Collocational Links in the Multilingual Mind: Frequency Intuitions and Association Strength Measures

Hakan Cangır[1], Cumhur Kuzu[2]

ORCID: [1]0000-0003-2589-2466, [2]0000-0002-6291-4807

[1]Ankara Üniversitesi, Yabancı Diller Yüksekokulu, Gölbaşı, 06830, Ankara

[2] Ankara Üniversitesi, Yabancı Diller Yüksekokulu, Gölbaşı, 06830, Ankara

[1]hcangir@ankara.edu.tr, [2]ckuzu@ankara.edu.tr

(Received 6 March 2022; Accepted 31 October 2022)

ABSTRACT: Collocations are ubiquitous in everyday language, which help language users sound more natural and ease the processing burden during language production. They are also commonly taught in foreign language classrooms and those learners who can master these prevalent word combinations and chunks get to use the language more fluently as they enjoy a processing advantage. There are different approaches to the definition of these frequently co-occurring word combinations, but the current study embraces the frequentist approach which adopts a corpus linguistic technique and tries to combine corpus linguistics and user intuition data in its interpretations. Collocations have been extensively investigated from many different angles using both corpus linguistics and psycholinguistics techniques so far. Having its spark in the theory of collocational priming, the present study attempts to (a) investigate the relationship between corpus-driven association measures and the collocational frequency judgements of L1 Turkish-L2 English-L3 Italian users, (b) analyse the possible effect of congruence across the languages on the frequency relationship, (c) question if there is a difference between L1 Italian users' and L3 Italian users' sensitivity towards collocational strength based on their collocational frequency intuitions in Italian and (d) find humble evidence for the psycholinguistic reality of collocations in the L3 Italian mental lexicon, if any. The results indicate that language users' frequency judgements in L2 English and L3 Italian correlate with an association measure (t-score) and the frequency of congruent items are judged more accurately. Additionally, L3 Italian users' frequency intuitions overlap with native Italian users to a certain

http://dx.doi.org/10.18492/dad.1083480

Dilbilim Araştırmaları Dergisi, 2022/2, 191-219

extent. The findings are discussed in the light of the earlier research on usage-based language use and frequency intuition.

Keywords: collocations, priming, frequency, multilingual

Çokdilli Zihinde Eşdizimli Sözcükler Arası Etkileşim: Kullanıcı Sıklık Tahminleri ve Derlem Tabanlı Sıklık

ÖZ: Eşdizimli sözcükler günlük dilde çok sık kullanılır ve dili kullananlar bu sayede kulağa daha doğal gelen cümleler kurar. Bu sözcükler yabancı dil öğretimi derslerinde de genellikle öğretilir ve zihinsel işlemleme daha az olduğu için bu sözcük öbeklerini iyi kullanan öğrenciler dili daha akıcı bir şekilde kullanabilirler. Bu sözcük öbekleri farklı şekillerde tanımlanmaktadır, ancak mevcut çalışma sıklık odaklı tanımlamayı benimsemekte ve analizine derlembilim yöntemlerini ve anadil ve yabancı dil kullanıcısı tahminlerini dahil etmektedir. Eşdizimli sözcükler daha önce derlembilim ve ruhdilbilim yöntemleri harmanlanarak farklı açılardan incelenmiştir. Çıkış noktası eşdizimli sözcüklerde önceleme teorisi olan bu çalışma (a) derlem tabanlı eşdizimli sözcük sıklık profilleri ile D1 Türkçe-D2 İngilizce ve D3 İtalyanca kullananların sözcük sıklığı tahminleri arasındaki ilişkiyi incelemeyi, (b) diller arası benzerliğin sıklık ilişkileri arasındaki etkisini saptamayı, (c) İtalyancayı D1 olarak konuşan kişiler ile D3 olarak kullanan kişilerin eşdizimli sözcüklerin sıklık profillerine olan duyarlılığı arasında bir fark olup olmadığını sorgulamayı ve (d) eşdizimli sözcüklerin D3 olarak İtalyanca kullanan öğrencilerin zihin sözlüğündeki yerini ruhdilbilimsel açıdan açıklamayı hedeflemektedir. Bulgular, D2 İngilizce – D3 İtalyanca kullanan kişilerin sözcük sıklık tahminleri ile derlem sıklık profilleri (t-değeri) arasında belli bir oranda örtüşme olduğunu göstermektedir. Ayrıca, D3 İtalyanca kullanan kişiler ile anadil olarak İtalyanca kullanan kişilerin eşdizimli sözcük sıklığı tahminleri belli oranda benzeşmektedir. Araştırma sonuçları, kullanıcı temelli dil kullanımı ve sıklık tahminleri alanyazını ışığında tartışılmıştır.

Anahtar Sözcükler: eşdizimli sözcükler, önceleme, sıklık, çokdillilik

1 Introduction

Studies to date have stated that collocations are important components of written and spoken production in L1 and L2 and a significant amount of L1 production consists of these formulas to make speech more fluent and writing more coherent and natural (Schmitt, 2010). It is also claimed that advanced L2 users enjoy a processing relief with the help of collocations, and they produce near native-like utterances. The association strength between the component parts of a collocation is an important factor for L2 users because their frequency of use in the vicinity of one another helps language users observe this pattern

thoroughly and internalize the combination more easily, which is in line with usage-based approaches to language use (Bybee, 2010). The fact that they tend to appear in each other's company also seems to entrench them more strongly in the non-native lexicon, which eases automatic production (Hoey, 2005). Ellis (2001) similarly asserts that word strings which language users frequently encounter are likely to form chunks in long-term memory. Studies (e.g., Wolter & Gyllstad, 2013) also indicate that if those word combinations are congruent in L1 and L2 (i.e, if they have a word-for-word overlap in the two languages), they are processed more easily and learned earlier than incongruent items. Hoey (2005) acknowledges the importance of collocations in language production. In his theory of collocational priming, he explains based on corpus evidence that each word constituting a node is likely to trigger its collocate (as in *heavy* primes *rain*) in the mental lexicon, and this may be the source of creative language systems. Hoey's collocational priming theory has its roots in corpus linguistics, but his claims have also been supported by psycholinguistic evidence.

To date, there has been attempts to test the existence of collocational priming in L1 English (Durrant & Doherty, 2010), L2 English (Öksüz et al., 2021), L1 Turkish (Cangır et al., 2017), L1 and L2 Italian (Lenci et al., 2021). These studies state that there are certain fundamental factors affecting the priming process. The frequency effect is one of the conclusive results in earlier research, which roughly indicates that more frequent word combinations are primed faster than less frequent ones (e.g., Wolter & Gyllstad, 2013; Wolter & Yamashita, 2018). Another variable which is claimed to play a critical role in the priming process is the congruence between the first and the second language. That is, word combinations which have one-to-one literal translations in the L1 and L2 are processed faster than non-congruent collocations (Wolter & Yamashita, 2015).

