Available online at www.jlls.org

JOURNAL OF LANGUAGE AND LINGUISTIC STUDIES

ISSN: 1305-578X

Journal of Language and Linguistic Studies, 13(2), 465-486; 2017

Investigating collocational priming in Turkish

Hakan Cangır [a*], S. Nalan Büyükkantarcıoğlu [b], Philip Durrant [c]

[a] *Ankara University, School of Foreign Languages, Gölbaşı, Ankara, 06830, Turkey*
[b] *Hacettepe University, Faculty of Letters, Department of English Linguistics, Ankara, 06800, Turkey*
[c] *University of Exeter, St Luke's Campus, Heavitree Road, Exeter, EX1 2LU, UK*

APA Citation:

Cangır, H., Büyükkantarcıoğlu, S. N., & Durrant, P. (2017). Investigating collocational priming in Turkish. *Journal of Language and Linguistic Studies, 13*(2), 465-486.

Submission Date:23/08/2017

Acceptance Date:13/09/2017

Abstract

Several attempts have been made to illustrate the organization of the monolingual mental lexicon and each model proposed so far has highlighted different aspects of lexical processing. What they have in common is the fact that their depictions rely on single lexical items and paradigmatic relations come to the fore in their explanations. Hoey's lexical priming theory (2005) tries to shed light on the issue of collocational processing in the internal lexicon from a cognitive and psycholinguistic perspective and its importance for our overall creative language production. A number of psycholinguistic studies have tested Hoey's theory as it relates to English, but work in other languages is limited. The present study broadens the scope of work in this area by investigating whether collocational priming also holds for speakers of Turkish. Furthermore, the possible influence of frequency and part of speech on collocational priming is scrutinized by exploring the correlations between response times in the priming experiment and these independent variables. The findings revealed a significant collocational priming effect for Turkish L1 users, in line with Hoey's claims. The regression analysis indicated frequency and part of speech as important predictors of processing duration. The correlation analysis also showed significant correlations between the response times and both word and collocational frequency. A tentative mental lexicon framework is proposed based on the findings of this research.

Keywords: Mental Lexicon; collocations; frequency; lexical priming; collocational priming

1. Introduction and Literature Review

As Sinclair (1991) and Hoey (2005) state, to shed on light on the principles behind the processing and acquisition of collocations, we need to look at them within a broader perspective of formulaic language as a whole.

'Formulaic language' has been defined as 'recurrent multi-word lexical items having a single meaning or function' and it is generally employed as an umbrella term for idioms, collocations, lexical bundles etc. (Schmitt, 2010). Writers have addressed the issue of formulaic language in many different ways and used different terms, often in inconsistent ways (Wray, 2002). Many researchers (e.g. Wray, 2002; Schmitt, 2010) acknowledge that formulaic language is one of the key components of language

* Corresponding author. Tel.: +90-312 485 05 07
E-mail address: hakancangir@gmail.com

mainly because of its pervasiveness in language use. Furthermore, meanings and functions are achieved by dint of formulaic language and the language users producing formulaic phrases in language production enjoys a processing advantage (Conklin and Schmitt, 2012).

The reason why researchers concentrate on formulaic language emerges from the viewpoint that formulas are basic language units (e.g. Conklin and Schmitt, 2012). This theoretical stance is affected by Sinclair's (1991) idiom principle and by pattern grammar (Hunston and Francis, 2000), and construction grammar (Goldberg, 2006). Sinclair claims that a language user knows a huge number of semi-preconstructed phrases, many of which are uttered in speech and can be observed in texts. It is even estimated that about half of fluent native text is shaped based on idiom principle.

Another rationale comes from the theoretical position that formulas seem to have a unique psycholinguistic status and that they have a vital role in language acquisition (Schmitt, 2010). The investigation of formulaic language is of importance due to the fact that there may be a link between the learners' use of formulaic language and their perceived proficiency in language (e.g. Staples, Egbert, Biber and McClair, 2013), though no conclusive results have been observed based on empirical research. However, it has been concluded by many researchers that formulaic sequences, statistically defined and extracted from a large and balanced corpora have indications for educational and psycholinguistic research and applications (Ellis and Simpson-Vlach, 2009).

