Dilbilim Dergisi
Journal of Linguistics

İSTANBUL UNIVERSITY PRESS

Dilbilim Dergisi - Journal of Linguistics 37, (2021): 45-66
DOI: 10.26650/jol.2021.972166 Research Article

Lexical Associations in the L1 Turkish Mental Lexicon: Can L1 Lexical Intuition and a Representative Corpus Guide Teaching of Turkish as a Foreign Language Materials?

Hakan CANGIR*

*(Lecturer, PhD), Ankara University, School of Foreign Languages, Ankara, Turkiye

ORCID: H.C. 0000-0003-2589-2466

Corresponding author:
Hakan CANGIR,
Ankara University, School of Foreign Languages, Ankara, Turkiye
E-mail: hcangir@ankara.edu.tr; hakancangir@gmail.com

Submitted: 15.07.2021
Revision Requested: 06.11.2021
Last Revision Received: 11.11.2021
Accepted: 17.12.2021

Citation: Cangir, H. (2021). Lexical associations in the L1 Turkish mental lexicon: Can L1 lexical intuition and a representative corpus guide teaching of Turkish as a foreign language materials. *Dilbilim Dergisi - Journal of Linguistics, 37,* 45-66.
https://doi.org/10.26650/jol.2021.972166

ABSTRACT

Studies to date using corpus linguistic and psycholinguistic approaches have investigated the collocational links in both the L1 and L2 brain from different angles. They have attempted to test the role collocations play in the mental lexicon and question how corpus data can guide us in our investigations. However, there have been almost no attempts to explore the relationship between the L1 Turkish users' subjective judgements of collocational use (frequency) and the association strength of collocational items in Turkish evidenced in representative corpora. This research, mainly with a corpus-linguistic approach, aims to (a) detect the possible relationship between the *collocational links* in the L1 Turkish mental lexicon and the *collocational frequency profiles* on the TNC (Turkish National Corpus) (b) and thus question if L1 speaker intuitions (familiarity) regarding collocational frequency and the frequency profiles of the items, as evidenced through representative corpora, can/should be exploited in an attempt to create target vocabulary lists or vocabulary teaching materials for teaching Turkish as a foreign language. The results indicate that the collocational links in L1 Turkish users' mental lexicon seem to resemble the lexical associations represented in the TNC. To be more precise, higher Delta-P (1 > word 2) scores are associated with stronger collocational intuitions. Additionally, the Delta-P (1 > word 2) as well as the CEFR[1] level are significant predictors of subjective judgements for frequent collocational use. The results have been discussed in light of the psycholinguistic research highlighting collocational processing, and some pedagogical conclusions have been drawn.

Keywords: Collocation, Frequency, Intuition, L1 Turkish, L2 Turkish Teaching

1 (Common European Framework of Reference for Languages - Council of Europe, 1996)

1. Introduction

The psycholinguistic reality of collocational frequency is a long-debated issue in the literature. The idea that high frequency word combinations as evidenced in representative corpora may have a form of independent representation in the mental lexicon has been discussed in research studies investigating first and second language acquisition (e.g., Wray, 2002; Tomasello, 2003) and language processing in L1 and L2 (Ellis, 2002; Cangır et al., 2017; Göymen and Aygüneş, 2020; Öksüz et al., 2021). However, some researchers (e.g., Herbst, 1996) disagree with the psycholinguistic reality of collocations stating that real-life coincidences can account for the frequency of word combinations and that collocations do not have robust and fixed representations in the mind. A second disagreement stems from the fact that representative corpora cannot match the linguistic experience of any particular language user. Hoey (2005) accepts the fact that every language user has a unique repertoire of collocational items in his mental lexicon and that corpora by itself cannot reflect that psycholinguistic capacity comprehensively. However, he also claims that corpora can be utilized to investigate the types of lexical input L1 users are likely to have encountered and the evidence from the corpora can be used to test some psycholinguistic claims (i.e., *priming*). As he states, when language users encounter a lexical item or lexical patterns in their L1, they subconsciously detect and record the linguistic context where the patterns are employed. Frequent encounters with these patterns help users recognize the features of the context. The more frequent the encounters are with the patterns, the more likely language users tend to identify the lexical items and their constituent parts (i.e., its collocations). Therefore, as native users of a language, we recognize the patterns and chunks, recognize their frequency of use, make generalizations, and store them as they are, so when we see a part of them, we retrieve the other members of the chunk as the coexisting nature makes them linked in our mental lexicon (Christiansen and Arnon, 2017). It is revealed in the related literature that the detection of language patterns via the integration of multidimensional posterior distributions is regarded as analytic processing (Wray, 2002), multiple cue integration (Christiansen, Conway and Curtin, 2005), and lexical priming (Hoey, 2005). It can then be hypothesized that those who are native Turkish language users with experience of chunk-based language acquisition in their first language are sensitive to co-occurring patterns in language.

It is commonly agreed among researchers and frequently stated in the literature that *collocations*, which are considered to be co-occurring word combinations, are pervasive in native language. It is also well-acknowledged that collocations are recurring patterns in language use that can be proven by representative corpora through some statistical measures, such as MI value and *t*-score and can be classified under the heading of *formulaic language* (Schmitt, 2010). Research (e.g., Durrant and Doherty, 2010) suggests that they have partial psycholinguistic reality in the mental lexicon. Despite the common ground on the omnipresence of co-occurring word combinations in native language use, researchers and theoreticians in the

field of psycholinguistics, corpus linguistics, and phraseology have controversial ideas regarding the definition of collocations. There have been different approaches as to how collocations should be defined in the literature. The phraseological approach states that the compositionality of the word combinations is what matters. On the other hand, the frequency-based approach disregards the compositionality of the word pairs and underlines the importance of statistical evidence gathered from representative corpora.

Users' lexical experience and corpus evidence may not correlate strongly. Research evidence indicates that corpus frequency may not always reflect individual experiences of a certain lexical item. Therefore, we may need both objective frequency values provided by representative corpora and subjective frequency intuitions of L1 users based on their user experiences as a proxy to investigate the collocational links in the L1 mental lexicon. If we can find an association between subjective and objective frequency measures, we may then come up with a more realistic approach to choosing what words to teach to L2 learners. There have been certain attempts trying to highlight the relationship between objective frequency measure and subjective frequency ratings of native and non-native speakers. Siyanova and Schmitt (2008) gave their participants a questionnaire and asked them to rate the target collocations according to their commonness. They concluded based on the results that native speakers' intuition correlated more strongly with BNC frequencies than the non-native speaker intuitions, particularly for low frequency collocations. Chen and Dong (2019) claim that L2 users frequency intuitions correlate with the corpus frequencies, for high frequency word combinations in particular. Siyanova-Chanturia and Spina (2015) also conclude that both native and non-native users of Italian are sensitive towards objective frequency profiles of collocations as represented on corpora. They emphasize the fact that non-native users of Italian are as successful as native users in determining the commonness of frequent collocations. However, they also add that advanced users of L2 Italian are as sensitive as native users when detecting low frequency collocations. In a recent study, Cangır (2021) also finds a strong correlation between the L2 English users' frequency intuitions (who are instructors of English) and a collocational frequency measure, *t*-score. It may be concluded that there is somewhat a consensus towards the positive correlation between subjective frequency ratings of high frequency collocations in particular and the association strength measures computed through corpora. The phenomenon has been investigated through the lenses of the English and Italian languages. However, there have been no attempts, to the writer's knowledge, to investigate the relationship between the subjective and objective frequency measures of lexical items in Turkish.

