relationship between component words of Turkish adjective-noun and noun-verb collocations. Future research should examine whether a processing advantage for collocations is maintained if prime words are presented for a shorter duration than 100 milliseconds, and how word-to-word statistical information, including individual word and phrasal frequencies and collocational strength interact with the choice of stimulus-onset-asynchrony. Finally, Durrant (2013) identified multimorphemic patterns within morphologically complex Turkish words; future research should examine how these patterns facilitate morphological processing.

It is also crucial to acknowledge the importance of combining corpus and experimental approaches to investigate the comprehension and production of multiword units (Rebuschat, et al., 2017;) For example, future experimental research should investigate speakers' sensitivity to various distributional information including frequency, dispersion, and exclusivity, as corpus linguistics provides researchers with measures of collocation strength that capture different dimensions of collocational relationships such as directionality (delta P), dispersion (Cohen's *d*), and exclusivity (log Dice and Mutual Information).

Conclusion

In summary, we have contrastively examined the effects of individual word and phrasal frequencies for processing adjective-noun collocations in English and Turkish. We found that the typology of language impacts language users' sensitivity to phrasal frequencies. Our findings align well with a single-system view of language processing (e.g. Baayen et al. 2013; Bybee, 1998; Christiansen & Chater, 2016; Onnis & Huettig, 2021) that combines information about individual words and larger units to explain language acquisition and processing. The findings are hard to align with proposals that multiword units are represented holistically (e.g. Wray, 2002). Methodologically, combining psycholinguistic experimentation with corpus analyses constitutes an ideal toolset for research into processing and learning of multiword units. Corpora provide direct information about the multiword units' usage patterns; the frequencies and association measures that capture various collocational properties. Experiments make it possible to test language users' sensitivity to these distributional properties.

Acknowledgements

We would like to thank Morten Christiansen and Aina Casaponsa for helpful discussion, as well as Leyla Marti, Belma Haznedar and Yasemin Bayyurt for their support with data collection. While this manuscript was written the final author was supported by the the LEAD Graduate School & Research Network, a project of the Excellence Initiative of the German federal and state governments.

Declaration of Competing interest

We have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

Aksan, Y., Aksan, M., Koltuksuz., A, Sezer, T., Mersinli, Ü., Demirhan, U., … Kurtoglu Ö. (2012). Construction of the Turkish National Corpus (TNC). *Proceedings of the 8th International Conference on Language Resources and Evaluation* (pp. 3223-3226). Istanbul, Turkey.

Allen, S. E. M. (2017). Polysynthesis in the acquisition of Inuit languages. In M. Fortescue, M. Mithun, & N. Evans (eds.), *Handbook of polysynthesis*. Oxford: Oxford University Press.

Arnon, I. (2021). The Starting Big approach to language learning. *Journal of Child Language 48*(5), 937-958. https://doi.org/10.1017/S0305000921000386

Arnon I., & Clark E, V. (2011). Why brush your teeth is better than – Children's word production is facilitated in familiar sentence – frames. *Language Learning and Development, 7*(2), 107-129. https://doi.org/10.1080/15475441.2010.505489

Arnon, I., & Christiansen, M. H. (2017). The Role of Multiword Building Blocks in Explaining L1–L2 Differences. *Topics in Cognitive Science, 9*(3), 621–636. https://doi.org/10.1111/tops.12271

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*(3), 377–400.

Arnon I., & Snider, N. (2010). More than words: Frequency effects for multi-word phrases. *Journal of Memory and Language, 62*(1), 67-82. https://doi.org/10.1016/j.jml.2009.09.005

Arnon, I., McCauley, S. M., & Christiansen, M. H. (2017). Digging up the building blocks of language: Age-of-acquisition affects for multiword phrases. *Journal of Memory and Language, 92,* 265-280.

Baayen, R. H., Hendrix, P., & Ramscar, M. (2013). Sidestepping the combinatorial explosion: An explanation of n-gram frequency effects based on naïve discriminative learning. *Language and Speech, 56,* 329-347.

Barr, D. J., Levy, R., Scheepers, C. & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language, 68*(3), 255-278. https://doi.org/10.1016/j.jml.2012.11.001

Bod, R. (2006). Exemplar-based syntax: How to get productivity from examples. *Linguistic Review, 23*(3), 291-300. https://doi.org/10.1515/TLR.2006.012

Biber, D. (2009). A corpus-driven approach to formulaic language in English. *International Journal of Corpus Linguistics*, *14*(3), 275-311. https://doi.org/10.1075/ijcl.14.3.08bib

Bien, H., Levelt, W. M. J. & Baayen, R. H. (2005). Frequency effects in compound production. *Proceedings of the National Academy of Sciences, 102*, 17876-17881. https://doi.org/10.1073/pnas.050843110

Brezina, V., McEnery, T., & Wattam, S. (2015). Collocations in context: a new perspective on collocation networks. *International Journal of Corpus Linguistics 20(2),* 139-173. https://doi.org/10.1075/ijcl.20.2.01bre

Bybee, J. (1998). The emergent lexicon. In *The 34th Chicago Linguistic Society* (pp. 421-435).

Chater, N., & Christiansen, M., H. (2018). Language acquisition as skill learning. *Current Opinion in Behavioural Science*, *21*, 205-208.

Chang, F., Dell G. S., & Bock, K. (2006). Becoming syntactic. *Psychological Review*, *113*(2), 234-272. https://psycnet.apa.org/doi/10.1037/0033-295X.113.2.234

Cangir, H., Büyükkantarcioglu, N. & P. Durrant (2017). Investigating collocational priming in Turkish. *Journal of Language and Linguistics Studies*, *13*(2), 465-486.

Carrol G., & Conklin K. (2020). Is all formulaic language created equal? Unpacking the processing advantage for different types of formulaic sequences. *Speech, 63*(1), https://doi.org/10.1177/0023830918823230

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/S0140525X1500031X

Christiansen, M., H., Contreras Kallens, P. & Trecca, F. (2022). Towards a comparative approach to language acquisition. *Current Directions in Psychological Science, 31*(2), 131-138. https://doi.org/10.1177/09637214211049229

Chomsky, N. (2000). *New horizons in the study of language and mind.* Cambridge University Press.

Courtney, E. H., & Saville-Troike, M. (2002). Learning to construct verbs in Navajo and Quechua. *Journal of Child Language, 29*(3), 623–654. https://doi.org/10.1017/S0305000902005160

Davies, M. (2004). *BYU-BNC* (Based on the British National Corpus from Oxford University Press). Available at http://corpus.byu.edu/bnc/

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: Addison, Wesley, Longman.