Given that formulaic language plays an important role in language processing and language acquisition and that collocations are regarded as a sub-category of this group, the current research, which investigates collocational priming in Turkish, approaches the issue of lexical processing from a syntagmatic perspective and attempts to come up with a tentative framework for the structuring of collocations in the internal lexicon.

As stated by Cruse (2000), the vocabulary of language is comprised of two main relations, which are paradigmatic and syntagmatic links. Based on this organization, collocations can be depicted under the syntagmatic branch together with other multi-word units, whereas synonyms, antonyms and hyponyms are classified in the paradigmatic end.

In addition to where collocations stand in the vocabulary knowledge organization, the definition of the term is also an important issue to consider and has been a controversial phenomenon in psycholinguistic, corpus linguistic and language acquisition research.

Firth (1957), who is considered as one of the first linguists to use the term collocation in its modern linguistic sense, says:

> Meaning by collocation is an abstraction at the syntagmatic level and is not directly concerned with the conceptual or idea approach to the meaning of words. One of the meanings of night is its collocability with *dark*, and, *of dark*, of course, collocation with *night*. (Firth, 1957: 196)

As is discussed in the previous section, collocations are commonly seen as a subcategory of formulaic language (Wray, 2002). Notwithstanding their apparently prevalent use in language, collocations are difficult to define (Wolter & Yamashita, 2014). Two commonly accepted approaches to the definition can be observed in the literature. The first one, the phraseological approach (Cowie, 1994; Howarth, 1998), asserts that a word cluster can be considered a genuine collocation on condition that one of the words in the cluster is non-compositional (i.e. non-transparent or opaque), which makes the combination semi-transparent. If both the members of the combination are fully compositional, the item is then called a `free combination` (as in "brush teeth") as far as the phraseological approach is concerned. If both the members are non-transparent or opaque, the cluster is named as `idiom` (as in "kick the bucket"). Benson et al. (1986) stated word combinations are grouped according to three

principal criteria; the level of cohesiveness, semantic transparency and frequency. The basic problem with the classification provided by the phraseological approach is the fact that it is challenging to decide on the boundaries between these categories.

The second acknowledged approach has close links with corpus linguistics and employs statistical measures to investigate the frequency of the co-occurrence of certain word patterns (Sinclair, 1991). The rationale behind the frequency approach originates from the idea that the more frequent word combinations exist together in written or spoken language, the more likely they are to be entrenched in the mental lexicon and can be seen as collocations. Native speakers of the language and even some advanced second language users produce these word combinations automatically and they enjoy a processing advantage, which eventually affects their fluency. According to Henriksen (2013), integrating the corpus approach into research appears to be logical because then you rely on objective criteria, such as frequency, range and span, rather than your own intuition about word pairs. As for the problems regarding this approach, as Howarth (1998) states it focuses on performance and take no notice of competence. Extracting word pairs from corpora based on frequency measures without paying attention to semantics could reveal word pairs that native speakers would not consider as a collocation, as in the case of English definite article 'the'. It appears to collocate with all the nouns due to its pervasive use in language and if researchers rely on corpus data only, the frequency measures are likely to misguide them in their analysis and interpretation if the primary aim is to explore the collocational processing in the mental lexicon. In other words, without considering the semantic aspect, corpus extracted word pairs tend to lack strong psycholinguistic legitimacy for the language users.

Given that each approach has its strengths and weaknesses, the current research applied both the strategies as complementary methods, in line with some earlier research (see Nesselhauf, 2005 for a discussion). Therefore, according to the current research, in order for a word combination to be considered as a collocation, it must be frequent at a certain level (benchmarks are given in the methodology section) and semi-transparent, an approach that was employed by some earlier research (e.g. Kjellmer, 1984; Kjellmer, 1987). As this study was conducted to set a baseline for a cross-linguistic investigation, the lexical items were adopted from the main experiment. Recurrent word combinations in two balanced corpora (Corpus of Contemporary American English and Turkish National Corpus) were detected with the help of association measures, which will be discussed in more details in the methodology section. After that, the list of collocations was fine-tuned based on their semantic features (i.e. compositionality). We believe that this mixed approach employed in deciding the word combinations to be used in the experiment was a sound move considering the pros and cons of each approach and their complementary nature.