There are also some studies which are primed by the collocational priming theory and explore the relationship between the collocational familiarity of language users by asking their frequency intuitions and corpus-driven objective frequency measures (i.e., association strength measures) to test the psycholinguistic reality of collocations from a different angle. Research in this regard state that there is a certain overlap between the L1 users' frequency intuition and corpora output. For instance, Hoffmann and Lehmann (2000) state that native speakers are good at detecting collocations represented in corpora. In addition, those studies focusing on L2 users also suggest that L2 English users' intuition for high frequency words seems to correlate positively with corpus-driven frequencies (Chen & Dong, 2019). Siyanova and Schmitt (2008) assert that exposure to L2 in a native setting can make a difference and have an impact on learner intuitions. Hoffmann and Lehmann (2000) argue that learner frequency intuitions are worth taking into account, although L1 users are more competent guessers. According to Siyanova-Chanturia and Spina (2015), L2

Italian users' frequency intuitions correlate significantly with native speaker intuitions, particularly for high frequency collocations. Fioravanti et al., (2021), on the other hand, concluded that native speakers of Italian performed better than L2 Italian learners at lexical acceptability judgement tasks, where participants were asked to judge word combinations with different semantic compositionality (i.e., free combinations, collocations and idioms) in terms of their fixedness. Although these studies do not claim that they use the priming methodology and their findings could be attributed to a lexical priming effect, we believe the theory of collocational priming, which suggests that words are primed to co-occur and that their frequency of use could affect the way they are stored and retrieved, and the subjective collocational frequency judgements of language users can intersect at certain points. In other words, if language users store frequency information of single words and word combinations in their lexicon and can intuitively differentiate between a frequent and infrequent lexical unit, we can speculate that those users are likely to respond to a frequent collocational item faster than a less frequent or a non-collocational item in a lexical decision task.

Those studies approaching the theory of collocational priming from a pedagogical angle refer to Lexical Teaching Approach (Lewis, 1993) while drawing language teaching related conclusions. In his approach, Lewis states that lexical chunks (e.g., collocations) appear to help learners in the acquisition process. Fluency is not achieved through a set of generative grammar rules and a string of isolated words, but by promptly accessing this inventory of chunks that are readily available to native speakers. That is why, language learning should mainly focus on mastering these chunks to sound more native-like and be more fluent in language production (Lewis, 1997).

Sparked by the *Collocational Priming Theory* (Hoey, 2005), which partly suggests that collocations are stored as single units in the lexicon resulting in a spreading activation once a component part is triggered, and more importantly in light of the findings of the frequency intuition research, we hypothesize that if component parts of a collocation are highly associated (i.e., more frequently used in each other's company), which can be evidenced based on representative corpora, they will be more robust in the mental lexicon (i.e., hypothetically more likely to prime one another). Language users, thus, can intuitively guess the word combinations' frequency more accurately, as they are exposed to these collocations more often and these lexical items are more readily available to retrieve and use in language production. In other words, the frequency of use can influence the intuitive frequency ratings of language users, which could tentatively indicate to what extent the collocational items are likely to be primed in the language users' mental lexicon. It should be noted that our study does not follow the priming paradigm standards and thus we do not argue that our findings will indicate a possible priming (or inhibition) effect, but rather show language

users' frequency intuitions for co-occurring lexical units, which could indicate how those users store word combinations.

Additionally, given the syntactic and morphological differences between Turkish (a non-Indo-European language) and English & Italian (two branches of Indo-European languages), we believe that the relationship between the language users' frequency intuition and association strength measures and the possible effect of the confounding factors (i.e., congruence) are worth investigating. This research is yet another attempt to investigate the possible relationship between subjective frequency intuitions of language users and the objective corpus frequency measures. However, to our knowledge, no research study has attempted to explore this relationship for L3 Italian users and scrutinize the effect of collocational overlap in Italian, English and Turkish.

2 Current Research

We assume based on earlier research that L1 users are sensitive towards word level and collocational frequency, as is represented in reference corpora (e.g., Cangır, 2021b; Durrant & Doherty, 2010). Research also contends that language users partly transfer their collocational knowledge in L1 to L2, particularly when the items in the two languages are congruent (e.g., Wolter & Gyllstad, 2011). As their proficiency improve, some collocational links between L1 and L2 seem to get more entrenched (Öksüz et al., 2021). If the collocations are non-congruent across L1 and L2, language users are likely to produce new schemas in L2 in their mental lexicon, which we think could feed into their L3 lexical processing particularly if the L2 and L3 are from the same language family and if they share certain linguistic features, such as syntax. That is why, we take the collocational profiles (i.e., association measures) of both English and Italian lexical items into account in our analysis.

Considering the findings of earlier research and our assumptions based on the available theoretical background in frequency intuition, we have the following research questions:

1. Is there a relationship between (a) L1 Turkish-L2 English-L3 Italian users' collocational frequency intuitions in English and Italian and (b) corpus-extracted association strength measures?
2. To what extent does congruence between the languages in question play a role in the process, if any?
3. To what extent is there a relationship between the L1 Italian users' and L3 Italian users' sensitivity to corpus-extracted association strength measures based on their collocational frequency intuitions in Italian?

3 Methodology

3.1 Item Development

The co-author of this article who is an experienced instructor of Italian extracted potential (V+N) collocational items from course books designed for teaching Italian as a foreign language. The collocational items were either represented in the vocabulary section of the textbook, particularly designed for teaching collocations or in the target lists at the end of every unit. The co-author focused only on the items presented as word combinations and did not make intentional choices to combine certain lexical items. V+N combinations were chosen for a reason. These types of collocations are the most commonly taught word combinations in foreign language textbooks. In addition, Turkish word (N+V) order is different from Italian and English word order (V+N), which could influence the processing burden and the familiarity of those items to Turkish native speakers. The extracted items (*N*=135) were compared against an Italian collocation dictionary (Tiberii, 2018). The items which we could validate with the help of the dictionary were kept on our target list. The first author who is an experienced instructor of English and the co-author worked together to find the English translations of the collocations. The items which were not intuitively translatable according to the first author were removed from the list. The chosen collocational items (*N*=120) were checked against a collocational dictionary in English (Macmillan Collocation Dictionary online[1]). The items which were not present in the dictionary were removed from the list (*N*=99). The two authors who are native speakers of Turkish worked together to find the Turkish translations of the target collocations. The authors grouped the target collocations based on their congruence (i.e., items which are congruent across the three languages and items which are congruent in English and Italian but incongruent in Turkish). 62 congruent items across all the languages and 37 incongruent items in Turkish were detected. To double-check the writers' intuition about congruence, three different instructors of Italian who are also advanced L2 English users were consulted. The authors explained the concept of congruence to the raters by giving examples. The items all three raters agreed on for congruence were kept in the list (Congruent=62, Incongruent=37). The repeated verbs (if more than 3) were removed from the list. (Congruent=36, Incongruent=25). As the final step of filtering, the authors exploited the Sketch Engine (itTenTen – Jakubíček et al., 2013) to extract collocational profiles of the target items. Those items with an MI (Mutual Information) score of at least 3.0

[1] See the dictionary link here:
https://www.macmillandictionary.com/browse/collocations/british/a/

and a t-score of at least 2.0 (Congruent=24, Incongruent=19) in Italian were kept in the list based on the recommendations by Schmitt (2010). The Italian Web corpus (itTenTen) is an Italian corpus with texts from the Internet (10+ billion words). Sketch Engine, which is a versatile corpus tool used to explore patterns and emerging usage in a language, provides access to TenTen corpora in many different languages. We used BNC (British National Corpus, 2007), TNC (Turkish National Corpus – Aksan et al., 2012) to extract the association strength measures for the English and Turkish items. To compute the association strength values, we extracted the raw frequency values of the component parts of collocations (i.e., expected frequency, observed frequency and the frequency with which two words co-occur with a span of 4) and took corpus size into account (see Appendix A to check how to compute t-score and MI score). We randomly chose 15 items for each category to have an equal number of collocations. We also paid attention to frequency dispersion. Samples are provided below (see the full list in Appendix B).