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

Elman, J. (2009). On the meaning of words and dinosaur bones: Lexical knowledge without a lexicon. *Cognitive Science, 33*(4), 547-582. https://doi.org/10.1111/j.1551-6709.2009.01023.x

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*(2), 143–188. https://doi.org/10.1017/S0272263102002024

Ellis, A, W., & Morrison, C. M. (1998). Real age-of-acquisition effects in lexical retrieval. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 24*(2), 515-523. https://doi.org/10.1037/0278-7393.24.2.515

Franck, J., Vigliocco, G., Antón-Méndez, Collina, S., Frauenfelder, U. (2008). The interplay of syntax and form in sentence production: A cross-linguistic study of form effects on agreement. Language and Cognitive Processes, *23*(3), 329-374 https://doi.org/10.1080/01690960701467993

Fox, J. & Weisberg, S. (2019). *An {R} Companion to Applied Regression, Third Edition.* Thousands Oaks CA. Sage

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

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

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

Hardie, A. (2012). CQPweb—Combining power, flexibility and usability in a corpus analysis tool. *International Journal of Corpus Linguistics, 17*(3), 380–409. https://doi.org/10.1075/ijcl.17.3.04har

Hernández, M., Costa, A., Arnon, I. (2016). More than words: Multiword frequency effects in non-native speakers. *Language, Cognition and Neuroscience, 31*(6), 785–800. https://doi.org/10.1080/23273798.2016.1152389

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

Jacobs, C. L. Dell, G. S., Benjamin, A. S., & Bannard, C. (2016). Part and whole linguistic experience affect recognition memory for multiword sequences. *Journal of Memory and Language, 87,* 38-58. https://doi.org/10.1016/j.jml.2015.11.001

Janssen, N., Barber, H. A. (2012). Phrase frequency effects in language production. *PloS ONE, 7*(3), e33202. https://doi.org/10.1371/journal.pone.0033202

Jolsvai, H., McCauley, S. M., & Christiansen, M. H. (2020). Meaningfulness beats frequency in multiword chunk processing, *Cognitive Science, 44*, e12885. https://doi.org/10.1111/cogs.12885

Kerz, E., Wiechmann, D., & Christiansen, M. H. (2019). Turning to multiple statistics: Second language processing of multiword sequences across registers. In A. Goel, C. Seifert & C. Freksa (Eds.), *Proceedings of the 41st Annual Conference of the Cognitive Science Society* (pp. 546-552). Austin TX: Cognitive Science Society

Kuznetsova A, Brockhoff PB, Christensen RHB (2017). lmerTest Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software, 82*(13), 1– 26. ” https://www.jstatsoft.org/article/view/v082i13/0”

Lenth, R. (2022). *Emmeans: Estimated marginal means, aka least-square means*. R package version 1.8.0

Lõo, K., Järviki, J. & Baayen, H. (2018). Whole-word frequency and inflectional paradigm size facilitate Estonian case-inflected noun processing. *Cognition, 175*, 20-25. https://doi.org/10.1016/j.cognition.2018.02.002

Mata, A., Percy, E. J., & Sherman, S. J. (2013). Adjective–noun order as representational structure: Native-language grammar influences perception of similarity and recognition memory. *Psychonomic Bulletin & Review, 2*, 193–197. https://doi.org/10.3758/s13423-013-0479-z

McCauley, S. M. & Christiansen, M. H. (2014). Acquiring formulaic language: A computational model. *Mental Lexicon, 9,* 419-436. https://doi.org/10.1075/ml.9.3.03mcc

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

McEnery, T., & Hardie, A. (2011). *Corpus linguistics: Method, theory and practice.* Cambridge, UK: Cambridge University Press.

Nagy, W., Anderson, R. C., Schommer, M., Scott, J. A., & Stallman, A. C. (1989). Morphological families in the internal lexicon. *Reading Research Quarterly, 24*(3), 262–282. https://doi.org/10.2307/747770

Onnis, L., & Huettig, F. (2021). Can prediction and retrodiction explain whether frequent multi-word phrases are accessed 'precompiled' from memory or compositionally constructed on the fly? *Brain Research* 147674 https://doi.org/10.1016/j.brainres.2021.147674

Öksüz, D. C. (2023, June 26). Collocational processing in typologically different languages, English and Turkish. Retrieved from osf.io/muwjz

Öksüz, D. C., Brezina, V., Rebuschat, P. (2021). Collocational processing in L1 and L2: The effects of word frequency, collocational frequency, and association. *Language Learning*, *71*(1), 55-98 https://doi.org/10.1111/lang.12427

Pinker, S. & Ullman, M. (2002). The past and future of the past tense. *Trends in Cognitive Science*, 6(11), 456-463. https://doi.org/10.1016/S1364-6613(02)01990-3

Peirce, J. W., Gray, J. R., Simpson, S., MacAskill, M. R., Höchenberger, R., Sogo, H., Kastman, E., & Lindeløv, J. (2019). PsychoPy2: Experiments in behaviour made easy. *Behaviour Research Methods*, *51*(1), 195-203. https://doi.org/10.3758/s13428-018-01193-y

Pullido, M. (2021). Individual chunking ability predicts efficient or shallow processing: Eye-tracking evidence from multiword units in relative clauses. *Frontiers in Psychology* 11:607621. doi:10.3389/fpsyg.2020.607621

R Core Team. (2022). *R: A language and environment for statistical computing. R Foundation for Statistical Computing*. Retrieved from: http://www.R-project.org

Rayson, P. & Garside, R. (1998). The CLAWS Web Tagger. *ICAME Journal*, *22*, 121-123. http://icame.uib.no/ij22/claws.pdf

Rebuschat, P. (2022). Implicit learning and language acquisition: Three approaches, one phenomenon. In A. S. Reber & R. Allen (Eds.) The cognitive unconscious: The first half-century. Oxford University Press.

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. https://doi.org/10.1111/lang.12243

Rebuschat, P., Monaghan, P., & Schoetensack, C. (2021). Learning vocabulary and grammar from cross-situational statistics. *Cognition*. 206, [104475] **https://doi.org/10.1016/j.cognition.2020.104475**

Powell, M. (2009). *The BOBYQA algorithm for bound constrained optimization without derivatives* [Unpublished manuscript]. Department of Applied Mathematics and Theoretical Physics, Centre for Mathematical Sciences, University of Cambridge.