The discussion so far have tried to shed light on the basic concepts, formulaic language and collocations to provide some basic insight into syntagmatic relations between words. The core paradigm employed in the study also needs explaining before giving details about the methodology.

Firth's famous saying "you shall know a word by the company it keeps" has been used and adopted by many linguists and the philosophy behind this notion has been discussed and enhanced in many aspects over the years. Having its roots in Firthian tradition, a new theory of lexical priming was proposed by Hoey (2005). The theory asserts that every word is mentally primed for collocational use and collocational priming is sensitive to the contexts where the lexical unit is encountered. The fact that a lexical item is employed in specific combinations in particular types of texts constitutes part of our knowledge of that lexical unit. According to his definition of the term, collocation:

> "Collocation is a psychological association between words which is evidenced by their occurrence together in a corpora more frequently than is rational in terms of random distribution" (2005, pp. 3-5)

Hoey (2005) further claims that priming can also be seen as the source of our creative language system. According to him, the grammatical categories assigned to lexical units are determined by lexically specific patterns of priming rather than an independently existing grammar. This view is in accord with the usage-based models, which are closely linked with Cognitive Linguistics and Construction Grammar (Barlow & Kemmer, 2000). The cognitive view of language postulates that language learning emerges from general practices of human inductive reasoning being applied to the specific problem of language (Tomasello, 2003). Unlike the Chomskyan view of language, cognitivists assert language acquisition device per se does not exist. Rather, language goes hand in hand with other cognitive processes though its cognitive content could vary. In addition, cognitive view of language posits that genes do not appear to be the mere source of language. On the contrary, the language emerges from the structure of adult language and the structure of social and cognitive skills (Ellis, 2001).

Considering his views and stance, one can deduce that Hoey is at odds with Generative Grammar (Chomsky, 1965) and approaches the issue from a psycholinguistic perspective. According to the Chomskyan view of language, the principal goal of linguistics is to investigate speakers' competence, which is also defined as the abstract system of linguistic knowledge, rather than linguistic performance. Chomsky is interested in the internalized (i-) language, not the externalized (e-) language. On the contrary, what Hoey and Sinclair concentrate on is the exploration of e- language enhanced by corpora. Sinclair states that scrutinizing competence and disregarding real life language in an attempt to escape the noise or the disorganization in language use does not make sense as the larger-scale corpora these days are powerful enough to help researchers to get a clear picture of real language use and find significant patterns of various language phenomena (1991, p. 103).

On the whole, Hoey thinks all the priming forms; lexical, textual, grammatical etc. accumulate as one is exposed to the real language around him. Because we have different language learning experiences, the priming effect can differ slightly for each person. However, those minor variations appear to be adjusted in time as we have more exposure since there needs to be some standards so that language users can comprehend each other through a common use of lexical units (2005, p. 9). These standards he says include education, traditions, the mass media and reference works like dictionaries (2005, pp. 181-182).

Hoey accepts that priming might harbour some conflicts. A basic example can be observed in the rules that are taught at school or in grammar books which seem to contradict with native speaker intuition. To give an example from Turkish, we can think of the "neither …. nor …" (*ne…..ne de….*) situation. Considering the negative form of the phrase, native speakers of Turkish are primed to use a negative verb at the end of this phrase (*Ne annesi ne de babasi ona yardim etmedi-* "Neither his mother nor his father did not help him") using their native speaker intuition; however, Turkish grammar states the opposite (*Ne annesi ne de babasi ona yardim etti-*"Neither his mother nor his father helped him"), which is the correct grammatical form of the sentence, according to prescriptive grammars.

There are some studies exploiting the collocational priming paradigm, which were conducted to find evidence for psycholinguistic notion of priming. Those studies mainly used experimental psycholinguistic techniques and tools, such as lexical decision, word naming, semantic association etc.