Table 1. Sample items

ITALIAN (V+N)	**ENGLISH (V+N)**	**TURKISH (N+V)**
CONGRUENT (same in all languages)		
affittare casa	rent (a) house	ev kirala-
INCONGRUENT (same in Italian and English, different in Turkish)		
correre rischio	run (the) risk	risk al-

Finally, we computed an independent samples T-test to report the difference between the association strength measures for congruent and non-congruent items. We wanted to make sure that the two variables (i.e., congruency and collocational strength) are not confounding to avoid possible misinterpretations. The results of the test can be seen in Table 2 and the descriptive raincloud plots can be seen in Appendix C.

Table 2. Independent Samples T-Test

	t	**df**	**p**
t_score_IT	0.636	28	0.530
MI_IT	1.877	28	0.071
t_score_ENG	-0.421	28	0.677
MI_ENG	0.563	28	0.578

Note. Student's t-test.

3.2 Association Strength Measures (AMs)

We have exploited two commonly used AMs in our study. T-score is claimed to foreground pure frequency and correlate to a certain extent with the raw frequency values of the component parts of a collocation (Durrant & Schmitt, 2009). We used this measure both to detect collocations in Italian during our filtering procedure and as a variable with potential to have an effect on the frequency judgements of our participants. Additionally, we used the MI score which is claimed to signify unique word combinations. They seem to exclusively be in company of one another and their component parts may not be very frequent by themselves. Research shows that these measures appear to complement each other, and they pave the way for a more comprehensive statistical analysis (Schmitt, 2010). Table 3 illustrates the mean association strength measures we have calculated for the three different languages. The AMs of all the collocations under investigation can be seen in Appendix D.

Table 3. AMs of Collocations in Three Languages

	congruence	N	Mean	SE	SD
t_score_IT	Incongruent	15	113.10	14.939	57.86
	Congruent	15	99.71	14.825	57.42
MI_IT	Incongruent	15	7.14	0.608	2.36
	Congruent	15	5.64	0.521	2.02
t_score_EN	Incongruent	15	12.86	2.805	10.87
	Congruent	15	14.59	2.984	11.56
MI_EN	Incongruent	15	2.86	0.479	1.86
	Congruent	15	2.50	0.421	1.63
t_score_TR	Incongruent	15	9.80	1.170	4.53
	Congruent	15	7.33	1.189	4.60
MI_TR	Incongruent	15	9.35	0.842	3.26
	Congruent	15	8.68	0.802	3.11

3.3 Instruments

We designed a simple (6 points – 0-5) Likert-scale questionnaire (i.e., acceptability judgement/rating task) and asked our participants to rate the collocations based on their frequency intuitions in L2 English and L3 Italian (0 indicating infrequent and 5 very frequent). We also used the same questionnaire in Italian and asked our native Italian participants to rate their familiarity with

the collocations under investigation and asked for their native speaker intuition in terms of frequency. A sample questionnaire screen can be seen in Appendix E.

3.4 Participants

79 L2 English - L3 Italian users (aged between 21 and 24) filled in the informed consent form and participated in the English, and Italian frequency judgement questionnaire. The participants in the experimental group, who are native Turkish users, are students in the Italian Language and Literature department at Ankara University (youngest in their 2nd year and oldest in their 4th year). To get accepted to the Italian Language and Literature departments in Turkey, students need to attend the official university entrance exam where they have to prove that they are beyond a certain level of English. If they can pass the entrance exam, they are assumed to be at least B1 level in English and they are accepted to the preparatory classes in the Italian language before they start their major in Italian. Our participants passed the entrance exam, had their preparatory Italian language education for a year and were ready to start taking their departmental classes in Italian. According to the requirements of the language school at Ankara University, they are considered at least B1 level in Italian (according to Common European Framework of Reference for Languages - CEFR) as they successfully achieved their training and passed the Italian language proficiency exam conducted at the end of each academic year. Additionally, 19 native Italians (aged between 30 and 55) took the frequency intuition questionnaire in Italian as the control group. The native participants are experienced instructors of Italian language in Turkey (N=13) and Italy (N=6).

3.5 Procedure

L1 Turkish – L2 English – L3 Italian participants were given the English and Italian version of the questionnaire where they had to judge the frequency of the target collocations intuitively. The co-author of the article explained the task to the attendees and made sure they understood the concept of collocation and frequency of use and co-occurrence. The participants conducted the questionnaire online and were monitored through a virtual classroom (Zoom). They completed the English and Italian version of the questionnaires on separate days and each session took around 10-15 minutes. As the items were taken from a coursebook the participants studied during their Italian preparatory year, we assumed all the items were comprehensible. The native Italian group were given the Italian version of the questionnaire only. We sent the potential participants

an e-mail explaining the purpose of the research and the requirements of the task. They conducted the task on their own through a shared link on Google forms.

The study has been approved by Ankara University Social Sciences Ethics Committee on January 24, 2022 with the decision number (3/33).

3.6 Method of Analysis

- We decided to exclude Turkish AMs from the analysis because nouns in Turkish have case endings and verbs have conjugations and that is why the frequency profiles are influenced by its agglutinating nature. Since we were trying to reach a more standard calculation of association measure, we took the bare forms of the verbs and nouns in Turkish into account and presented them descriptively. To avoid the effect of agglutination on the AMs, we did not include these measures in the final regression model.
- We investigated the effect of congruence and collocational frequency in English and Italian on the intuitive frequency ratings of L2 English and L3 Italian users through a linear regression model.
- We explored the possible predictors of native Italians' collocational frequency judgements with a linear regression model. Our assumption is that association strength measures can explain a certain extent of native speakers' collocational frequency intuitions.
- We further analysed the correlation between (a) the native Italian speakers' and non-native users' frequency judgements, (b) the intuitive frequency ratings and association strength measures in Italian.

4 Results

In this section, we present the results of two separate data sets. The first data set (in 4.1) shows the results of L2 English-L3 Italian participants, and the following data set (in 4.2) displays the output of the native speakers, which we used as the control data.

4.1 Results for L2 English - L3 Italian Users

We ran a linear regression model using jamovi (2021) to investigate the association between the intuitive frequency ratings of our participants and the independent variables; congruence, and AMs. Since we have data both in English and Italian, we initially analyzed the predictive power of AMs in English on the frequency ratings of our L2 English and L3 Italian subjects. By doing so, we

wanted to test the hypothesis that L2 English users' frequency intuition correlate with corpus output to a certain extent (e.g., Cangır, 2021a).