The study of fixed expressions like collocations and the influence of frequency on the structuring of the mental lexicon is closely associated with usage-based models of language (Kemmer and Barlow, 2000). Such models suggest that a speaker's language system is shaped by their language experience or exposure throughout their lifetime. Studies (e.g., Durrant and Doherty, 2010) incorporating corpus and psycholinguistic evidence appear to prove this

relationship between the frequencies of occurrence of various aspects of the language and their representation in the native speaker mind.

Given this need in the design of corpus-informed and pedagogically sound vocabulary teaching materials and the claims in the literature regarding the effect of frequency of use on the mental representation of lexis, a more rational method can be devised in an attempt to extract the most functional words to teach for L2 users with the help of corpora and L1 (and L2) user experience.

Problem statement

Textbooks written to teach Turkish as a foreign language tend to disregard multi-word units, such as collocations in their syllabi, and there has been no attempt to build learner collocation dictionaries for Turkish language (Karadağ, 2020). In other words, students learning Turkish as a foreign language lack the explicit instruction of collocations through modern course books or dictionaries. However, native speakers of L1 Turkish have intuitive judgements regarding what nouns may follow certain adjectives which could also be taught to L2 users of Turkish to help them communicate better and more effectively in Turkish and sound more natural. Earlier research on native speaker intuitions about word level and collocational frequency indicate that there is a relationship between subjective judgements of collocational frequency (or familiarity) and association strength measures (Cangır, 2021), and it must be investigated further from the eyes of understudied languages, such as Turkish, to gather more conclusive results and enlighten the designers of language teaching materials. When constructing target vocabulary lists or designing materials for vocabulary teaching, designers can be informed by both corpora and native speaker intuitions. Furthermore, as the research indicates advanced L2 user intuitions and experiences can/should also feed into the process.

2. Method

2.1. Item Extraction

Three Turkish as a Foreign Language textbooks (*Yeni İstanbul*, *Hitit*, and *Yedi İklim*) have been investigated for their vocabulary teaching approach. Those books were scrutinized because they were designed and published by the two most established universities in Turkiye (İstanbul and Ankara University) and the only government institution (Yunus Emre) whose main aim is to teach Turkish language and help spread Turkish culture. Upon investigation, no evidence was found for collocation teaching, and it was detected that no emphasis was given on common nouns following certain adjectives. Adjectives were extracted from the target vocabulary lists of the two textbooks (Bölükbaş Kaya and Yılmaz, 2020) by the same publishing house (İstanbul University) written with the aim of teaching Turkish to foreigners. The two levels of the book (A2-B1) were chosen to test the effect of two different levels. The book was chosen

because it was one of the most commonly taught books in schools teaching Turkish as a foreign language and was the only book providing the learners with a target vocabulary list at the end of each unit with an emphasis on formulaic expressions. It should also be noted that the item extraction in this study was purposefully done through course books and their vocabulary lists as the starting point of this research study was the lack of emphasis on the use collocations in those textbooks and the potential use (benefits) of corpus as well as L1 speaker intuitions in guiding material designers and instructors as evidenced in earlier research.

2.2. Association Measures (AMs)

AMs are mathematical formulas indicating how strongly two words are connected to one another, and they provide evidence from a representative corpus. One way of doing collocational analysis is the raw frequency approach. That is, one can rely solely on the number of items two words tend to be in the company of one another. However, AMs go beyond that approach and take into account some more advanced parameters (Gablasova et al., 2017). There are certain well-acknowledged AMs commonly used in related studies, t-score, MI, MI2, and Delta-P. Each has its own strengths and weaknesses, and some research studies tend to use them as complementary variables in their analysis (e.g., Cangır, 2018). Four different AMs were utilized in this study and the headings below attempt to give further details about each measure.

2.2.1. T-Score

As Schmitt (2010) states, the t-score seems to favour items with very high frequency. The fact that it penalizes low frequency items with exclusive uses almost always in the company of one another makes the measurement susceptible to criticism. Another drawback of the calculation is that it does not take the corpus size into account, which makes it hard to produce comparable results. The t-score is calculated as an adjusted value of collocation frequency based on the raw frequency from which random co-occurrence frequency is subtracted. This is then divided by the square root of the raw frequency (Gablasova et al., 2017, p. 8).

2.2.2. MI-MI2

Unlike the t-score, MI-score favours exclusivity, and it has the potential to foreground the word pairs with a strong tendency to co-occur in the company of each other. “The MI-score uses a logarithmic scale to express the ratio between the frequency of the collocation and the frequency of random co-occurrence of the two words in the combination” (Church and Hanks, 1990). Its inclination to favour exclusive lexical items can also be regarded as its weakness as the calculation disregards high frequent items. MI2, which is an alternate version of MI with a slight modification, is considered to be the corrected version as the collocational frequency is squared during the calculation (Gablasova et al., 2017).

2.2.3. Delta-P

This AM (Gries, 2013) is of special importance for the current study since it has a directional nature. To be more precise, the calculations provide two different numbers indicating the bidirectional effect of each word in a collocation on one another (i.e., word 1 > word 2 and word 2 > word 1). Given the design of the current research, where the participants are provided with a cue word (i.e., a node) through which they are expected to form a collocation, the directional influence of the node and the potential collocate becomes significant.

2.3. Participants

Thirty-five native speakers of Turkish took part in the online questionnaire. The majority of the participants (96%) were academicians at different universities in Turkiye. All the subjects were L1 Turkish users mainly from the capital city of Turkiye, Ankara (*N*=20). There were also native Turkish participants from other cities in Turkiye (*N*=13) and abroad (*N*=2). Their ages range from 27 to 53 (*Mean*=35.5). They were informed about the overall purpose of the research, and all took part in the study voluntarily.

2.4. Instrument

A simple questionnaire was designed to gather participants' collocational preferences. They were given a node word (i.e., an adjective) followed by a blank space and were asked to write the top three collocates (i.e., nouns) based on their native speaker intuitions (See Appendix A for a sample screen of the questionnaire). Some questions regarding the participants' educational background and age were also present in the questionnaire.

2.5. Research Steps

- L1 Turkish users wrote the top three collocates of the provided nodes (i.e., adjective) intuitively.
- Two of the collocates they mostly agreed on were extracted from the spreadsheet for each target item, and the number of the instances of a preferred collocate was used as an intuitive frequency value in the correlation and regression analyses as the dependent variable. For instance, if 30 participants wrote 'telefon-*phone*' as the collocate of the node 'akıllı-*smart*,' then the subjective frequency value was set to 30 for the collocation 'akıllı telefon-*smart phone*.'
- The following step was to calculate the association strength measures of the detected collocational items. To achieve that, Turkish National Corpus (Aksan et al., 2012) was exploited through which the researcher gathered word level frequency profiles and the number collocational instances. The AMs were calculated separately using a spreadsheet.
- The relationship between subjective frequency values and corpus-extracted AMs were calculated using a correlation analysis. The analysis was conducted with the help of

the open resource statistical software, Jamovi (2021). The software builds on the R programming language (R Core Team, 2021) and provides the researchers with a user-friendly interface.

- Using the subjective frequency values as the dependent variable and the AMs as the predictive variables, a regression model was designed in an attempt to detect the best predictors of subjective collocational frequency using the same software.