In one of those studies by Durrant and Doherty (2010), evidence for collocational priming was found and the writers claimed that their findings were partly in line with Hoey's (2005) lexical

priming theory. It was the first research proposing a frequency based collocational priming explanation independent of psychological association. However, in their second experiment, they found inconsistent results with the first application. The results indicated that there was a priming effect for the associated word pairs but not for high frequency collocations. Therefore, they were cautious in their interpretation and called for further research. In an earlier study by McKoon and Ratcliff (1992), a weak priming effect was detected for high frequency collocations. The researchers reported their limitations as a small size corpus and a lack of a psychological association measure. Thus, they tentatively suggested a possible priming influence and avoided making strong claims.

There have been many other attempts to shed light on the processing of collocations in L1 and L2. Some researchers, Wray (2002, 2008) in particular, claimed that native speakers (NS) process collocations or formulaic phrases as chunks, whereas non-native speakers (NNS) decompose the whole into its single units to process. However, some others (e.g. Durrant and Schmitt, 2010) disagreed with Wray's stance claiming that NS and NNS do not differ in their approach to the acquisition of collocations. Rather, NNS process collocations differently in that they have insufficient language input and limited exposure.

The studies discussed above attempted to test the hypothesis that words are primed to co-occur or question if they are stored as chunks in the mental lexicon, an idea different versions of which have been proposed and discussed for a long time (e.g. Sinclair, 1987; Ellis 2001; Hoey, 2005). However, no research to the researchers' knowledge to this date has considered a typologically different language in its investigation and approached the issue of collocational priming from this angle. Having this notion in mind, the writers of the current research seeks to answer the research questions below:

a- Does collocational priming exist in Turkish?

b- To what extent does frequency play a role in collocational priming, if any?

To this end, a monolingual priming experiment including a lexical decision task was designed following the standards of the paradigm. The details of the approach are provided in the following section.

2. Method

2.1. Overall Design

The application was a lexical decision task including a balanced number of collocations, non-collocations, and some filler items to balance the proportion of the target items with the control and non-word items (with a relatedness proportion of 0.24 and a non-word ratio of 0.27). To be more precise, for each collocational item (e.g. *soğuk savaş* – "cold war"), there was one non-collocation with the same target word but a different prime word with the same word length (+/-1) and a similar prime word frequency, (e.g. *uzak savaş* – "far war"), a filler non-collocation consisting of random words with the same target word length (+/-1), (*geniş nefret* – "broad hatred"), and a non-word pair consisting of a random prime word followed by a non-word made up by the Turkish L1 members of the research team (e.g. *çukur sagit* – "hollow sagit"). Additionally, having relatedness proportion and non-word ratio concerns, the team came up with ten more non-collocation items and non-word items including made-up words (i.e. fillers) with similar word length with the other items. Eventually, only the mean response times for the collocate (e.g. *soğuk savaş* – "cold war") and corresponding non-

collocate items (e.g. *uzak savaş* – "far war") were investigated in the regression and correlation analyses and the response times of all the other lexical items were ignored intentionally due to the design of the current research.

The relatedness proportion stands for the ratio of accompanying prime–target lexical items out of all the lexical items. It is claimed that the bigger the relatedness proportion is, the stronger the semantic priming is (de Groot, 1984). That's why, a standard level (lower than 0.25) mentioned in Jiang (2012) was adopted. The non-word ratio is the proportion of non-words to all the collocational, non-collocational items and unrelated word pairs (see Altarriba and Basnight-Brown, 2007 for a discussion).

The stimulus onset asynchrony (SOA), which is described as the time interval between the prime word and the onset of the target word, was set to 100 milliseconds to comply with the standards of the priming paradigm based on the discussion by Jiang (2012). The remote version of DMDX[1] was used in the current research since one of the researchers was abroad during the actual application. The research team compiled the priming experiment script together with a simple batch file so that the test could run automatically on each participant's screen and send the results of the experiment to the team as an e-mail. The lexical items were presented at a random order. The subjects were guided through a web-interface designed for this research only, which includes all the details about the procedure and the necessary steps. Example items from the priming experiment are shown in Table 1:

Table 1. A sample DMDX screen

SCREEN 1 * (500 ms)	SCREEN 2 ######### (200 ms)	SCREEN 3 prime word (100 ms)	SCREEN 4 target word (response is recorded)	Item type
*	#########	yapmak	HATA	Collocation
*	#########	almak	HATA	Non-collocation
*	#########	dürtmek	PAZI	Filler
*	#########	çarpmak	LATİ	Non-word

After the priming experiment, the subjects took an online end of test questionnaire answering questions about vision, dexterity and priming items. They were asked if they were able to consciously see the priming items flashed before the target words for 100 milliseconds and whether they detected a pattern between the stimulus and the target to make sure the collocational processing was automatic and they were not making use of any conscious strategies during lexical processing.

The output of the lexical decision task and the frequency values (i.e. the difference between the mean response times of collocate and non-collocate items only and the relationship between the mean response times and frequency measures) were analysed using the Statistical Package for the Social Sciences (SPSS) 23 software.

[1] a software developed at Monash University and at the University of Arizona by K. I. Forster and J. C. Forster (2003) and provided as an open-source tool

2.2. Participants

41 native speakers of Turkish (27 female and 14 male) took part in the study. Participants were either undergraduate students at Ankara University (N=28) or lecturers from different universities in Ankara (N=13). They were aged between 18 and 55.

Several instruments were used during the process. A digit span test was employed in an attempt to evaluate the possible participants' short term memory and make sure that they can keep a lexical item they see on a computer screen in their mind for a required period of time. The test is used as a standard procedure in psycholinguistic experiments and it was conducted through a simple java application in which participants were asked to recollect the numbers presented to them and write them on the screen accurately. The application provided a digit span score in the end indicating an overview of the participants' short term verbal memory. All the subjects scored 6 and above in the test and they took the monolingual collocational priming experiment, which was conducted with the help of the DMDX software.

2.3. Item development

An important issue for the current study was to extract the collocational items from corpora based on both statistical and semantic aspects, the rationale of which has been discussed in the previous section. Because this monolingual experiment was the first step of a cross-linguistic priming study, the researchers made use of the Contemporary American English Corpus (COCA), (Davies, 2008-) and Turkish National Corpus (TNC), (Aksan et al., 2012) in combination. First, around 70 V+N and 70 ADJ+N collocations were chosen from the COCA list (Davies, 2008-) of English collocations, which provided only the MI values of all the word combinations on the list as a frequency measure. The *t*-scores of all those collocational items were also computed separately with the help of a spreadsheet developed by Philip Durrant. The chosen collocations were required to have an MI score of at least 3.0 and a *t*-score of 2.0, which was mentioned as a benchmark in some research (Schmitt, 2010) and to be semantically semi-transparent. The research team chose the semi-transparent collocations based on their native speaker intuitions and then two objective eyes were asked to confirm the semantic opaqueness of the items. Once the items that were semi-transparent were chosen, they were cross-checked with their Turkish counterparts on TNC to make sure they had an MI score of at least 3.0. Together with the MI score, which has its weaknesses like any other association measures, *t*-score was also integrated into the study as a complementary frequency measure. The items were fine-tuned so that they had an MI score of at least 3.0 and a *t*-score of at least 2.0, both in Turkish and English, to comply with the standard benchmark values in Schmitt (2010). Additionally, the research team made sure that the chosen items in Turkish and English had no case marking since it is believed by many prominent linguists (e.g. Hoey, 2005; Sinclair, 1991) that lemmatization tends to fail to reflect essential differences in collocational preferences between different forms of a lemma. The decision can also be attributed to Durrant's (2014) findings indicating that the difference between lemmatized and non-lemmatized frequency values in terms of their correlation with the learner knowledge of collocations is vague. Additionally, the trial version of TNC didn't allow for a part of speech search, which made the possible lemmatization goal hard to achieve.