Table 4. Regression for items in English

Model Coefficients - Intuition_ENG

Predictor	**Estimate**	**SE**	**t**	**p**
Intercept [a]	3.4365	0.2553	13.46	< .001
t_score_ENG	0.0378	0.0136	2.78	0.010
MI_ENG	-0.1199	0.0874	-1.37	0.182
congruence:				
congruent – non-congruent	0.5931	0.2360	2.51	0.018

[a] Represents reference level - R: 63, R^2: 39

As is seen in the regression table, congruence plays the most significant role when students respond to the collocations in English, $R^2 = .39$, $F(1,26) = 6.32$, $p < .05$. The congruence variable by itself explains approximately 30% of the variance in the regression model. That is to say, the participants said they were more familiar with the collocations in English if they were congruent with the Turkish collocations although the word order in Turkish is in the opposite direction (N+V). In other words, the subjects rated the congruent collocations considerably more frequent than the incongruent items, which could indicate that those congruent items in Turkish and English are linked more strongly in their mental lexicon. T-score was another significant predictor for intuitive frequency ratings and explained 39% of the variance in combination with the congruence variable, $F(1,26) = 7.73$, $p < .01$. As the numbers indicate a positive association, we can contend that higher AMs (t-score) signify higher ratings by the participants. We couldn't find the same relationship between MI score and frequency judgements, and this finding can be attributed to the English level of our participants. To be more precise, as the MI score favours unique word combinations with infrequent lexical items, our participants might have misjudged the frequency profiles as those items are not represented in their

lexicon as chunks. Since we see the hints of a potential negative association based on the statistical estimate, though non-significant, we may speculate that our participants could rate collocations in English considerably less frequent, particularly if they have much higher MI values than the ones we had in our investigation. We know from earlier research that there is a positive relationship between language proficiency and the use of collocations with strong MI values (e.g., Bestgen, 2017). However, we accept that a more detailed analysis is needed to validate this assumption. Given that there was a positive relationship between the frequency ratings (t-score) and the corpus extracted AMs in English, we went on to test the existence of this type of relationship between the intuitive frequency judgements of the same participants who use Italian as their L3, and the AMs taken from a representative corpus of Italian. Table 5 shows the regression output for the lexical items in Italian.

Table 5. Regression for items in Italian

Model Coefficients - Intuition_IT

Predictor	**Estimate**	**SE**	**t**	**p**
Intercept [a]	3.98708	0.22692	17.57	< .001
congruence:				
congruent – non-congruent	0.68790	0.12185	5.65	< .001
t_score_IT	0.00227	0.00104	2.18	0.030
MI_IT	-0.13654	0.02739	-4.99	< .001

[a] Represents reference level - R: 70, R^2: 49

The results indicate that congruence still has the most significant effect on the frequency ratings of the L3 Italian participants, R^2 =.49, F(1,26) = 9.63, p < .001. In other words, congruence is associated with higher frequency ratings or participants had stronger and more accurate judgements when the collocations were congruent across the languages. This is in line with the output of the earlier analysis with the English items, which could indicate that L2 English and L3 Italian users still have access to their L1 Turkish and transfer their collocational knowledge from their native language as those collocational items are highly connected in their mental lexicon. Besides the congruence effect, the numbers indicate that t-score is again another significant predictor of intuitive frequency judgement, R^2 =.49, F(1,26) = 1.44, p = .03. The participants are sensitive towards more frequent collocational items as represented in the TenTen (Jakubíček et al., 2013) corpus. In other words, they provided higher frequency

ratings for more frequent collocational items. This finding also partly supports earlier studies which claim that L2 users of English can accurately judge the association strength of high frequency collocations (e.g., Chen & Dong, 2019). Another noteworthy finding and possibly a counter-intuitive result is the negative association between the MI score and the intuitive frequency ratings of the participants $R^2 = .49$, $F(1,26) = 7.51$, $p < .001$. To be more precise, the higher the MI scores are, the less powerful intuitive frequency ratings collocations have. We attribute this finding to the fact that language users lower than a certain level are less sensitive towards more unique word combinations and that these exclusive word combinations are not entrenched in the participants' L3 Italian lexicon even when they are congruent with the native language.

To support our data visually, we also present the dispersion of the frequency ratings of our participants in English and Italian and illustrate the effect of congruence in each language.

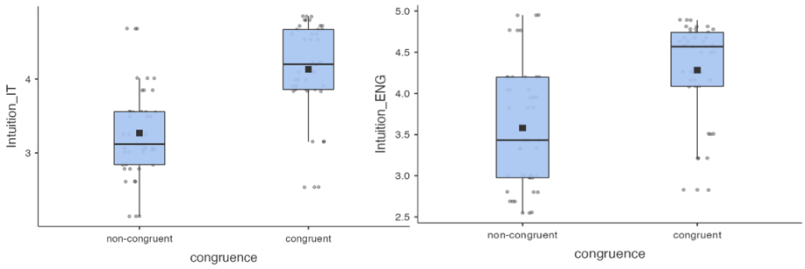


Figure 1. Dispersion of Intuitive Frequency Ratings

The figures indicate that the participants rated the congruent items more frequent than the incongruent items for each language. Additionally, the standard deviation is particularly lower for the congruent items, which suggests that participants tended to agree on the intuitive frequency ratings of congruent items. Additionally, the range of the frequency ratings in English seems to be larger than the ratings in Italian. As the starting point of the item development phase was the Italian collocations, which were extracted from a coursebook designed for teaching Italian as a foreign language, the frequency profiles of the target items in Italian could be more balanced, which in turn could arguably have affected the L3 Italian users' intuitive frequency ratings. It should also be noted

that the participants studied this coursebook in their preparatory year, which could have also influenced the way they have responded to the questionnaire.

4.2 Results for Native Italians (control group)

In an attempt to detect the possible predictors of collocational frequency intuitions of native Italians and explore the possible relationship between native speaker and non-native speaker frequency intuitions, we ran another linear regression model and a follow-up correlational analysis. The dependent variable was the average frequency ratings of the participants for each Italian collocation under investigation and the independent variables were the association strength measures (t-score and MI value) for the lexical items in Italian only. Table 6 presents the variables under investigation and their influence on the frequency judgements of native Italian participants.

Table 6. Regression for Native Italian Intuition

Model Coefficients - Intuition_Native

Predictor	**Estimate**	**SE**	**t**	**p**
Intercept [a]	3.71036	0.44752	8.29	< .001
t_score_IT	0.00311	0.00245	1.27	0.032
MI_IT	0.06329	0.06105	1.04	0.309

[a] R .31 / R^2 .09

The table indicates that t-score can partly predict native speakers' frequency intuitions ($F(2,27) = 1.14$, $p < .05$), with an R^2 of .09. As our variables can explain a small proportion of the variance, some other confounding factors may be playing a role. Other more advanced association measures which could reveal some bidirectional lexical relationships (e.g., delta-P) can show us a different picture given the nature of native speaker mental lexicon and its organization, as stated in earlier studies (Cangır, 2021a; Gries, 2013). Another noteworthy finding in this table can be seen as the lack of MI score effect on frequency judgement of the users. One could assume that native speaker frequency judgements should match with any type of collocational frequency value. However, we should note that MI score favors unique word combinations, component parts of which are not very frequent, but almost always in the near vicinity of one another. We might thus speculate that those rare word

combinations with high MI scores are not labelled as frequent collocations as each unit by itself is not commonly and frequently used in language. The fact that MI score cannot predict native speakers' frequency intuition can also be attributed to the source of the chosen collocations and their frequency dispersion. A more controlled and a more comprehensive list of items (i.e., with items extracted not only from course books) may yield more reliable and different results indicating the MI score as a potential significant predictor.