Reali, F., & Christiansen, M H. (2007). Processing of relative clauses is made easier by frequency of occurrence. *Journal of Memory and Language, 57*(1), 1-23. https://doi.org/10.1016/J.JML.2006.08.014

Shao, Z., van Paridon, J., Poletiek, F., & Meyer, A. S. (2019). Effects of phrase and word frequencies in noun phrase production. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *45*(1), 147-165. https://doi.org/10.1037/xlm0000570

Skarabela, B., Ota, M., O'Connor, R., & Arnon, I. (2021). 'Clap your hands' or 'Take your hands'? One-year-olds distinguish between frequent and infrequent multiword phrases. *Cognition, 211,* 104612. https://doi.org/10.1016/j.cognition.2021.104612

Siegelman, N., & Arnon, I. (2015). The advantage of starting big: Learning from unsegmented input facilitates mastery of grammatical gender in an artificial language. *Journal of Memory and Language*, *85*, 60–75. https://doi.org/10.1016/j.jml.2015.07.003

Smith, N. J. & Levy, R. (2013). The effect of word predictability on reading time is logarithmic. *Cognition, 128,* 302-319. https://doi.org/10.1016/j.cognition.2013.02.013

Smith, E. E., & Osherson, D. N. (1984). Conceptual combination with prototype concepts. *Cognitive Science, 8,* 337–361. https://doi.org/10.1207/s15516709cog0804_2

Siyanova-Chanturia, A., Conklin, K., Cafarra, S., Kaan, E. & van Heuven W. J. (2017). Representation and processing of multiword expressions in the brain. *Brain and Language, 175*, 111-122 http://dx.doi.org/10.1016/j.bandl.2017.10.004

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

Stoll, S., Abbot-Smith, K., & Lieven, E. (2009). Lexically restricted utterances in Russian, German, and English child-directed speech. *Cognitive Science, 33*(1), 75–103. https://doi.org/10.1111/j.1551-6709. 2008.01004.x

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

Vilkaité, L. (2016). Are nonadjacent collocations processed faster? *Journal of Experimental Psychology: Learning Memory, and Cognition, 42*(10), 1632-1642. https://doi.org/10.1037/xlm0000259

Vilkaité-Lozdiené (2022). Do different morphological forms of collocations show comparable processing facilitation? *Journal of Experimental Psychology: Learning Memory, and Cognition.* Advance Online Publication. https://doi.org/10.1037/xlm0001130

Wray, A. (2008). Formulaic language: pushing the boundaries. Oxford: *Oxford University Press.*

Williams, J., & Rebuschat, P. (2022). Implicit learning and SLA: A cognitive psychology perspective. In Godfroid, & H. Hopp (Eds,). *The Routledge Handbook of Second Language Acquisition and Psycholinguistics.* (The Routledge Handbooks in Second Language Acquisition). Routledge.

Winter, B. (2020). *Statistics for linguists: An Introduction using R.* New York, Taylor & Francis.

Winter, B., Perlman, M., & Majid, A. (2018). Vision dominates in perceptual language: English sensory vocabulary is optimized for usage. *Cognition, 179,* 213-220 https://doi.org/10.1016/j.cognition.2018.05.008

Wolter, B., & Yamashita, J. (2018). Word frequency, collocational frequency, L1 *Studies in Second Language Acquisition, 40(2),* https://doi.org/10.1017/S0272263117000237

Yi, W., Lu, S., & Ma, G. (2017). Frequency, contingency and online processing of multiword sequences: An eye-tracking study. *Second Language Research, 33,* 519–549. https://doi.org/10.1177/0267658317708009

Appendix A

Log Dice score calculated as follows:

$$Log\ Dice = 14 + log_2 \frac{2\ x\ O_{11}}{R_1 + C_1}$$

14 defines the theoretical maximum of the score from which a certain number is subtracted. Therefore, Log Dice has a theoretical maximum of 14. The binary logarithm will always be negative or zero because the ratio of $\frac{2\ x\ O_{11}}{R_1 + C_1}$ will be smaller than or equal to zero. It represents the $\frac{O_{11}}{R_1}$ and $\frac{O_{11}}{C_1}$.

Appendix B

A Full List of Collocations analyzed in Study 1

Table B1. High-frequency collocations

High-frequency Collocations (English)	High-frequency Collocations (Turkish)
Long time	Uzun zaman
Short time	Kisa zaman
Good time	Iyi zaman
Right time	Dogru zaman
Present time	Simdiki zaman
Spare time	Bos zaman
Extra time	Ekstra zaman
Sufficient time	Yeterli zaman
Bad time	Kötü zaman
Appropriate time	Uygun zaman
Other day	Diger gün
Following day	Ertesi gün
Previous day	Önceki gün
Good day	Iyi gün
Whole day	Bütün gün
Full day	Tam gün
Final day	Son gün
Bad day	Kötü gün
New day	Yeni gün
Special day	Özel gün
Long way	Uzun yol
Only way	Tek yol
Easy way	Kolay yol
Right way	Dogru yol
Short way	Kisa yol
Straightforward way	Düzgün yol
Safe way	Güvenli yol
Middle way	Orta yol
Suitable way	Uygun yol
Basic way	Temel yol
Young man	Genç adam
Old man	Yasli adam
Little man	Küçük adam
Big man	Büyük adam
Poor man	Fakir adam

Good man	Iyi adam
White man	Beyaz adam
Married man	Evli adam
Rich man	Zengin adam
Thin man	Zayif adam
Real world	Gerçek dünya
New world	Yeni dünya
Outside world	Dis dünya
Whole world	Bütün dünya
Modern world	Modern dünya
Old world	Eski dünya
Physical world	Fiziki dünya
Developing world	Gelisen dünya
Inner world	Iç dünya
Free world	Özgür dünya

Table B2. Mid-frequency collocations

Mid-frequency collocations (English)	Mid-frequency Collocations (Turkish)
Whole family	Bütün aile
Large family	Genis aile
Nuclear family	Çekirdek aile
Small family	Küçük aile
Happy family	Mutlu aile
Good family	Iyi aile
Traditional family	Geleneksel aile
Poor family	Fakir aile
Wealthy family	Zengin aile
Modern family	Modern aile
Whole country	Bütün ülke
Foreign country	Yabanci ülke
Developing country	Gelisen ülke
Poor country	Fakir ülke
Free country	Bagimsiz ülke
Different country	Farkli ülke
Independent country	Bagimsiz ülke
Rich country	Zengin ülke
Neighbouring country	Komsu ülke
Democratic country	Democratik ülke
Modern society	Modern toplum
Capitalist society	Kapitalist toplum
	Democratik toplum