3. Results

As a preliminary analysis, the descriptive details of the AMs under investigation have been computed to present their mean scores and the standard deviations. Table 1 illustrates the mean AMs in two different proficiency levels.

Table 1: Association Strength Measures

	Level	t	MI	MI2	Delta-P (word 1 > word 2)	Delta-P (word 2 > word 1)
N	A2	40	40	40	40	40
	B1	40	40	40	40	40
Mean	A2	5.14	7.22	10.2	0.0201	0.0286
	B1	4.20	7.53	9.22	0.0295	0.00897
Median	A2	4.03	6.69	9.72	0.0153	0.00290
	B1	3.16	7.40	8.73	0.0172	0.00335
Standard deviation	A2	4.14	2.52	2.15	0.0220	0.0780
	B1	3.61	2.43	2.13	0.0393	0.0174

Overall, the numbers indicate that the t-score is higher in the A2 than in the B1 level as it favours pure frequency and the nodes and collocates in this level tend to have more frequent instances in the TNC. There is not much difference in terms of the MI-score between the two levels when the items are taken into account as a whole though some exceptions exist in certain collocational items. Although one can see a slightly bigger difference between the mean scores of Delta-P in the A2 and B1 levels, the only significant difference between the two levels is for the MI2-score (p=.04). Figure 1 illustrates the visual representation of the difference in the two different proficiency levels. (See also Appendix C for the illustrations of the other AMs).

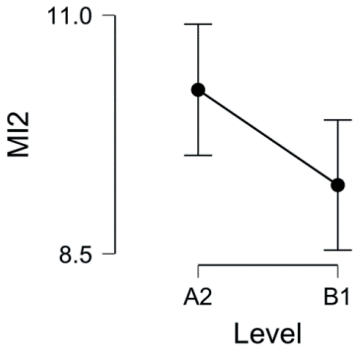


Figure 1. MI2 in A2 vs. B1

Table 2 shows the collocates the L1 Turkish participants mainly agreed on in two different levels and reports both the number of responses and their percentages to give the readers a general insight into the overall predictive power of the node words.

Table 2: Collocate Intuitions for A2

Node	Collocates	N. of Responses	Node	Collocates	N. of Responses
1. Akıllı *(smart)*	Telefon (*phone*)	21 (60%)	11. Sıradan (*ordinary*)	Gün (*day*)	15 (42.8%)
	Çocuk (*kid*)	15 (42.8%)		İnsan (*human*)	13 (37.1%)
2. Düşünceli *(thoughtful)*	İnsan (*person*)	23 (65.7%)	12. Ölümsüz (*immortal*)	Eser (*art*)	12 (34.2%)
	Davranış (*attitude*)	7 (20%)		Aşk *(love)*	9 (25.7%)
3. Faydalı *(useful)*	Bilgi (*information*)	17 (48.5%)	13. Sihirli (*magical*)	Değnek *(stick)*	19 (54.2%)
	Besin (*food*)	6 (17.1%)		Dokunuş (*touch*)	9 (25.7%)
4. Güvenilir *(reliable)*	İnsan (*person*)	13 (37.1%)	14. Bilimsel (*scientific*)	Çalışma *(study)*	12 (34.2%)
	Arkadaş (*friend*)	11 (31.4%)		Makale *(article)*	12 (34.2%)
5. Huzurlu *(peaceful)*	Ortam (*environment*)	19 (54.2%)	15. Günlük (*daily*)	Süt *(milk)*	13 (37.1%)
	Ev (*house*)	18 (51.4%)		Yaşam *(life)*	6 (17.1%)
6. Lezzetli *(delicious)*	Yemek (*meal*)	32 (91.4%)	16. Karanlık *(dark)*	Oda (*room*)	24 (68.5%)
	Tarif (*recipe*)	6 (17.1%)		Sokak (*street*)	9 (25.7%)
7. Neşeli *(cheerful)*	Çocuk (*kid*)	15 (42.8%)	17. Polisiye *(detective)*	Roman *(novel)*	28 (80%)
	Günler *(days)*	11 (31.4%)		Film *(film)*	16 (45.7%)
8. Cömert *(generous)*	İnsan *(person)*	15 (42.8%)	18. Yöresel *(regional)*	Yemek *(food)*	24 (68.5%)
	Davranış *(attitude)*	6 (17.1%)		Kıyafet *(clothing)*	10 (28.5%)
9. Fırsatçı *(opportunist)*	İnsan *(person)*	12 (34.2%)	19. Yüksek *(high)*	Lisans *(degree)*	13 (37.1%)
	Esnaf *(tradesman)*	8 (22.8%)		Bina *(building)*	8 (22.8%)
10. Konforlu *(comfortable)*	Ev *(house)*	16 (45.7%)	20. Yoğun *(intensive)*	Bakım *(care)*	9 (25.7%)
	Araba *(car)*	14 (40%)		Trafik *(traffic)*	8 (22.8%)

Collocational items with the highest number of responses (e.g., lezzetli yemek-*delicious food and* polisiye roman-*detective novel*) seem to have relatively high MI-scores (8.30 and 11.03, respectively) as evidenced in a representative corpus. In other words, it could tentatively be said that the higher the MI-score, the more likely the participants tended to agree on a collocate item. It can then be claimed that the exclusivity of the collocation (e.g., polisiye roman-*detective novel*) as represented in TNC appears to have a psycholinguistic reality, and the exclusive use in everyday life (rather than pure frequency) seems to affect the mental lexicon structuring of L1 Turkish users. Table 3 reports on the same issue in a different CEFR level.

Table 3: Collocate Intuitions for B1

Node	Collocates	N. of Responses	Node	Collocates	N. of Responses
1. Büyüleyici *(mesmerizing)*	Manzara *(view)*	11 (31.4%)	11. Akılcı *(rationalist)*	Çözüm *(solution)*	10 (28.5%)
	Güzellik *(beauty)*	7 (20%)		Yaklaşım *(attitude)*	10 (28.5%)
2. Endişeli *(anxious)*	Bekleyiş *(wait)*	10 (28.5%)	12. Yaygın *(widespread)*	İnanış *(belief)*	10 (28.5%)
	Bakış *(look)*	9 (25.7%)		Kullanım *(use)*	9 (25.7%)
3. Masrafsız *(inexpensive)*	Araba *(car)*	12 (34.2%)	13. Şaşırtıcı *(surprising)*	Sonuç *(result)*	8 (22.8%)
	Kredi *(loan)*	7 (20%)		Cevap *(answer)*	6 (17.1%)
4. Yapışık *(siamese, attached)*	İkiz *(twin)*	34 (97.1%)	14. Gerçekçi *(realistic)*	Yaklaşım *(approach)*	14 (40%)
	Ev *(house)*	6 (17.1%)		İnsan *(person)*	8 (22.8%)
5. Asgari *(minimum)*	Ücret *(wage)*	35 (100%)	15. Yorucu *(tiring)*	İş *(work)*	20 (57.1%)
	Müşterek *(denominator)*	7 (20%)		Yolculuk *(journey)*	8 (22.8%)
6. Dolgun *(high)*	Maaş *(salary)*	27 (77.1%)	16. Talihsiz *(unfortunate)*	Kaza *(accident)*	10 (28.5%)
	Dudak *(lips)*	17 (48.5%)		Açıklama *(statement)*	5 (14.2%)
7. Elverişsiz *(unfavourable)*	Hava *(weather)*	15 (42.8%)	17. Bağımsız *(independent)*	Ülke *(country)*	13 (37.1%)
	Toprak *(soil)*	9 (25.7%)		Devlet *(government)*	6 (17.1%)
8. Özenli *(attentive)*	İş *(work)*	12 (34.2%)	18. Çarpıcı *(striking)*	Sonuç *(result)*	8 (22.8%)
	Çalışma *(study)*	7 (20%)		Gerçek *(truth)*	6 (17.1%)
9. Zahmetli *(difficult)*	İş *(work)*	29 (82.8%)	19. Karşılıklı *(mutual)*	Anlayış *(understanding)*	13 (37.1%)
	Yemek *(food)*	15 (42.8%)		Konuşma *(conversation)*	4 (11.4%)
10. Verimli *(fertile)*	Toprak *(land)*	18 (51.4%)	20. İkiyüzlü *(two-faced)*	İnsan *(person)*	27 (77.1%)
	Çalışma *(work)*	14 (40%)		Davranış *(behaviour)*	6 (17.1%)