We also conducted a further correlational analysis to see the relationship between the intuitive frequency judgements of both native Italians and non-native users.

Table 7. Correlation Result for Intuitions

		Intuition_Native		**Intuition_IT**
Intuition_Native	Pearson's r	—		
	p-value	—		
Intuition_IT	Pearson's r	0.726	***	—
	p-value	< .001		—

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

The numbers indicate that there is a strong positive correlation between native and non-native users' judgement of collocation frequency.

5 Discussion

According to the usage-based approaches to language, multiword expressions, such as collocations are acquired as patterns of language from language usage. Our knowledge of these patterns seems to ease language processing and underlies fluent language production (Ellis & Ogden, 2017). Ellis (2012) also holds that sufficient exposure to these expressions (e.g., collocations) in L2 is likely to encourage automatic access to them in the mental lexicon and

their frequency of use in different contexts make them more entrenched in the lexicon.

Earlier research indicates that native speakers and L2 users are sensitive towards high frequency lexical items and claim that collocations may have psycholinguistic reality. Priming studies show that collocational priming seems to exist in L1 Turkish (Cangır et al., 2017), L2 English (Öksüz et al., 2021), and L2 Italian (Lenci et al., 2021), which signals how these word combinations are represented and processed as chunks or how the components parts trigger one another in the mental lexicon. These studies also assert that frequency profiles of lexical items and their congruence across the languages significantly influence collocational processing. Other studies investigating the relationship between learners' acceptability judgements and collocational frequency ratings indicate that L1 users' frequency ratings correlate with representative corpus output to a certain extent (Siyanova & Schmitt, 2008). Some also claim that L2 users' frequency judgements tend to correlate positively with objective corpus frequency (e.g., Cangır, 2021a). However, no study to date has explored the collocational frequency judgements of L2 English and L3 Italian users.

Based on our findings, it can be claimed that L1 is still playing a significant role in how collocations are perceived in L2 and L3 as the evidence in our research suggests that congruent collocations are rated more frequent by L2 English and L3 Italian users, who are native Turkish users. This could be a clear indication that those collocations are linked in their mental lexicon more strongly and that the subjects transfer their L1 Turkish experience into their L2 and L3 learning and use. The findings could also indicate that L2 English - L3 Italian users tend to transfer lexical information from their L1 and thus are more inclined to acquire the congruent vocabulary first. In other words, those word combinations that are congruent across the languages are more simply entrenched in the L3 Italian mental lexicon and thus users are capable of retrieving those lexical items more easily, which make them more familiar, salient, and frequent for L3 Italian users. The findings regarding L1 effect is in line with earlier research indicating that congruent collocations are processed faster and appear to be more robust in the L2 mental lexicon (Wolter & Gyllstad, 2011, 2013; Wolter & Yamashita, 2018). We may also observe some research findings exploiting the cross-linguistic collocational priming paradigm, which partly supports the claims here. For instance, Cangır and Durrant (2021c) found evidence for cross-linguistic collocational priming for congruent adj+noun collocations in Turkish and English, particularly in L1-L2 direction. The connectionist account of language processing can help us explain this cross-linguistic spreading activation since lexical nodes in different languages are connected to a certain extent, which could result in retrieving word combinations as chunks and help users intuitively label them as more frequent. Supporters of connectionism (e.g., Christiansen & Chater, 2001) assert that the

interconnections in the brain store lexical information in the form of a network. The representation of words or word clusters involve interconnections between different but interrelated modules, e.g., phonological, semantic, or orthographic sections. The meaning of a word and the company it keeps is formed based on the convergence of these networks. Thus, a well-known word and its syntagmatic and paradigmatic association with other words require a complex network of interconnections. In a lexical network, the nodes are systematically structured in such a way that any node can trigger or inhibit other nodes at various levels due to numerous factors and this interaction can also be cross-linguistic to a certain extent (Ellis, 2003).

Another noteworthy finding was the strong relationship between the intuitive frequency ratings of native and non-native users of Italian. We believe it may stem from the fact that the word combinations in our study were extracted from language teaching materials and the native Italian participants are dominantly instructors of Italian as a foreign language. To put it differently, the native speakers in our study are mostly teachers of Italian who teach Italian as a foreign language and thus are also familiar with the collocational patterns coursebooks mainly integrate in their vocabulary teaching sections. Language learners and teachers' language experience seem to be similar due to the same language they are exposed to in the language classroom through the semi-authentic language learning materials. This could affect the language context the participants are in and thus their lexical background, which is commonly shared in the language classroom. There are also some studies in the literature which conclude that (particularly) advanced users of a language can be as sensitive towards lexical frequency as native speakers and that native and non-native users' lexical frequency intuitions may correlate to a certain extent especially for high frequency word pairs (Siyanova-Chanturia & Spina, 2015). We can also assert based on our findings and the similar claims of the related research that our findings are compatible with usage-based language learning accounts (Bybee, 2010; Ellis, 2019).

We found a significant correlation between t-score and the intuitive frequency ratings. We can argue that non-native users' experience with the collocations and their familiarity with these frequent word combinations may partly correspond to a representative corpus. One could also claim that learners' main exposure to collocation is through semi-authentic coursebooks particularly designed for teaching languages to foreigners and that is why it may fail to reflect the natural language native speakers would use in authentic contexts. Representative corpora, on the other hand, consist of texts from a native environment and may not perfectly match with learners' vocabulary knowledge, which they build mainly with the help of material designed for foreign language learning. However, since we have found that an association strength measure (t-score) has a partial effect on language users' frequency intuitions, we conclude

that there may be a certain overlap between learners' lexical background formed mainly through coursebooks and the language use represented in corpora. As Bod (2006) states, language should be regarded as a statistical accumulation of experiences that changes every time a particular utterance is encountered. The frequency of these encounters seems to shape the mental lexicon of both L1 and L2 users.

Our results also showed that MI score had an important impact on the participants' frequency judgements; however, the detected correlation was negative for the L2 English – L3 Italian group. The higher the MI score, the less frequent a collocation was rated in our study. In other words, our participants' judgements for collocations with higher MI score were not as accurate as for the items with a lower MI score. Studies in the literature exploiting MI score as a variable state that only advanced non-native users are sensitive towards unique word combinations and the use of collocations with a high MI score in language production is claimed to be an indication of higher proficiency (e.g., Kyle & Crossley, 2016). Since our participants are not advanced L3 Italian users, we may tentatively claim that they may be less familiar with the items that have high MI scores (i.e., more unique word combinations).