Democratic society	Yeni toplum
New society	Kirsal toplum
Rural society	Özgür toplum
Free society	Sosyalist toplum
Socialist society	Burjuva toplum
Bourgeois Society	Feodal toplum
Feudal society	Sivil toplum
Divided society	Iyi sonuç
Good result	Olumlu sonuç
Positive result	Benzer sonuç
Similar result	Farkli sonuç
Different result	Önemli sonuç
Important result	Olumsuz sonuç
Negative result	Kötü sonuç
Bad result	Ilginç sonuç
Interesting result	Basarili sonuç
Successful result	Dogru sonuç
Correct result	Yeni sanayi
New industry	Yerel sanayi
Local industry	Agir sanayi
Heavy industry	Önemli sanayi
Important industry	Modern sanayi
Modern industry	Genis sanayi
Large industry	Ulusal sanayi
National industry	Hafif sanayi
Light industry	Geleneksel sanayi
Traditional industry	Otomotiv sanayi
Automative industry	

Table B3. Low-frequency collocations

Low-frequency collocations (English)	Low-frequency collocations (Turkish)
Public holiday	Resmi tatil
Long holiday	Uzun tatil
Popular holiday	Populer tatil
Short holiday	Kisa tatil
Cheap holiday	Ucuz tatil
Previous holiday	Önceki tatil
Nice holiday	Güzel tatil
Alternative holiday	Alternatif tatil
Luxury holiday	Lüks tatil
Religious holiday	Dini tatil

Tropical forest	Tropik orman
Natural forest	Dogal orman
Dense forest	Yogun orman
New forest	Yeni orman
Remaining forest	Kalan orman
Brown forest	Kahverengi orman
Old forest	Eski orman
High forest	Yüksek orman
Short distance	Kisa mesafe
Long distance	Uzun mesafe
Far distance	Uzak mesafe
Social distance	Sosyal mesafe
Little distance	Küçük mesafe
Equal distance	Esit mesafe
Correct distance	Dogru mesafe
Physical distance	Fiziki mesafe
Large distance	Büyük mesafe
Middle distance	Orta mesafe
New wood	Yeni tahta
Old wood	Eski tahta
Dry wood	Kuru tahta
Painted wood	Boyali tahta
White wood	Beyaz tahta
Fine wood	Güzel tahta
Thick wood	Kalin tahta
Rotten wood	Çürük tahta
White cloth	Beyaz tahta
Damp cloth	Islak tahta
Woollen cloth	Yünlü bez
Wowen cloth	Dokuma bez
Clean cloth	Temiz bez
Red cloth	Kirmizi bez
Dry cloth	Kuru bez
Thick cloth	Kalin bez

Appendix C

Table C1. Model 1 Table

Fixed effects	***b***	***SE***	***t***	***Df***	***p***
Intercept	1.712	0.208	8.206	97.899	<.001
Language: English vs Turkish	-0.123	0.056	-2.169	553.952	.030
Lemmatization Status: Form vs Inf. Family	0.182	0.057	3.163	555.314	.000
Adjective frequency	0.198	0.031	6.283	558.818	<.001
Language: Eng. vs Tur x Lemmatization Status: Form vs Inf. Family	0.394	0.080	4.895	553.945	<.001

Note. R^2 marginal = .110; R^2 conditional = .598

Figure C1

Interaction between Item type (High-, low-frequency, baseline) and Noun frequency counts

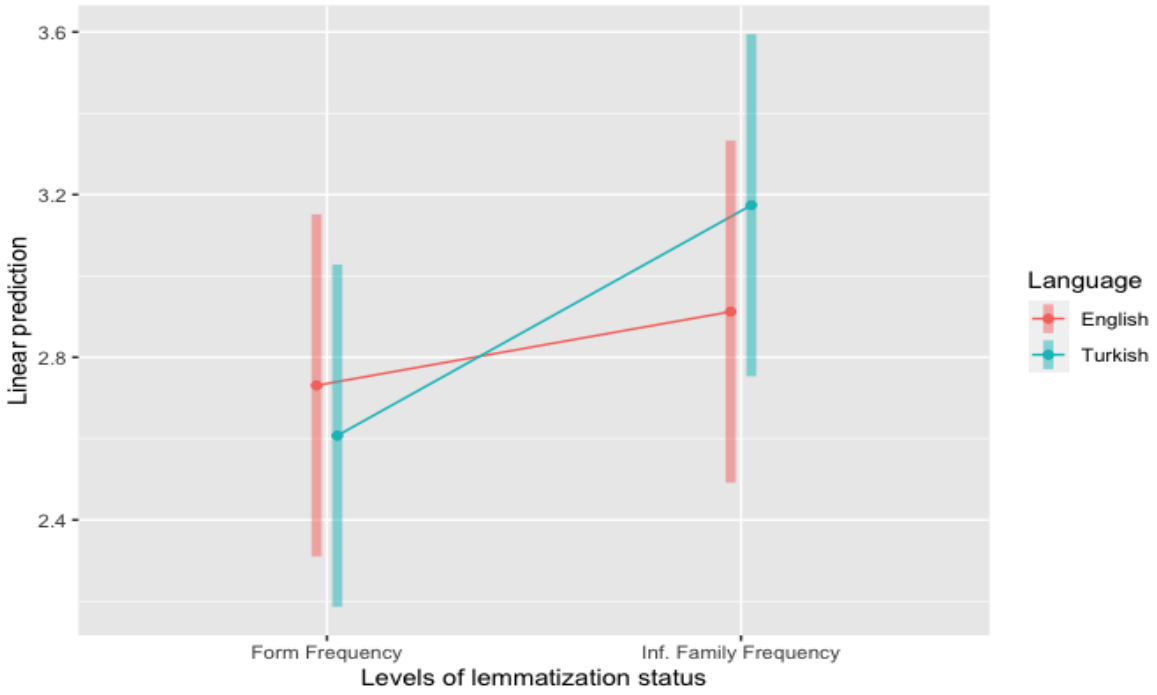


Table C2. Model 2 Table

Fixed effects	***b***	***SE***	***t***	***Df***	***p***
Intercept	6.694	0.653	10.250	369.575	<.001
Language: English vs Turkish	-0.529	0.210	-2.518	558.192	.012
Lemmatization Status: Form vs Inf. Family	0.383	0.212	1.802	562.506	0.072
Adjective frequency	-0.481	0.116	-4.141	572.003	<.001
Language: Eng. vs Tur x Lemmatization Status: Form vs Inf. Family	0.689	0.297	2.316	558.163	.020

Appendix D

Experimental Materials

Table D1 Selected nouns for extracting a large sample of adjective-noun to determine the frequency and log Dice cut-off scores

Noun nodes	Raw frequency counts	Relative frequency counts per million words
People	121,591	1,084.64
Government	61,798	551.26
Life	54,907	489.79
Children	45,641	407.14
Water	34,325	306.19
Power	31,560	281.53
Report	23,598	210.50
Paper	16,899	150.75
Material	13,315	118.78
Officer	8,961	79.94

To determine the frequency and log Dice cut-off scores for high- and low-frequency collocations, we selected 10 nouns from various raw frequency counts with a high score of 121,591 (*people*), and a low score 8,961 (*officer*). We paid a special attention to select the nouns from various frequency bands to be able to reliably establish threshold frequency counts. we extracted a pool of 4718 adjective-noun combinations and examined their frequency distributions.