With regard to the type of collocations chosen for the current study, Verb+Noun (V+N) and Adjective+Noun (ADJ+N), which were investigated comprehensively by previous research as well (e.g. Siyanova&Schmitt, 2008; Fan, 2009; Barfield&Gyllstad, 2009; Wolter, 2006; Wolter&Gyllstad, 2011 etc.), were selected for a specific and a unique purpose. This study is part of a larger, cross-linguistic (Turkish-English), study for which the fact that adjective-noun word order is similar between the two language but verb-noun word order is not will be important. Because the current

research was employed as a starting point for the follow-up cross-linguistic experiment, it adopted the same lexical items extracted and categorized for the cross-linguistic investigation so that the findings can be reliable and comparable. Furthermore, although the word order for V+N collocations in Turkish is the opposite of English language (N+V), the researcher preferred the English word order in the priming experiment to have a comparable data for the planned future study. For instance, when the collocation was *ışık tutmak* – "shed light", the prime word was *tutmak* – "shed" and the following target word was *ışık* – "light" in the priming script.

Taking into account all these factors, the research team came up with thirty ADJ+N and thirty V+N items with no case marking that were chosen strategically for cross-linguistic investigation purposes to be used in another experiment and the same items were employed in the current experiment to be able to have comparable data in the end.

Although it was not part of the item development procedure, another association measure, Delta P (ΔP) by Gries (2013) was integrated into final the analysis to test the possible bidirectional activation of collocational networks. (See the complete list of items in Appendix A and how MI, *t* and ΔP values were computed in Appendix B)

3. Results

Every subject confirmed that they had normal or corrected to normal vision. All the participants were right hand dominant except for a single subject who was dominant in both hands. Below is an overview of the participants' biographical information, dexterity and vision.

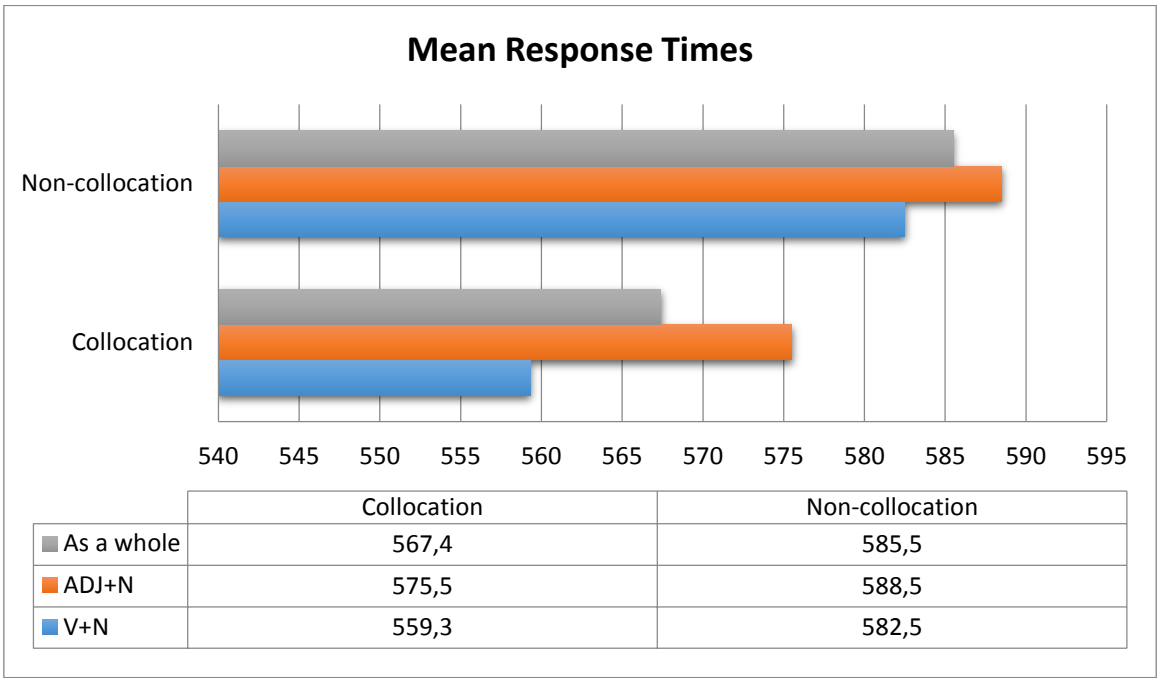
Table 2. Summary of participants' biographical information