Previous research has helped us acknowledge the psycholinguistic reality of collocations in the second language users' mental lexicon. There are some studies revealing a relationship between corpus-extracted frequency values and both L1 (e.g., Durrant & Doherty, 2010) and L2 users' frequency intuitions (e.g., Cangır, 2021a). We can deduce based on these findings that statistical frequency as evidenced in corpora could be associated with the user experience to a certain extent, which should be represented in language teaching material. Given the research results stating that high-frequency collocations tend to be represented as chunks in the competent users' mind, which could trigger accurate frequency judgements, we can talk of the importance of instruction for multi-word units such as collocations in second and foreign language classes. We must put special emphasis on the unique instances of collocations where the constituent parts almost always trigger one another (i.e., with higher MI scores). Some studies indicate that more advanced language users are more comprehensible in their language production with the help of collocations with higher MI scores (e.g., Saito & Liu, 2022). Coursebooks may integrate more unique combinations into their vocabulary lists which are conventionally used in everyday language and which native speakers of Italian are sensitive towards. Furthermore, as congruence has been found to play a significant role in the way collocations are perceived and rated by L3 users of Italian, more attention could be given to incongruent collocations in language teaching (Nesselhauf, 2003). Additionally, when designing target vocabulary lists, corpus data, native speaker intuition and learner experience can be taken into account. Merging different approaches

might create a more pedagogically convenient and empirically driven teaching materials (Cangır, 2021a; Jones & Durrant, 2010).

Future research may incorporate more lexical items extracted directly from corpora and with a wider range of AMs (see Gablasova et al., 2017 for further details regarding AMs). This could yield a different perspective and help draw a more reliable conclusion regarding how collocations are processed in the native and non-native mental lexicons or how and to what extent native and non-native frequency intuitions may differ. Additionally, a within-priming experiment exploiting the same or similar lexical items in both the languages can help us explore the phenomenon more comprehensively. Therefore, we believe intuition data needs to be supported by psycholinguistic evidence and future studies might consider investigating response times of language users, which could reveal how those collocations are processed and if/how languages interact during collocational processing.

6 Conclusion

In this research, we tried to test the earlier hypotheses regarding the possible positive relationship between the L2 English users' collocational frequency intuitions in English and corpus extracted AMs and explore the effect of congruence. We also attempted to extend our investigation and examine these associations from the L3 Italian users' perspective. The L3 Italian users are native Turkish speakers who have learned English as their L2. Our results indicate that (a) congruence and t-score are strong indicators of frequency intuitions of L2 English users, in accordance with earlier research (Cangır, 2021a) and (b) the same two variables are strong predictors of frequency intuitions of L3 Italian users. We conclude that congruence plays a significant role in the way collocations are structured and represented in the mental lexicon of both L2 and L3 users. In other words, the L1 effect can be seen in the frequency intuitions of non-native language users. Finally, we tentatively claim that AMs signifying how strongly word combinations are glued to one another in corpora, are also represented in the non-native users' mental lexicon to a certain degree and that this must have some pedagogical implications. As usage-based and exemplar-based theories (Ellis, 2012; Ellis, 2019; Goldberg, 1995; Tomasello, 2006) claim, mental representation of lexis is largely determined by language use and thus by the strength of association between word pairs and expressions. This is something coursebook designers need to consider more seriously when creating material and more importance should be given to the explicit instruction of multi-word units, such as collocations in language classrooms.

Appendix A. How to calculate association strength measures

1.

T-score $\frac{O_{11} - E_{11}}{\sqrt{O_{11}}}$

Explanation: "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.

MI (mutual information) $\log_2 \frac{O_{11}}{E_{11}}$

Explanation: "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 & Hanks, 1990, as cited in Gablasova et al., 2017, p. 8).

(Calculations have been taken from Brezina, 2018, p. 72)

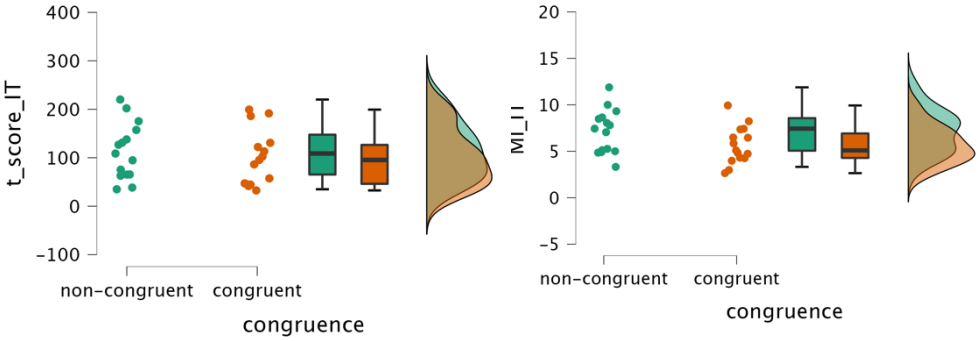
Appendix B. List of collocations

Item No.	ITALIAN (V+N)	ENGLISH (V+N)	TURKISH (N+V)
	CONGRUENT (same in all languages)		
1	affittare (una) casa	rent (a) house	ev kirala-
2	attirare (l') attenzione	draw attention	dikkat çek-
3	dire (una) bugia	tell (a) lie	yalan söyle-
4	cambiare (il) discorso	change (the) subject	konuyu değiştir-
5	alzare (la) voce	raise voice	sesini yükselt-
6	cercare (una) soluzione	seek (a) solution	çözüm ara-
7	guidare (la) macchina	drive (a) car	araba sür-
8	fare (un) errore	make (a) mistake	hata yap-
9	girare (la) pagina	turn (the) page	sayfayı çevir-
10	guadagnare fama	gain fame	ün kazan-
11	imparare (una) lingua	learn (a) language	dil öğren-
12	passare (il) tempo	pass time	zaman geçir-
13	prendere (la) forma	take (the) form	şekil al-
14	firmare (un) contratto	sign (a) contract	kontrat imzala-
15	trovare (una) via	find (a) way	yol bul-
	INCONGRUENT (same in Italian and English, different in Turkish)		
16	correre (il) rischio	run (the) risk	risk al-
17	causare danno	cause damage	zarar ver-
18	creare (una) relazione	create (a) relationship	ilişki kur-
19	dare (una) mano	lend (a) hand	el at-
20	perdere peso	lose weight	[2]kilo ver-
21	porre (l') accento	put emphasis	vurgu yap-
22	soddisfare (un) bisogno	satisfy (a) need	ihtiyacı karşıla-
23	sollevare (il) dubbio	raise doubt	şüphe uyandır-
24	riconoscere (il) valore	recognize value	değerini bil-
25	rispettare (la) legge	respect law	yasaya uy-
26	segnare (un) goal	score (a) goal	gol at-
27	rompere (il) silenzio	break (the) silence	sessizliği boz-
28	prendere (una) multa	get (a) ticket	ceza ye-
29	inventare (una) scusa	invent (an) excuse	bahane uydur-
30	prendersi cura	take care	dikkat et-