As the level (suggested by the target coursebooks) goes up, the number of the nodes with the same collocate responses appears to increase as in 'asgari ücret-*minimun wage*.' This agreement among the respondents is also reflected in the detected association measures, i.e., MI-score. To be more precise, the higher the agreement, the higher the MI-score is, which can be attributed to the uniqueness of the collocations in question (e.g., asgari ücret-*minimum wage*,

yapışık ikiz-*siamese twins*) and thus the conventionality of the related word combinations. Additionally, the predictive power of the node on the collocate seems to be stronger for the items in B1 level as evidenced by the Delta-P (word 1 > word 2) value. Figure 2 represents the difference between the two levels. (See also Appedix C for a visual representation of the other AMs).

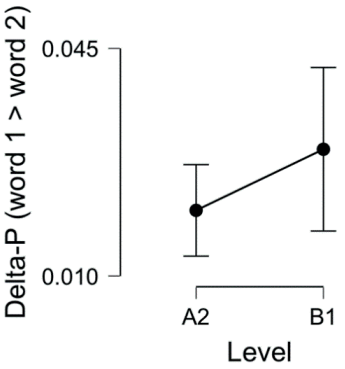


Figure 2. Delta-P (word 1 > word 2) for A2 vs. B1

To investigate the relationship between the L1 users' collocation intuitions (i.e., their familiarity with the combinations) and corpus evidence, a correlation analysis was computed. Table 4 illustrates both significant and non-significant correlations between the intuition scores (i.e., the mean number of responses) and the AMs.

Table 4: Correlation Matrix for Both Levels

		Intution	t-score	MI	MI2	Delta-P (word 1 > word 2)	Delta-P (word 2 > word 1)
Intution	Pearson's r	—					
	p-value	—					
t-score	Pearson's r	0.148	—				
	p-value	0.191	—				
MI	Pearson's r	0.188	0.076	—			
	p-value	0.096	0.504	—			
MI2	Pearson's r	-0.157	0.628***	-0.295**	—		
	p-value	0.164	< .001	0.008	—		
Delta-P (word 1 > word 2)	Pearson's r	0.443***	0.308**	0.416***	-0.349**	—	
	p-value	< .001	0.005	< .001	0.002	—	
Delta-P (word 2 > word 1)	Pearson's r	0.107	0.504***	0.440***	0.268*	0.121	—
	p-value	0.345	< .001	< .001	0.016	0.285	—

Note: * p < .05, ** p < .01, *** p < .001

The numbers indicate that there is a strong positive correlation between the intuition scores and Delta-P word 1 > word 2 (r= .44, p<.001). In other words, as the Delta-P scores increased, number of responses went up (i.e., the more the participants agreed on a collocate). This finding validates the writer's assumptions based on the structure of the task. The participants were given the node word and were expected to guess the following collocate. Thus, the higher the predictive power of the noun on the adjective (i.e., the higher the Delta-P), the more easily and confidently the participants responded to the task and their answers seemed to correlate with the corpus output to a greater extent.

To see the picture from a slightly different angle, I split the data into two different levels and computed the correlational analysis again.

Table 5: Correlation Matrix for A2

		Intution	t-score	MI	MI2	Delta-P (word 1 > word 2)	Delta-P (word 2 > word 1)
Intution	Pearson's r	—					
	p-value	—					
t-score	Pearson's r	0.033	—				
	p-value	0.839	—				
MI	Pearson's r	0.110	0.071	—			
	p-value	0.498	0.662	—			
MI2	Pearson's r	-0.229	0.682***	-0.246	—		
	p-value	0.155	< .001	0.126	—		
Delta-P (word 1 > word 2)	Pearson's r	0.522***	0.134	0.423**	-0.351*	—	
	p-value	< .001	0.410	0.007	0.026	—	
Delta-P (word 2 > word 1)	Pearson's r	0.057	0.564***	0.554***	0.278	0.130	—
	p-value	0.726	< .001	< .001	0.083	0.423	—

Note: * p < .05, ** p < .01, *** p < .001

As far as the A2 level items are concerned, the results do not show a different picture. It was still the Delta-P word 1 > word 2 highlighting a strong relationship with the participant responses (r= .52, p<.001). To be more precise, as the predictive power of the adjectives on the nouns increase in the participants' lexicon, the Delta-P word 1 > word 2 score also goes up, which indicates the positive relationship between the subjective judgements and the objective measures. However, there was a slight difference considering the B1 level items.

Table 6: Correlation Matrix for B1

		Intution	t-score	MI	MI2	Delta-P (word 1 > word 2)	Delta-P (word 2 > word 1)
Intution	Pearson's r	—					
	p-value	—					
t-score	Pearson's r	0.226	—				
	p-value	0.161	—				
MI	Pearson's r	0.280	0.100	—			
	p-value	0.080	0.540	—			
MI2	Pearson's r	-0.182	0.554***	-0.334*	—		
	p-value	0.261	< .001	0.035	—		
Delta-P (word 1 > word 2)	Pearson's r	0.469**	0.492**	0.434**	-0.334*	—	
	p-value	0.002	0.001	0.005	0.035	—	
Delta-P (word 2 > word 1)	Pearson's r	0.275	0.554***	0.467**	0.298	0.466**	—
	p-value	0.086	< .001	0.002	0.062	0.002	—

Note: * p < .05, ** p < .01, *** p < .001

Table 6 shows that Delta-P word 1 > word 2 still plays the most important role displaying a strong positive correlation (r=0.46, p=0.002). Another noteworthy result could be the near significance for the MI and Delta-P (word 2 > word 1). Although the p values (r=0.28, p=0.08 and r=0.27, p=0.08, respectively) do not signify a statistically significant correlation, their potential to indicate relationships may encourage further investigation and future research with more items or more participants can reach to stronger conclusions.

Before proceeding with the general linear model, a multicollinearity check was computed for the potential independent variables to avoid strong correlations between the AMs which could shadow some important findings as the results may become uninterpretable. As Myers (1990) states, variance inflation factors (VIFs) that are greater than 5.0 indicate critical levels of multicollinearity where the coefficients are poorly estimated, and the p-values are questionable. The collinearity analysis revealed that none of the VIF values were greater than 5.0 and the highest tolerance was 0.92, so all the potential predictors were kept for the final analysis.