Figure D1

Frequency distributions of the extracted adjective-noun combinations in the British National Corpus (2007).

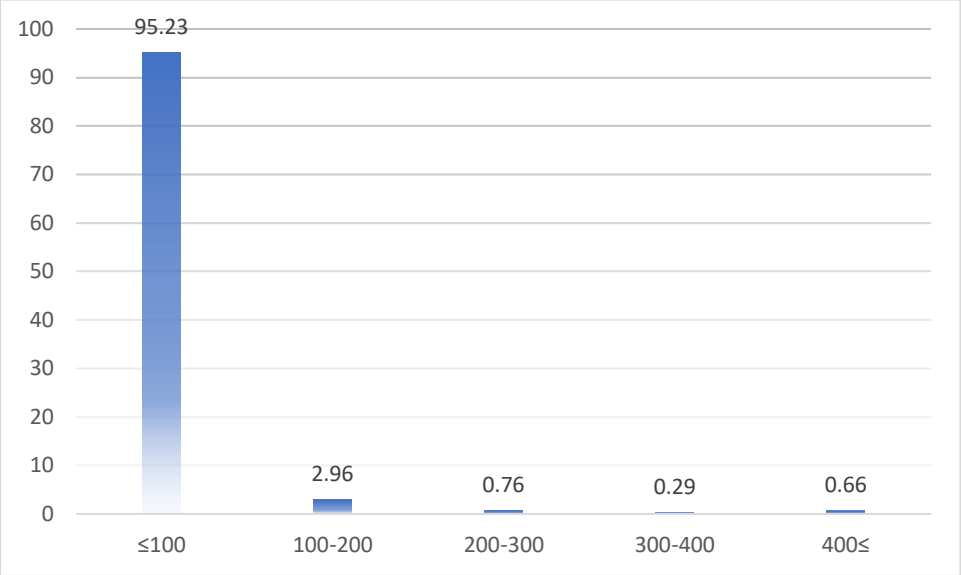
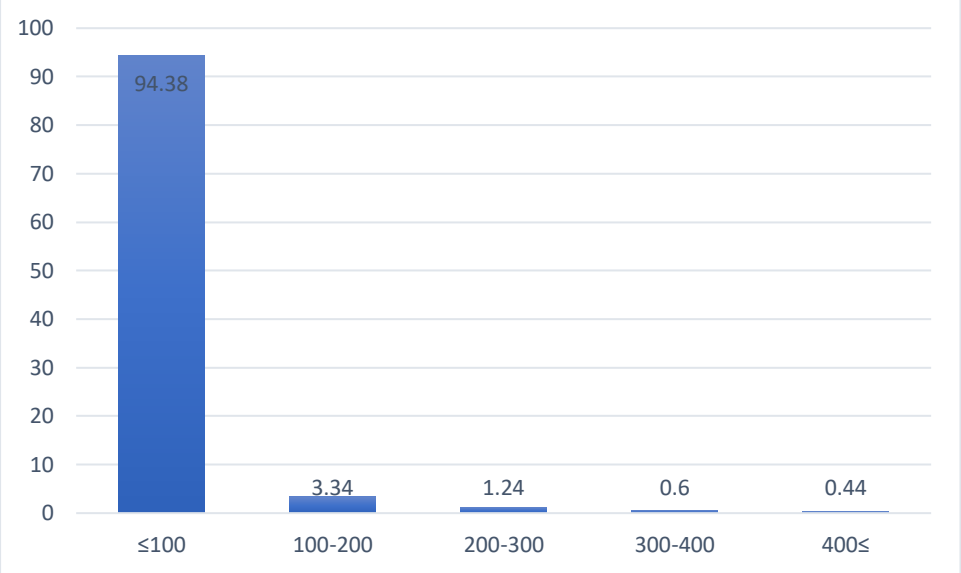


Figure D2

Frequency distributions of the extracted adjective-noun combinations in the Turkish National Corpus.



As shown in Figure D1 and D2, frequency counts of the adjective-noun combinations follow Zipf-like skewed distributions in the BNC and the TNC, with a small number of high-

frequency collocations, and a very large number of low-frequency adjective-noun combinations.

Experimental materials in both languages

Table D2 Complete list of English items

Adjective-noun pairs	Adjective frequency counts	Noun frequency counts	Phrasal frequency Counts	Frequency band
Long time	5.69	6.13	4.61	High-frequency
Young people	5.45	6.03	4.53	High-frequency
Recent years	5.14	5.89	4.40	High-frequency
Local government	5.61	5.74	4.48	High-frequency
Old man	5.67	5.71	4.34	High-frequency
Labour party	5.38	5.54	4.59	High-frequency
Hot water	4.89	5.48	3.97	High-frequency
Left hand	5.60	5.47	3.92	High-frequency
Economic development	5.31	5.45	3.98	High-frequency
Nuclear power	4.85	5.44	4.08	High-frequency
Blue eyes	4.95	5.38	3.90	High-frequency
Common law	5.24	5.37	4.01	High-frequency
Main road	5.34	5.37	3.82	High-frequency
High level	5.53	5.37	3.97	High-frequency
Social policy	5.56	5.36	3.89	High-frequency
Mental health	4.70	5.33	3.93	High-frequency
Prime minister	5.03	5.32	4.93	High-frequency
Annual report	4.85	5.32	3.75	High-frequency
Front door	5.27	5.31	4.16	High-frequency
European community	5.25	5.31	4.26	High-frequency
Free trade	5.26	5.24	3.71	High-frequency
Middle class	5.05	5.20	3.92	High-frequency
White paper	5.31	5.17	4.00	High-frequency
Bad news	5.12	5.10	3.78	High-frequency
Dark hair	5.05	5.08	3.49	High-frequency