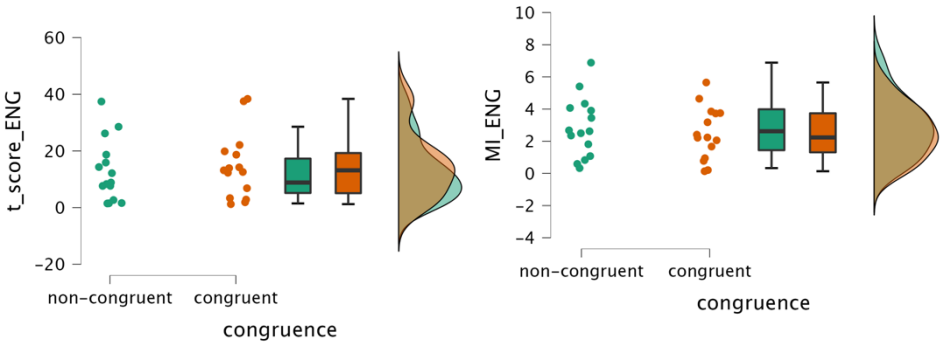
[2] We are aware of the word combination "kilo kaybet-" in Turkish, but the TNC indicates that it is not a frequent noun+verb collocation (with low MI and t-scores). Additionally, we think that noun+noun "kilo kaybı", which is the only frequent collocation according to TNC, has a negative connotation in Turkish (i.e., used in a different sense) and thus still can be considered incongruent in English and Italian.

Appendix C. Raincloud Plots

Italian:



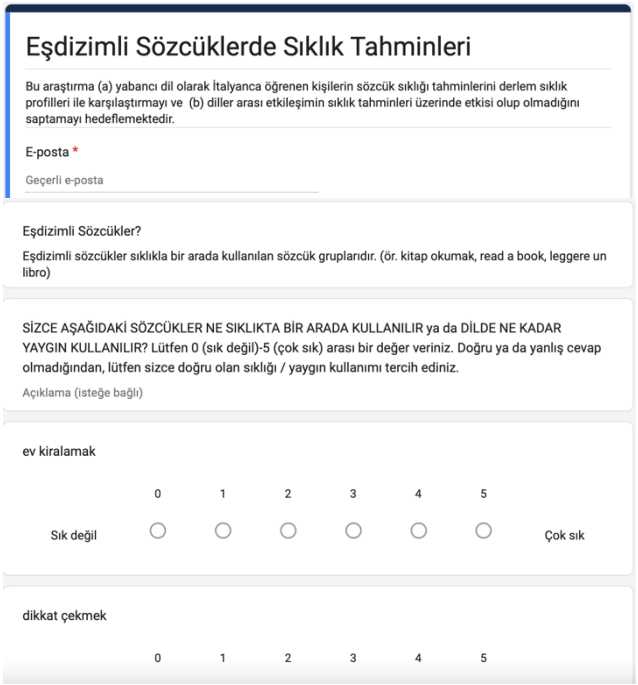
English:



Appendix D. AMs of all Collocational Items

No	t_score Italian	MI Italian	t-score English	MI English	t_score Turkish	MI Turkish
1	57,3954302	6,48575002	2,82750649	1,67011202	3,45975176	9,6372982
2	186,14205	9,93020433	38,374718	5,65169823	16,7604787	9,01866703
3	44,4817766	5,85123409	13,9276244	3,72563973	11,4480281	8,1270118
4	45,2386884	4,31369419	12,2986893	2,05778543	8,87443555	9,33538448
5	102,509885	7,3532242	12,5778523	2,23552247	2,63371927	7,78065264
6	122,170957	4,83163826	6,84097081	2,21566802	6,67933671	7,8603504
7	47,1034661	4,72920461	19,8502822	3,17804383	3,4542255	8,45432131
8	113,143545	2,9797526	37,5385409	3,84983683	4,3150126	4,83099451
9	41,5798837	3,9786769	1,21844858	0,13463824	5,83038081	13,3177509
10	32,5506834	7,42345835	3,33732993	3,74872389	2,9886331	8,04397984
11	86,2638554	6,46359039	13,1607411	2,40756304	7,42143417	6,91804649
12	199,297921	4,25456085	1,88059637	0,20697506	7,51136693	4,44048443
13	191,598667	5,09703337	14,2171279	0,77727583	6,20017458	5,66812898
14	130,774209	8,23483208	18,664363	4,64354056	5,09896826	16,6021137
15	95,4072765	2,6478093	22,097673	0,94701348	17,3053943	10,1624051
16	202,060017	8,4857016	15,8937361	2,6762442	10,4317087	7,53935923
17	175,339488	9,31055955	28,5334066	5,40464599	10,9151695	7,01870496
18	62,8968936	3,32189401	1,46988133	0,32886221	12,0082077	8,49452972
19	220,137689	4,85871394	7,67632653	2,62198466	8,43726114	7,46517611
20	131,244474	7,05032012	18,6880199	3,90476809	10,0480664	7,61721858
21	137,810469	9,99508227	8,31194971	1,81288461	10,3251891	9,09686568
22	126,83774	7,43786222	12,1550177	3,45022129	7,92296606	9,11770494
23	65,2270078	7,80008926	8,84634654	2,35264549	6,85504315	13,4527688
24	94,7383343	4,90474496	1,60870379	0,58952414	8,11389679	9,64576636
25	65,5439513	5,00313902	1,51615152	0,83052453	3,15976775	10,2991149
26	108,700207	11,8817019	26,1772123	6,88386873	19,9972949	12,8520187
27	75,3101872	8,64836337	14,3190894	4,3332644	12,2470916	15,0657421
28	38,4540579	5,2603956	7,65180366	1,07560524	2,82463441	3,22739241
29	34,9528304	8,04570673	2,6597751	4,06787844	6,78182807	13,7220474
30	157,321386	5,12025433	37,4403565	2,49748336	16,9895266	5,70958686

Appendix E. Sample Frequency Judgement Questionnaire



Eşdizimli Sözcüklerde Sıklık Tahminleri

Bu araştırma (a) yabancı dil olarak İtalyanca öğrenen kişilerin sözcük sıklığı tahminlerini derlem sıklık profilleri ile karşılaştırmayı ve (b) diller arası etkileşimin sıklık tahminleri üzerinde etkisi olup olmadığını saptamayı hedeflemektedir.

E-posta *

Geçerli e-posta

Eşdizimli Sözcükler?

Eşdizimli sözcükler sıklıkla bir arada kullanılan sözcük gruplarıdır. (ör. kitap okumak, read a book, leggere un libro)

SİZCE AŞAĞIDAKİ SÖZCÜKLER NE SIKLIKTA BİR ARADA KULLANILIR ya da DİLDE NE KADAR YAYGIN KULLANILIR? Lütfen 0 (sık değil)-5 (çok sık) arası bir değer veriniz. Doğru ya da yanlış cevap olmadığından, lütfen sizce doğru olan sıklığı / yaygın kullanımı tercih ediniz.