Table 7: General Linear Model for both Levels

				95% Confidence Interval					
Names	**Effect**	**Estimate**	**SE**	**Lower**	**Upper**	**β**	**df**	**t**	**p**
(Intercept)	(Intercept)	1.0668	0.0210	1.0250	1.10862	0.000	77	50.81	<0.001
Level	B1 - (A2, B1)	-0.0491	0.0212	-0.0914	-0.00685	-0.230	77	-2.31	0.023
Delta-P (word 1 > word 2)	Delta-P (word 1 > word 2)	3.1910	0.6676	1.8617	4.52025	0.477	77	4.78	<0.001
Goodness of fit: R 0.49, R^2 0.25									

Delta-P dominates the model mainly due to the nature of the given task to the participants as discussed earlier. Still, the finding proves that an association strength measure (Delta-P) can predict the intuitive collocational judgements of L1 Turkish users (β=.47, p<.001). As the table indicates, the level of the target items also has an effect on the subjective judgements of the participants (β=.23, p=.02). The two variables explain 25% of the variance. Since a significant effect was detected, two separate general linear models were computed to see the possible difference between A2 and B1 level lexical items. Table 8 and 9 highlights the significant predictors of L1 Turkish intuitions about collocational frequency.

Table 8: General Linear Model for A2

			95% Confidence Interval					
Names	**Estimate**	**SE**	**Lower**	**Upper**	**β**	**df**	**t**	**p**
(Intercept)	1.10	0.0260	1.05	1.15	0.000	38	42.37	<0.001
Delta-P (word 1 > word 2)	4.52	1.1961	2.10	6.94	0.522	38	3.78	<0.001
Goodness of fit: R .58, R^2 .34								

Table 9: General Linear Model for B1

			95% Confidence Interval					
Names	**Estimate**	**SE**	**Lower**	**Upper**	**β**	**df**	**t**	**p**
(Intercept)	12.5	1.04	10.5	14.6	0.000	38	12.12	<0.001
Delta-P (word 1 > word 2)	112.9	26.68	58.9	166.9	0.566	38	4.23	<0.001
Goodness of fit: R .56, R^2 .32								

Delta-P still dominates the model and can explain 34 and 32 % of the variance by itself, respectively. In both models, Delta-P (word 1 > word 2) can be seen as the only significant predictor of frequency intuition of L1 Turkish users (β=.52, p<.001 and β=.56, p<.001). Regardless of the level, the Delta-P (word 1 > word 2) and the participants' intuitions for word combinations are strongly associated. When one goes up, the other one also follows. In other words, the way the (adj+noun) lexical partners are associated in the lexicon is reflected through corpus output. The predictive power of the adjectives on the nouns in the mental lexicon seems to overlap with the predictive power evidenced by a corpus output (i.e., Delta-P word 1 > word 2). A different task (e.g., a frequency rating task) can yield partly different results, so future research can replicate the research design by manipulating an acceptability judgement task (See Öksüz, 2019 for the methodological design).

4. Discussion

The current study tentatively claims that the stronger the Mutual Information of a collocation is, the stronger the participants' predictions regarding the nouns following the adjectives are. In other words, the exclusivity of the chosen collocational items is represented in the mental lexicon of the L1 Turkish participants. Their lexical choices based on their intuitions and the strength of word associations can be given as evidence to this. This finding is partly in line with earlier research (e.g., Siyanova and Schmitt, 2008; Siyanova-Chanturia and Spina, 2015) indicating that L1 users are sensitive towards the frequency of word combinations, the constituent parts of which are exclusively in company of one another (i.e., with higher MI value). The study also finds evidence for a positive correlation between the L1 Turkish users' collocational intuitions and the association strength measure (Delta-P word 1 > word 2) indicating how strongly a node triggers a collocate item (ADJ+N in this research). In similar studies with partly different methodologies (e.g., Cangır, 2021), researchers found evidence for a positive correlation between L1 and L2 user intuitions and objective frequency means. However, the findings of the current research seem to extend the previous view and could be regarded as another steppingstone because they provide evidence for a directional spreading activation from the node to the collocate with the help of the Delta-P value, which previous studies with similar aims lack. To be more precise, it can be claimed that the stronger the Delta-P from the node to the collocate, the more robust the spreading activation in the mental lexicon is as there is a strong positive correlation between the AM and the participants' lexical choices.

Although the current research does not follow the standard psycholinguistic research norms and apply an online psycholinguistic methodology (e.g., priming), I speculate that the findings through the corpus and the participants' intuitive responses, which reveal the rate of their familiarity with a word combination may still explain some interesting psycholinguistic facts. Therefore, the discussion here builds on the tentative psycholinguistic evidence with the help of the corpus output and attempts to explain its potential reflections on Turkish language teaching.

Earlier studies (e.g., Baayen et al., 2016; Brysbaert et al., 2018) using psycholinguistic methods have shown that single word frequency can be regarded as an important factor in both L1 and L2 lexical processing and how the mental lexicon is structured. To be more precise, the higher-frequency values single words have, the faster and more accurate participants respond to these lexical items in online experiments. Research studies scrutinizing word combinations also indicate that corpus evidence has the potential to reflect the psycholinguistic representation of collocations in native speakers' mind. As Gablasova et al. (2017) state, we can get direct information from corpora regarding the formulaic patterns, which are produced by native and non-native users. Additionally, Rebuschat and Williams (2012) suggest that corpora can provide us with indirect evidence indicating users' (learners) experience with language use, which seems to influence the representation of language in their mental lexicon. In addition,

there is an agreement among researchers (e.g., Ellis et al., 2015; González Fernández and Schmitt, 2015) regarding the effect of frequency on collocational processing and that both L1 and L2 speakers appear to be sensitive towards the association strength between lexical units.

In light of the earlier claims in the literature and the detected relationship between lexical intuitions and the AMs, the current study gives us hints about how the L1 mental lexicon is shaped and what role collocations might play in its structuring by investigating how strongly certain adjectives trigger their constituent nouns. It is hypothesized that the target adjectives are likely to prime the most conventional and familiar nouns (i.e., with stronger associations) in the L1 lexicon. Priming studies seem to support that assumption of psycholinguistic reality of collocations both in Turkish (Cangır et al., 2017; Öksüz et al., 2020; Cangır and Durrant, 2021) and other languages (Durrant and Doherty, 2010). The psycholinguistic reality of collocations, which is also partly evidenced through the humble findings of this research study, has the potential to guide the applications within the context of language teaching.

From a pedagogical perspective, it seems that native speaker intuitions and corpus output (i.e., Delta-P / word 1 > word 2) strongly correlate, which can also guide the textbook designers. If the aim is to design materials which include authentic texts and functional language to help learners master the language as it is used by native speakers, textbook designers may need to exploit representative corpora and consult L1 user intuitions (or experts like language instructors) as a complementary step. There is also research indicating that we need to be cautious with coursebooks writers' preferences for target vocabulary and that learner and native speaker experience may not be reflected in the coursebooks (Cangır, 2021; Jones and Durrant, 2010). There is also a need for an L2 Turkish learner corpus which can benefit both the students learning Turkish as a second language, teachers, and the decision makers. Using these methods, designers can come up with the most useful words to foreground in their materials. It is commonly accepted that multi-word units, such as collocations, is an important aspect of vocabulary learning, and vocabulary teaching materials need to be adapted or extended to include them (Scott, 2019).