Human rights	5.23	5.05	4.08	High-frequency
Rapid growth	4.50	5.05	3.47	High-frequency
Medical treatment	4.91	5.03	3.61	High-frequency
Vast majority	4.63	4.94	3.88	High-frequency
Senior officer	4.86	4.90	3.53	High-frequency
Inner world	4.59	5.70	2.85	Low-frequency
Difficult life	5.28	5.69	3.11	Low-frequency
Lovely house	4.73	5.64	2.79	Low-frequency
Warm place	4.77	5.62	2.86	Low-frequency
Poor children	5.11	5.60	3.63	Low-frequency
Similar case	5.21	5.60	2.82	Low-frequency
Whole company	5.42	5.54	2.76	Low-frequency
Certain point	5.28	5.54	3.00	Low-frequency
Vital information	4.65	5.53	2.98	Low-frequency
Small head	5.58	5.50	2.81	Low-frequency
Foreign business	5.15	5.49	2.68	Low-frequency
Round face	5.43	5.46	2.85	Low-frequency
Great service	5.61	5.43	2.78	Low-frequency
Tiny room	4.66	5.40	2.75	Low-frequency
Special court	5.28	5.40	2.64	Low-frequency
Easy question	5.10	5.35	2.49	Low-frequency
Elderly mother	4.64	5.33	2.41	Low-frequency
Physical body	4.92	5.34	2.74	Low-frequency
Strong voice	5.14	5.34	2.67	Low-frequency
Important city	5.53	5.30	2.57	Low-frequency
Fair idea	4.90	5.27	2.43	Low-frequency
Good land	5.85	5.27	2.79	Low-frequency
Extra hours	4.92	5.21	2.65	Low-frequency
Full authority	5.40	5.21	2.65	Low-frequency
Only friend	6.12	5.16	2.66	Low-frequency
New award	6.04	5.12	2.93	Low-frequency
Away game	5.62	5.11	2.32	Low-frequency
Soft material	4.71	5.07	2.36	Low-frequency

Suitable software	4.72	4.92	2.04	Low-frequency
Dry glass	4.75	4.92	2.07	Low-frequency
Dirty time	4.37	6.13	1.41	Baseline
Sudden people	4.56	6.03	1.54	Baseline
Nice years	5.05	5.89	1.54	Baseline
Deep government	4.95	5.74	1.23	Baseline
Far man	5.54	5.71	1.90	Baseline
Hard party	5.29	5.54	0.00	Baseline
Late water	5.23	5.48	1.69	Baseline
Current hand	5.10	5.47	1.64	Baseline
Able development	5.42	5.45	0.94	Baseline
Sure power	5.32	5.44	1.54	Baseline
General eyes	5.53	5.38	0.00	Baseline
Regular law	4.82	5.37	1.41	Baseline
Basic road	4.99	5.37	1.23	Baseline
Big level	5.34	5.37	1.54	Baseline
Huge policy	4.83	5.36	1.23	Baseline
Single health	5.20	5.33	1.23	Baseline
Real minister	5.30	5.32	1.23	Baseline
Red report	5.11	5.32	1.54	Baseline
Political door	5.42	5.31	1.23	Baseline
Short community	5.24	5.31	1.54	Baseline
Clear trade	5.34	5.24	1.69	Baseline
Public class	5.53	5.20	1.64	Baseline
Outside paper	5.27	5.17	0.00	Baseline
True news	5.19	5.10	1.41	Baseline
British hair	5.49	5.08	1.23	Baseline
Low rights	5.17	5.05	0.00	Baseline
Chief growth	4.99	5.05	0.00	Baseline
Wide treatment	5.02	5.03	1.23	Baseline
Light majority	5.30	4.94	0.94	Baseline
Total officer	5.19	4.90	0.94	Baseline
Necessary world	5.20	5.70	1.64	Baseline

Final life	5.13	5.69	1.90	Baseline
Fast house	4.81	5.64	1.23	Baseline
Firm place	5.03	5.62	1.64	Baseline
Official children	4.93	5.60	1.23	Baseline
Dead case	5.02	5.60	1.41	Baseline
Green company	5.10	5.54	1.72	Baseline
Top point	5.35	5.54	1.23	Baseline
United information	5.23	5.53	1.54	Baseline
Various head	5.13	5.50	1.79	Baseline
Married business	4.93	5.49	0.94	Baseline
National face	5.53	5.46	1.79	Baseline
Natural service	5.09	5.43	1.41	Baseline
Key room	5.04	5.40	1.41	Baseline
Happy court	5.00	5.40	1.23	Baseline
French question	5.17	5.35	1.23	Baseline
Significant mother	5.02	5.33	0.00	Baseline
Early body	5.47	5.34	1.23	Baseline
Sorry voice	4.98	5.34	1.23	Baseline
Serious city	5.03	5.30	1.23	Baseline
Cold idea	5.01	5.27	0.94	Baseline
Ready land	4.94	5.27	1.41	Baseline
Large hours	5.48	5.21	1.54	Baseline
Future authority	5.29	5.21	1.23	Baseline
Wrong friend	5.14	5.16	1.41	Baseline
Obvious award	4.86	5.12	0.00	Baseline
Individual game	5.22	5.11	0.94	Baseline
Past material	5.35	5.07	1.23	Baseline
Little software	5.74	4.92	0.94	Baseline
Male glass	4.98	4.92	0.00	Baseline

Table D3. Complete list of Turkish items

Adjective-noun pairs	Adjective frequency counts	Noun frequency counts	Phrasal frequency counts	Frequency band
Genç adam	5.61	5.69	4.22	High-frequency
Sivil toplum	5.09	5.36	4.67	High frequency
Ertesi gün	4.99	6.03	4.77	High frequency
Milli egitim	5.45	5.7	4.61	High frequency
Bati avrupa	5.32	5.57	4.12	High frequency
Soguk savas	5.13	5.36	4.13	High frequency
Yerel yönetim	5.23	5.38	4.16	High frequency
Bilimsel arastirma	5.08	5.29	3.72	High frequency
Uzun dönem	5.84	5.21	3.99	High frequency
Dis ticaret	5.52	5.25	4.59	High frequency
Erken seçim	5.11	5.18	3.64	High frequency
Orta sinif	5.42	5.15	3.87	High frequency
Kimyasal madde	4.76	5.16	3.56	High frequency
Genel baskan	5.86	5.16	4.43	High frequency
Yakin iliski	5.58	5.17	3.84	High frequency
Yabanci sermaye	5.48	5.24	4.25	High frequency
Sosyal güvenlik	5.67	5.29	4.65	High frequency
Geçen yil	5.62	5.88	4.66	High frequency
Kisa süre	5.62	5.65	4.81	High frequency
Kücük kiz	5.85	5.54	4.25	High frequency
Yogun bakim	5.21	4.52	3.46	High frequency
Olumlu yanit	5.21	4.99	3.59	High frequency
Agir ceza	5.42	5.03	3.89	High frequency
Serbest meslek	5.17	5.03	3.57	High frequency
Yüksek faiz	5.79	5	3.73	High frequency
Kirmizi sarap	5.15	4.59	3.36	High frequency
Güçlü kadin	5.33	5.79	3.16	Low frequency
Küresel dünya	4.87	5.78	3.07	Low frequency
Yalniz çocuk	5.46	5.67	3.49	Low frequency
Çabuk karar	4.96	5.58	3.06	Low frequency