Açıklama (isteğe bağlı)

ev kiralamak

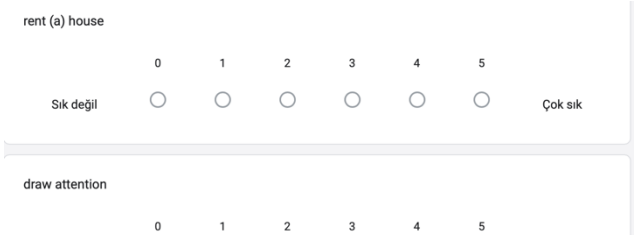
0 1 2 3 4 5

Sık değil Çok sık

dikkat çekmek

0 1 2 3 4 5

English version was conducted separately



rent (a) house

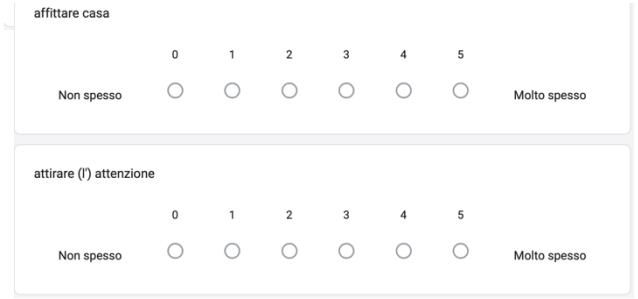
0 1 2 3 4 5

Sık değil Çok sık

draw attention

0 1 2 3 4 5

Italian version was conducted separately



affittare casa

0 1 2 3 4 5

Non spesso Molto spesso

attirare (l') attenzione

0 1 2 3 4 5

Non spesso Molto spesso

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).European Language.

Bestgen, Y. (2017). Beyond single-word measures: L2 writing assessment, lexical richness and formulaic competence. *System*, *69*, 65-78. https://doi.org/10.1016/j.system.2017.08.004

Bod, R. (2006). Exemplar-based syntax: How to get productivity from exemplars. *The Linguistic Review*, *23*, 291–320.

Brezina, V. (2018). *Statistics in corpus linguistic: A practical guide*. Cambridge University Press.

Bybee, J. (2010). *Language, usage and cognition*. Cambridge University Press. https://doi.org/10.1017/CBO9780511750526

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. (2021a). Objective and subjective collocational frequency Association strength measures and EFL teacher intuitions. *Pedagogical Linguistics*. *2*(1), 64–91.

Cangır, H. (2021b). 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?. *Journal of Linguistics*, (37), 45-66. DOI: 10.26650/jol.2021.972166

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

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.2019.10273884

Christiansen, M. H., & Chater, N. (2001). Connectionist psycholinguistics: Capturing the empirical data. *Trends in Cognitive Sciences, 5*(2), 82–88. https://doi.org/10.1016/S1364-6613(00)01600-4

Durrant, P., & Schmitt, N. (2009). To what extent do native and non-native writers make use of collocations? *IRAL*, *47*, 157–177. https://doi.org/10.1515/iral.2009.007

Durrant, P., & Doherty, A. (2010). Are high-frequency collocations psychologicallyreal? Investigating the thesis of collocational priming.

Corpus Linguistics and Linguistic Theory, *6*, 125–155. https://doi.org/10.1515/CLLT.2010.06

Ellis, N. C. (2001). Memory for language. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 33-68). Cambridge University Press.

Ellis, N. C. (2003). Constructions, Chunking, and Connectionism: The Emergence of Second Language Structure. In *The Handbook of Second Language Acquisition* (pp. 63–103). https://doi.org/https://doi.org/10.1002/9780470756492.ch4

Ellis, N. C. (2012). Formulaic language and second language acquisition: Zipf and the phrasal teddy bear. *Annual Review of Applied Linguistics*, *32*(2012), 17–44. https://doi.org/10.1017/S0267190512000025

Ellis, N. C., & Ogden, D. C. (2017). Thinking about multiword constructions: Usage-based approaches to acquisition and processing. *Topics in Cognitive Science*, *9*(3), 604–620. https://doi.org/10.1111/tops.12256

Ellis, N. C. (2019). Usage-based theories of Construction Grammar: Triangulating Corpus Linguistics and Psycholinguistics. In Jesse Egbert & Paul Baker (Eds.), *Using Corpus Methods to Triangulate Linguistic Analysis* (pp. 239-267). Routledge

Fioravanti, I., Senaldi, M. S. G., Lenci, A., & Siyanova-Chanturia, A. (2021). Lexical fixedness and compositionality in L1 speakers' and L2 learners' intuitions about word combinations: Evidence from Italian. *Second Language Research*, *37*(2), 291–322. https://doi.org/10.1177/0267658320941560

Gablasova, D., Brezina, V., & McEnery, T. (2017). Exploring learner language through corpora: Comparing and interpreting corpus frequency information. *Language Learning*, *67*(S1), 130–154. https://doi.org/10.1111/lang.12226

Goldberg, A. (1995). *Constructions: A construction grammar approach to argument structure*. Chicago University Press.

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

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

Hoffmann, S., & Lehmann, H. M. (2000). Collocational Evidence from the British National Corpus. In J. M. Kirk (Ed.), *Corpora Galore: Analyses and Techniques in Describing English*. Amsterdam: Rodopi.

Jakubíček, M., Kilgarriff, A., Kovář, V., Rychlý, P., & Suchomel, V. (2013). The TenTen corpus family. In *7th International Corpus Linguistics Conference CL* (pp. 125-127).

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* (pp. 387-400). Routledge.

Kyle, K., & Crossley, S. (2016). The relationship between lexical sophistication and independent and source-based writing. *Journal of Second Language Writing*, *34*, 12–24.

Lenci, A., Fioravanti, I., & Siyanova-Chanturia, A. (2021, August 15–20). *Collocational priming: Are collocations psychologically real for L1 and L2 speakers of Italian?* [Paper presentation]. American Association for Applied Linguistics.

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

Lewis, M. (1997). "*Implementing the lexical approach: Putting theory into practice*." Hove, England: Language Teaching Publications.

Nesselhauf, N. (2003). The use of collocations by advanced learners of English and some implications for teaching. *Applied Linguistics*, *24*(2), 223-242

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

Saito, K., & Liu, Y. (2022). Roles of collocation in L2 oral proficiency revisited: Different tasks, L1 vs. L2 raters, and cross-sectional vs. longitudinal analyses. *Second Language Research*, *38*(3), 531–554. https://doi.org/10.1177/0267658320988055

Schmitt, N. (2010). *Researching vocabulary: A vocabulary research manual.* New York: Palgrave Press. https://doi.org/10.1057/9780230293977

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

The British National Corpus, version 3 (BNC XML Edition). (2007). Distributed by Oxford University Computing Services on behalf of the BNC Consortium. URL: http://www.natcorp.ox.ac.uk/

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

Tiberii, P. (2018). *Dizionario delle collocazioni. Le combinazioni delle parole in italiano*. Bologna: Zanichelli editore.

Tomasello, M. (2006). Acquiring Linguistic Constructions. In D. Kuhn, R. S. Siegler, W. Damon, & R. M. Lerner (Eds.), *Handbook of child*

psychology: Cognition, perception, and language (pp. 255–298). John Wiley & Sons Inc.

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

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

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

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