A quick review of the course books commonly used to teach Turkish as a foreign language (e.g., *Yeni* İstanbul, Hitit) reveals that there is almost no emphasis on the teaching of multi-word expressions, such as collocations. Additionally, there are no specific sections exploiting the use of chunks which can guide learners in acquiring the second language more naturally and using it more functionally. Students need to be aware of the existence and importance of multi-word expressions like collocations (Schmitt, 2010). They need guidance to remember that learning words in groups (or in chunks) and making generalisations based on our observations regarding recurrent patterns is a cognitive process we have been through in L1 acquisition and that they as L2 learners of Turkish must go through the same process if they want to acquire the language more naturally and reach a near native-like level. Studies in this regard (e.g., Scott, 2019; Lewis, 1993) state that acquiring words in chunks help L2 users sound more

native-like and helps them enjoy a processing advantage while producing the language, which requires great mental flexibility.

A recent study by Karadağ (2020) suggests that there is limited research exploring the acquisition and teaching of collocations to L2 Turkish users; however, studies investigating the most effective approaches to teaching Turkish as a second language abound. She adds that L2 Turkish learners do not pay conscious attention to the use of collocations and thus they need explicit instruction. She also suggests using corpora for designing materials for teaching Turkish as a foreign language program. The use of corpus to validate intuition and to design in-class activities has also been suggested by (Cangır, 2021). In addition, research (e.g., Çelik, 2011) indicates that learners can also benefit from using a corpus to observe the natural language through concordance lines, have better knowledge of formulas (e.g., collocations), and sound more fluent and natural in their second or non-native language.

Future research may want to exploit online methods emphasizing the processing of collocations both in the Turkish monolingual and Turkish-English bilingual minds and discuss the findings within the scope of the materials written to teach Turkish as a foreign language.

5. Conclusion

If certain word pairs are highly associated in the native mind as evidenced by their vocabulary choices and the output from the corpus and if the native speakers enjoy a processing relief by using these formulas like collocations, my claim is that these word combinations should be targeted in the language teaching materials. That is how second language learners can also acquire the language more effectively and use it more naturally. An effective way to detect the target words and fixed expressions, like collocations to teach in a foreign language context, could be to use a corpus-informed and pedagogically sound method. That is to say, objective and subjective frequency measures can be merged to find the most efficient approach in designing vocabulary teaching materials. This research has attempted to show the correlation between the native speaker intuitions about frequency and the association measures indicating collocational frequency with the help of a reference corpus (TNC) and thus present a corpus-driven and a pedagogically convenient approach to designing materials for the teaching of Turkish as a foreign language course. It is hoped that this study could trigger more research in the field of corpus linguistics for pedagogical purposes in the Turkish context, which will guide decision makers and material designers in Turkiye and other related contexts.

Ethics Committee Approval: This study was approved by the Ankara University Social Sciences Sub-Ethics Committee (31.05.2021 / 181).
Peer-review: Externally peer-reviewed.
Conflict of Interest: The author has no conflict of interest to declare.
Grant Support: The author declared that this study has received no financial support.

References

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

Baayen, R.H., Milin, P., & Ramscar, M. (2016). Frequency in lexical processing. *Aphasiology 30*, 1174–1220. http://dx.doi.org/10.1080/ 02687038.2016.1147767

Bölükbaş Kaya, F., & Yılmaz, M. Y. (Ed.). (2020). *Yeni İstanbul uluslararası öğrenciler için Türkçe öğretim seti* [New Istanbul Turkish for international students course book]. Istanbul, Turkiye: Kültür Sanat Basımevi.

Brysbaert, M., Mandera, P., & Keuleers, E. (2018). The word frequency effect in word processing: An updated review. *Current Directions in Psychological Science*, *27*(1), 45–50.

Cangır, H., Büyükkantarcıoğlu, S. N., & Durrant, P. (2017). Investigating collocational priming in Turkish. *Journal of Language and Linguistic Studies*, *13*(132), 465–486.

Cangır, H. (2018*). Investigating the relationship between L1 and L2 collocational processing in the bilingual mental lexicon* (Doctoral dissertation, Hacettepe University). Retrieved from http://www.openaccess.hacettepe.edu.tr:8080/xmlui/handle/11655/4986

Cangır, H., & Durrant. P. (2021). Collocational networks in the L1 Turkish-L2 English mental lexicon. *Lingua, 258*. https://doi.org/10.1016/j.lingua.2021.103057

Cangır, H. (2021). Objective and subjective frequency. *Pedagogical Linguistics*, *2*(1), 64–91.

Council of Europe (1996). The Common European framework of reference for languages: Learning, teaching and assessment. http://www.coe.int/t/dg4/linguistic/Source /Framework_EN.pdf. Accessed 1 February 2021.

Chen, X. & Dong, Y. (2019). Evaluating objective and subjective frequency measures in L2 lexical processing. *Lingua*, *230*, 102738. https://doi.org/10.1016/j.lingua.2019.102738

Christiansen, M. H., Conway, C. M., & Curtin, S. (2005). Multiple-cue integration in language acquisition: A connectionist model of speech segmentation and rule-like behavior. In J.W. Minett & W.S.-Y. Wang (Eds.), *Language acquisition, change and emergence: Essays in evolutionary linguistics* (pp. 205-249). Hong Kong: City University of Hong Kong Press.

Christiansen, M. H., & Arnon, I. (2017). More than words: The role of multiword sequences in language learning and use. *Topics in Cognitive Science*, 1–10. https://doi.org/10.1111/tops.12274

Church, K. W., & Hanks, P. (1990). Word association norms, mutual information, and lexicography. *Computational Linguistics*, *16*(1), 22–29.

Çelik, S. (2011). Developing collocational competence through web based concordance activities. *Novitas-ROYAL (Research on Youth and Language)*, *5*(2). Retrieved from https://dergipark.org.tr/en/pub/novroy/issue/10814/130373

Durrant, P., & Doherty, A. (2010). Are high-frequency collocations psychologically real? Investigating the thesis of collocational priming. *Corpus Linguistics and Linguistic Theory*, *6*(2), 125--155.

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

Ellis, N. C., Simpson-Vlach, R., Römer, U., Brook O'Donnell, M., & Wulff, S. (2015). Learner corpora and formulaic language in second language acquisition research. *Cambridge Handbook of Learner Corpus Research*, 357–378. https://doi.org/10.1017/CBO9781139649414.016

Gablasova, D., Brezina, V., & McEnery, T. (2017). Collocations in corpus-based language learning research: Identifying, comparing, and interpreting the evidence. *Language Learning*. *67*(1), 155–179. https://doi.org/10.1111/lang.12225

González Fernández, B., & Schmitt, N. (2015). How much collocation knowledge do L2 learners have? The effects of frequency and amount of exposure. *ITL International Journal of Applied Linguistics*, *166*, 94–126.

Göymen, D., & Aygüneş, M. (2020). The processing of speech formulas on Turkish: A masked priming study. *Dilbilim Araştırmaları Dergisi*, *31*(2), 207-230.

Gries, S. T. (2013). 50-something years of work on collocations: What is or should be next ... *International Journal of Corpus Linguistics*, *18*(1), 137–166. https://doi.org/10.1075/ijcl.18.1.09gri

Herbst, T. (1996). What are collocations: sandy beeches or false teeth? *English Studies, 77*(4), 379-393.