Kötü durum	5.46	5.51	3.2	Low frequency
Aci haber	5.16	5.34	2.85	Low frequency
Siyasi hayat	5.45	5.35	2.75	Low frequency
Görsel sanat	4.51	5.31	2.51	Low frequency
Nitelikli isçi	4.53	5.25	2.69	Low frequency
Haksiz vergi	4.62	5.27	4.14	Low frequency
Sürekli görev	5.5	5.24	2.77	Low frequency
Askeri destek	5.31	5.2	2.78	Low frequency
Kalin kitap	4.79	5.3	2.69	Low frequency
Yanlis tercih	5.38	5.23	2.56	Low frequency
Ulusal hukuk	5.35	5.21	2.89	Low frequency
Belirli konu	5.23	5.33	2.74	Low frequency
Farkli düsünce	5.79	5.19	3.04	Low frequency
Kolay iletisim	5.4	5.2	2.75	Low frequency
Ilgili ögrenci	5.92	5.1	2.88	Low frequency
Fazla üretim	5.88	5.42	3.21	Low frequency
Güzel göz	5.84	5.49	2.97	Low frequency
Benzer ülke	5.33	5.42	2.69	Low frequency
Tüm hafta	5.88	5.34	2.94	Low frequency
Özel örnek	5.89	5.32	2.88	Low frequency
Ana yemek	5.47	5.27	3	Low frequency
Ortak sonuç	5.43	5.29	2.74	Low frequency
Kuzey adam	5.09	5.69	0	Baseline
Hizli toplum	5.31	5.36	0	Baseline
Ileri gün	5.41	6.03	0	Baseline
Kara egitim	5.22	5.7	0	Baseline
Mümkün avrupa	5.56	5.57	0	Baseline
Uygun savas	5.64	5.36	0	Baseline
Hos yönetim	5.09	5.38	0	Baseline
Canli arastirma	5.06	5.29	0	Baseline
Ünlü dönem	5.21	5.21	0	Baseline
Derin ticaret	5.23	5.25	0	Baseline
Geçici seçim	4.89	5.18	0	Baseline

Açik sinif	5.65	5.15	0	Baseline
Yasli madde	5.18	5.16	0	Baseline
Önemli baskan	6.07	5.16	0	Baseline
Beyaz iliski	5.34	5.17	0	Baseline
Yazili sermaye	4.99	5.24	0	Baseline
Demokratik güvenlik	5.1	5.29	0	Baseline
Gerçek yil	5.57	5.88	0	Baseline
Hazir süre	5.15	5.65	0	Baseline
Egemen kiz	4.93	5.54	0	Baseline
Karanlik bakim	5	4.52	0	Baseline
Kültürel yanit	5.2	4.99	0	Baseline
Bos ceza	5.25	5.03	0	Baseline
Emin meslek	5.01	5.03	0	Baseline
Çagdas faiz	5.03	5	0	Baseline
Zor sarap	5.46	4.59	0	Baseline
Temel kadin	5.66	5.79	0	Baseline
Sicak dünya	5.21	5.78	0	Baseline
Kesin çocuk	5.17	5.67	0	Baseline
Bagli karar	5.58	5.58	0	Baseline
Resmi durum	5.27	5.51	0	Baseline
Temiz haber	4.87	5.34	0	Baseline
Dogru hayat	5.96	5.35	0	Baseline
Sert sanat	5.04	5.31	0	Baseline
Büyük isçi	6.24	5.25	0	Baseline
Mutlu vergi	5.24	5.27	0	Baseline
Yavas görev	5.35	5.24	0	Baseline
Rahat destek	5.14	5.2	0	Baseline
Dogal kitap	5.42	5.3	0	Baseline
Tam tercih	5.82	5.23	0	Baseline
Genis hukuk	5.43	5.21	0	Baseline
Yesil konu	5.11	5.33	0	Baseline
Koca düsünce	4.98	5.19	0	Baseline
Eski iletisim	5.74	5.2	0	Baseline

Yasal ögrenci	5	5.1	0	Baseline
Uzak üretim	5.27	5.42	0	Baseline
Geleneksel göz	5.08	5.49	0	Baseline
Hafif ülke	5	5.42	0	Baseline
Basit hafta	5.08	5.34	0	Baseline
Bütün örnek	6.08	5.32	0	Baseline
Merkez yemek	5.2	5.27	0	Baseline
Ince sonuç	5.22	5.29	0	Baseline

Appendix E

Visual illustration of the Response Times

As our dependent variable we used log transformed response times rather than raw response times. As can be seen in Figure 2, log transformed response times (on the right) were considerably closer to the normal distribution than raw response times (on the left). Also raw response times created convergence issues.