Hoey, M. (2005). *Lexical Priming: A New Theory of Words and Language* (1st ed.). Routledge. https://doi.org/10.4324/9780203327630

The jamovi project (2021). *jamovi* (Version 1.6) [Computer Software]. Retrieved from https://www.jamovi.org

Jones, M., & Durrant, P. (2010). *What can a corpus tell us about vocabulary teaching materials?* In M. McCarthy and A. O'Keeffe (Eds.) The Routledge Handbook of Corpus Linguistics: London: Routledge: 387–400. https://doi.org/10.4324/9780203856949-28

Karadağ, B. M. (2020). Yabancılara Türkçe öğretiminde eşdizimlilik: Bir model önerisi. *Journal of Linguistics*, *34*, 27–56. https://doi.org/10.26650/jol.2020.003

Kemmer, S., & Barlow, M. (2000). Introduction: A usage-based conception of language. In Barlow, M. & Kemmer, S. (Eds.), *Usage-based models of language* (pp. vii–xxviii). Stanford, CA: CSLI

Lewis, M. (1993). *The Lexical Approach*. Hove: Language Teaching Publications.

Myers, R. (1990). *Classical and modern regression with applications* (2 ed.). Duxbury, Boston.

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

Öksüz, D. C. (2019). *Collocational processing in typologically different languages, English and Turkish: Evidence from corpora and psycholinguistic experimentation* (Doctoral dissertation). Lancaster University Retrieved from https://eprints.lancs.ac.uk/id/eprint/139672

R Core Team (2021). *R: A Language and environment for statistical computing*. (Version 4.0) [Computer software]. Retrieved from https://cran.r-project.org. (R packages retrieved from MRAN snapshot 2021-04-01).

Rebuschat, P., & Williams, J. N. (Eds.). (2012). *Statistical learning and language acquisition*. Berlin: deGruyter Mouton.

Schmitt, N., & Dunham, B. (1999). Exploring native and non-native intuitions of word frequency. *Second Language Research*, *15*, 389–411.

Schmitt, N. (2010). *Researching vocabulary: A vocabulary research manual*. London, UK: Palgrave Press.

Scott, T. (2019). *Learning language in chunks*. Part of the Cambridge Papers in ELT series. Cambridge: Cambridge University Press.

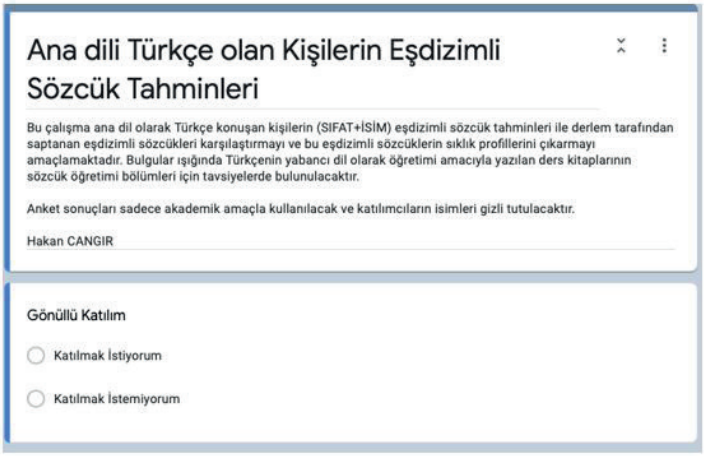
Siyanova, A., & Schmitt, N. (2008). L2 learner production and processing of collocation: A multi-study perspective. *Canadian Modern Language Review*, *64*, 429–458. https://doi.org/10.3138/cmlr.64.3.429

Siyanova-Chanturia, A., & Spina, S. (2015). Investigation of native speaker and second language learner intuition of collocation frequency. *Language Learning*, *65*(3), 533–562. https://doi.org/10.1111/lang.12125

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

Wray, A. (2002). *Formulaic language and the lexicon*. Cambridge, UK: Cambridge University Press.

Appendix A
Questionnaire Sample Screen


Ana dili Türkçe olan Kişilerin Eşdizimli Sözcük Tahminleri

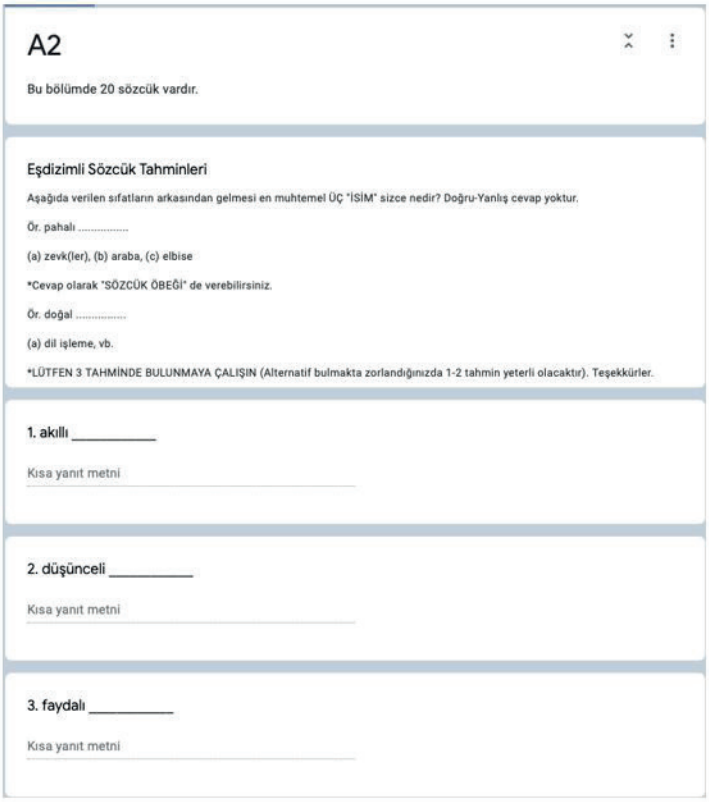
Bu çalışma ana dil olarak Türkçe konuşan kişilerin (SIFAT+İSİM) eşdizimli sözcük tahminleri ile derlem tarafından saptanan eşdizimli sözcükleri karşılaştırmayı ve bu eşdizimli sözcüklerin sıklık profillerini çıkarmayı amaçlamaktadır. Bulgular ışığında Türkçenin yabancı dil olarak öğretimi amacıyla yazılan ders kitaplarının sözcük öğretimi bölümleri için tavsiyelerde bulunulacaktır.

Anket sonuçları sadece akademik amaçla kullanılacak ve katılımcıların isimleri gizli tutulacaktır.

Hakan CANGIR

Gönüllü Katılım

- Katılmak İstiyorum
- Katılmak İstemiyorum


A2

Bu bölümde 20 sözcük vardır.

Eşdizimli Sözcük Tahminleri

Aşağıda verilen sıfatların arkasından gelmesi en muhtemel ÜÇ "İSİM" sizce nedir? Doğru-Yanlış cevap yoktur.

Ör. pahalı .....

(a) zevk(ler), (b) araba, (c) elbise

*Cevap olarak "SÖZCÜK ÖBEĞİ" de verebilirsiniz.

Ör. doğal .....

(a) dil işleme, vb.

*LÜTFEN 3 TAHMİNDE BULUNMAYA ÇALIŞIN (Alternatif bulmakta zorlandığınızda 1-2 tahmin yeterli olacaktır). Teşekkürler.

1. akıllı _____

Kısa yanıt metni

2. düşünceli _____

Kısa yanıt metni

3. faydalı _____

Kısa yanıt metni

Appendix B

w1	*w2*	*t-score*	*MI*	*MI2*	*Delta-P (word 1 > word 2)*	*Delta-P (word 2 > word 1)*
akıllı	telefon	3,346771	4,883826	8,468789	0,004403402	0,001482212
akıllı	çocuk	7,454022	5,557143	11,2989	0,021561371	0,002395391
düşünceli	insan	3,151469	4,32782	9,161716	0,017596573	0,000223865
düşünceli	davranış	1,974864	6,314117	9,194945	0,006649453	0,000920975
faydalı	bilgi	4,73878	4,934536	10,82024	0,012515382	0,001083206
faydalı	besin	1,958473	5,589818	10,85104	0,002111639	0,001726365
güvenilir	insan	4,556395	3,837948	10,79275	0,012265102	0,000478084
güvenilir	arkadaş	2,161691	4,909959	10,82192	0,002655968	0,001043863
huzurlu	ortam	4,785498	8,858314	9,630591	0,027387616	0,007658258
huzurlu	ev	2,39167	5,40478	9,690317	0,006991013	0,000684105
lezzetli	yemek	5,726397	8,304824	8,922066	0,059919857	0,00341554
lezzetli	tarif	1,718748	7,024558	9,084912	0,005422569	0,001399663
neşeli	çocuk	4,332998	5,006371	10,33417	0,014570122	0,000817664
neşeli	günler	3,57442	6,855708	10,34899	0,009690295	0,003013524
cömert	insan	3,047092	4,778927	8,553051	0,024394494	0,000206378
cömert	davranış	1,71275	6,48769	8,603833	0,007510363	0,000691728
fırsatçı	insan	1,945132	5,18789	6,797493	0,03269137	8,33E-05
fırsatçı	esnaf	0,997426	8,602012	6,8705	0,008381754	0,00091008
konforlu	ev	1,985251	7,083268	6,872283	0,031764099	0,000332088
konforlu	araba	1,731196	10,98522	6,895707	0,02398822	0,004997593
sıradan	gün	6,315349	3,665446	11,69055	0,012753682	0,000783881
sıradan	insan	11,21663	5,397906	11,6161	0,037961098	0,00276011
ölümsüz	eser	3,311372	9,302447	8,940321	0,021408752	0,006381727
ölümsüz	aşk	3,719343	7,389573	8,923212	0,027127971	0,001687533
sihirli	değnek	6,855324	14,33928	9,593501	0,054396451	0,353367337
sihirli	dokunuş	1,731824	12,90137	9,744852	0,003471828	0,130417793
bilimsel	çalışma	10,12994	5,957421	12,62511	0,015983385	0,007875097
bilimsel	makale	5,817282	8,736546	12,65693	0,005198385	0,054799198
günlük	süt	6,007424	6,335188	12,88104	0,004798113	0,011985546
günlük	yaşam	12,43331	6,524992	12,83485	0,020520914	0,01369422
karanlık	oda	5,709253	7,345977	12,29732	0,006432724	0,01626908
karanlık	sokak	5,130143	6,298618	12,30078	0,005228414	0,007820122
polisiye	roman	6,554306	11,03149	8,013853	0,126783908	0,013996113
polisiye	film	2,983339	7,492367	8,327716	0,026401409	0,001197984
yöresel	yemek	3,159465	10,13466	8,239927	0,031028472	0,0071367

yöresel	kıyafet	2,998899	11,4115	8,249136	0,027940227	0,017301516
yüksek	lisans	23,19516	9,086543	14,89979	0,017046792	0,338795501
yüksek	bina	4,637861	4,229704	14,94729	0,000718575	0,011089167
yoğun	bakım	12,30668	9,120501	12,95257	0,01844331	0,090209245
yoğun	trafik	8,002931	7,086362	12,98353	0,007842996	0,021902731
büyüleyici	manzara	1,410307	8,499939	8,224092	0,006582467	0,002158563
büyüleyici	güzellik	1,406426	7,504631	8,224118	0,006564352	0,001079837
endişeli	bekleyiş	3,604977	12,61628	8,187525	0,041003189	0,039268926
endişeli	bakış	2,825424	9,879255	8,234626	0,025209954	0,005884919
masrafsız	araba	0,998577	9,45727	4,325297	0,045389902	0,000304743
masrafsız	kredi	1,730479	10,10591	4,036602	0,136239954	0,000478018
yapışık	ikiz	2,827679	11,88408	7,334171	0,04518594	0,013197985
yapışık	ev	1,729401	9,352469	7,418319	0,016923283	0,002279672
asgari	ücret	19,04486	11,27286	10,07844	0,209627782	0,084489884
asgari	müşterek	2,820119	8,411207	10,74404	0,004608188	0,011593908
dolgun	maaş	2,643384	10,1265	8,028923	0,025431912	0,006060569
dolgun	dudak	3,315367	11,36417	7,985522	0,039985043	0,014299082
elverişsiz	hava	2,437842	7,716323	6,79855	0,048548652	0,000508112
elverişsiz	toprak	1,397402	6,39437	6,895491	0,016066903	0,000201777
özenli	iş	0,792408	2,26818	8,290985	0,002515598	2,37E-05
özenli	çalışma	1,685241	5,209518	8,271968	0,009266478	0,000223833
zahmetli	iş	4,766287	7,342757	7,422683	0,106318174	0,000684873
zahmetli	yemek	1,38621	5,658235	7,721492	0,009118163	0,000210044
verimli	toprak	6,549674	6,310487	11,4415	0,015149261	0,004435894
verimli	çalışma	9,139476	6,847325	11,39939	0,029381652	0,006461477
akılcı	çözüm	3,847786	7,264012	9,460251	0,020303305	0,00221199
akılcı	yaklaşım	4,347628	8,595233	9,444059	0,025819	0,005587306
yaygın	inanış	3,156832	9,181531	12,49186	0,001727321	0,06611131
yaygın	kullanım	7,094347	7,244913	12,47137	0,008766358	0,017186811
şaşırtıcı	sonuç	6,03867	7,108033	10,35672	0,026541037	0,003740852
şaşırtıcı	cevap	0,728324	1,880043	10,43283	0,00052626	7,32E-05
gerçekçi	yaklaşım	4,767305	7,393324	10,96462	0,011185965	0,006740779
gerçekçi	insan	4,058718	3,435291	10,97011	0,008880563	0,000396326
yorucu	iş	3,52634	5,508375	8,639041	0,029987045	0,000372714
yorucu	yolculuk	3,157158	9,270752	8,659109	0,023546921	0,005159789
talihsiz	kaza	3,457989	9,146509	8,798722	0,025541443	0,005235737
talihsiz	açıklama	2,436108	7,516032	8,836393	0,012723403	0,001684821
bağımsız	ülke	7,17001	4,912846	12,73687	0,007669671	0,003985928

bağımsız	devlet	14,64672	5,704029	12,66604	0,031547754	0,007000071
çarpıcı	sonuç	5,416615	6,49773	10,67759	0,017483215	0,002993124
çarpıcı	gerçek	3,124734	4,111355	10,71055	0,006107199	0,000545483
karşılıklı	anlayış	7,254731	8,1642	12,08375	0,011880314	0,025079504
karşılıklı	konuşma	6,499295	6,817362	12,09037	0,009586699	0,009807267
ikiyüzlü	insan	3,10104	5,69039	7,574798	0,046697089	0,000210031
ikiyüzlü	davranış	1,72179	7,399153	7,672851	0,014201141	0,000695379

Appendix C
Descriptive plots