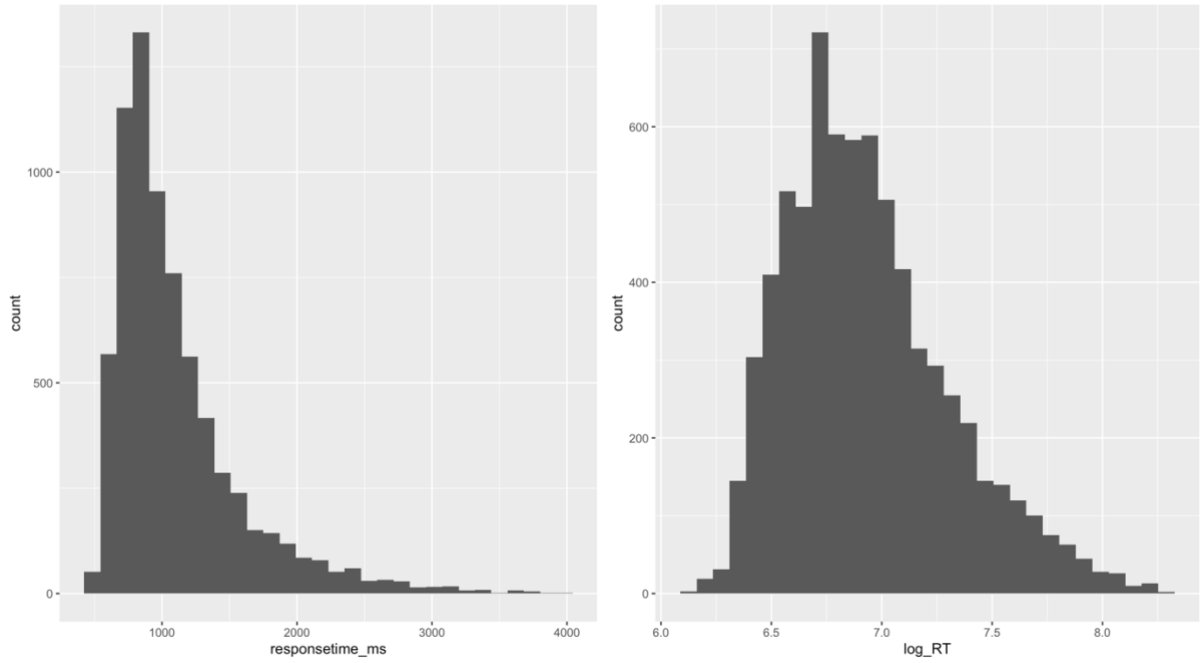


Figure D1. Comparing the distributions of log transformed and raw response times.

Additional Model Tables

Table E1. Maximal Model 3: Investigating English and Turkish speakers' sensitivity to individual word and collocation frequency count

Fixed effects	***b***	***SE***	***t***	***Df***	***p***
Intercept	6.954	0.020	338.81	116.481	<.001
English vs Turkish	0.075	0.040	1.872	108.807	.063
Collocation frequency	-0.165	0.009	-17.061	174.474	<.001
Nouns' order of occurrence	0.000	0.006	0.016	6952.013	.987
Lemma noun frequency	-0.010	0.013	-0.782	259.489	.434
Adjective frequency	0.017	0.007	2.421	221.145	.016
Noun frequency	0.022	0.012	1.811	237.440	.071
Length (Number of letters)	0.030	0.006	4.322	216.843	<.001
English vs Turkish x Collocation frequency	-0.066	0.021	-3.120	201.441	.002
Collocation frequency x adjective frequency	0.008	0.014	0.595	216.059	.552
English vs Turkish x Noun frequency	0.009	0.024	0.399	241.553	.690
Collocation frequency x Noun frequency	-0.004	0.010	-0.400	700.845	.689

English vs Turkish x Adjective frequency	0.008	0.014	0.595	216.069	.552
English vs Turkish x Lemma noun frequency	-0.006	0.026	-0.242	267.053	.808
Collocation frequency x Lemma Noun frequency	-0.000	0.010	-0.005	212.004	.966

Table E2. Additional Model 3: Investigating English and Turkish speakers' sensitivity to individual word and collocation frequency count (Inflectional family frequency)

Fixed effects	***b***	***SE***	***t***	***Df***	***p***
Intercept	6.934	0.019	352.112	97.114	<.001
English vs Turkish	0.088	0.039	2.247	99.494	.026
Lemma collocation frequency	-0.143	0.011	-12.731	231.921	<.001
Adjective frequency	0.019	0.008	3.635	209.931	<.001
Noun frequency	0.016	0.008	2.010	200.973	.045
Length (Number of letters)	0.029	0.007	3.890	204.325	<.001
English vs Turkish x Collocation frequency	-0.122	0.021	-5.585	231.785	<.001
Collocation frequency x Noun frequency	-0.007	0.007	-1.043	182.261	.298

Table E3 *Mixed effects Model 3, investigating English and Turkish speakers' sensitivity to individual word and collocation frequency counts excluding baseline items*

Fixed effects	***b***	***SE***	**95% CI**	***t***	***Df***	***p***
Intercept	6.968	0.022	[6.924, 7.011]	314.095	149.443.	<.001
English vs Turkish	0.091	0.043	[0.005, 0.177]	2.093	145.932	.038
Collocation frequency	-0.188	0.017	[-0.222, -0.155]	-11.019	141.237	<.001
Adjective frequency	0.019	0.008	[0.003, 0.035]	2.441	116.374	.016
Noun frequency	0.014	0.013	[-0.011, 0.040]	1.084	115.946	.280
Length (Number of letters)	0.030	0.008	[0.013, 0.047]	3.593	115.803	<.001
English vs Turkish x Collocation frequency	-0.089	0.032	[-0.152, -0.026]	-2.767	143.114	.006
Collocation frequency x Noun frequency	-0.000	0.012	[-0.016, 0.009]	-0.011	114.809	0.991

Note. R^2 marginal = .119; R^2 conditional =.332

Table E4. Maximal Model 4: Comparing English and Turkish speakers' response times for high-frequency, low-frequency, and baseline items

Fixed effects	***b***	***SE***	***t***	***Df***	***p***
Intercept	7.098	0.023	304.245	97.919	<.001
English vs Turkish	0.050	0.046	1.08	99.212	.279
Item Type: High frequency vs. Baseline	-0.367	0.019	-18.59	159.372	<.001
Item Type: Low frequency vs. Baseline	-0.183	0.019	-9.43	160.600	<.001
Nouns' order of occurrence	-0.001	0.006	-0.163	6982.642	.876
Adjective frequency	0.009	0.010	0.837	229.771	.403
Noun frequency	0.005	0.009	0.546	209.915	.585
Length (Number of letters)	0.031	0.006	5.028	224.238	<.001
English vs Turkish x High frequency (vs. Baseline)	-0.002	0.041	-0.067	173.448	.949
English vs Turkish x Low frequency (vs. Baseline)	-0.003	0.039	-0.098	163.891	.922
English vs Turkish x Adjective frequency	-0.000	0.013	-0.057	220.065	.954
English vs Turkish x Noun frequency	-0.014	0.013	-1.101	203.387	.272
High frequency vs. Baseline x Adjective frequency	0.018	0.018	1.018	200.311	0.309

Low frequency vs. Baseline x Adjective frequency	-0.004	0.014	-0.283	224.772	.777
High frequency vs. Baseline x Noun frequency	-0.016	0.014	-1.145	193.537	.256
Low frequency vs. Baseline x Noun frequency	-0.060	0.020	-2.96	218.550	.003
English vs Turkish x First vs second occurrence of nouns	-0.018	0.013	-1.356	6983.010	.175