GROUP	Age[a]	Dexterity (R/L/B)	Gender (M/F)	Vision
Turkish ONLY *(N=41)*	Mean: 24.4	R 40/97.6% B 1/2.4%	F 27/65.9% M 14/34.1%	No serious issues

[a] range=18-55

Subjects took a digit span test before the experiment and everybody scored 6 or more (*Mean=7.5*, *Range=6-9*), which was regarded as sufficient for a normal short term verbal memory. Participants at Ankara University and other universities (Hacettepe, METU etc.) took the remote version of the priming test. They were asked to take the test in a silent environment where they can focus on the task only and nobody will interrupt them. Despite the fact that 41 subjects took the priming test, the results of twenty eight participants were deemed to be consistent and worth investigating further on the grounds that some participants had more than 20% error rate, which is considered as a threshold in research adopting priming paradigm (Jiang, 2000). Moreover, the response times faster than 200 milliseconds, slower than 2000 milliseconds, and the items with more than 2.0 standard deviation were removed from the overall data in an attempt to adhere to the priming paradigm standards. It is commonly though that a language user cannot decide if a lexical item is a word or not in less than 200 milliseconds in a lexical decision task and if he/she does, it means he/she is not paying attention to the

task, which makes the results unreliable. If the participants are spending more than 2000 milliseconds processing a lexical item, that could indicate a strategic attempt, which needs to be avoided in experiments aiming for automatic priming effect. The results of the monolingual Turkish priming experiment are displayed in Figure 1 as an overview and in more details in Table 3.



	Collocation	Non-collocation
As a whole	567,4	585,5
ADJ+N	575,5	588,5
V+N	559,3	582,5

Figure 1. Mean response times in milliseconds general view

Table 3. Mean response times in milliseconds, standard deviations in parenthesis and error rates in square brackets

Number of lexical items	Collocation RT	Non-collocates RT	Priming Effect
60 items (120 total)	As a whole 567.4 (40.14) [1.52%]	As a whole 585.5 (38.89) [1.46%]	18.1 *p=.001, r=.41
30 items (60 total)	V+N 559.3 (33.54) [1.84%]	V+N 582.5 (34.92) [1.23%]	23.2 *p=.009, r=.46
30 items (60 total)	ADJ+N 575.5 (44.91) [1.24%]	ADJ+N 588.5 (42.89) [1.7%]	13.0 *p=.05, r=.36

*The significance level is .05

Based on the difference between the mean response times of the collocate and non-collocate items, it can be deduced that the stimuli primes the target if the word combination is a collocation, which indicates eventually that collocational priming appears to exist in the Turkish language. The priming effect for each condition is statistically significant at the level of $p < .05$. Although the priming effect in the ADJ+N group is also significant, it appears that participants responded to the lexical items faster

if they are part of a V+N collocation, which resulted in a considerably stronger priming effect in this group. The possible reasons behind this fast processing will be discussed in the next section.

As for the effect sizes of each category, when all the items were merged, the effect size of the priming effect was strong at the level of *r=.41*. However, when each part of speech group was analysed on its own, V+N collocations reflected a strong effect size of *r=.46*, whereas the ADJ+N collocations demonstrated a medium effect size (*r=.36)*. On the whole, when the mean response times of only the non-collocate items of each group are observed, it can be seen that there is not a big gap between them; however, when the collocate items are considered, one can conclude that the mean response durations are remarkably lower than the non-collocate ones and the V+N items were processed faster than the ADJ+N items by the Turkish participants.

Another issue to note is that error rates for each category were low due to the fact that outliers were trimmed during the data categorization and analysis process, thus it can be claimed that the results seem to be relatively reliable in that participants paid enough attention to the experiment and the response times that are out of the priming paradigm standards have been eliminated.

In an attempt to answer the second research question, a correlation and a regression analyses were conducted, the results of which could reveal a possible relationship between the dependent variable, mean response time and the association measures and part of speech exploited as independent variables. Furthermore, the regression analysis indicated the possible significant indicators of the mean response time in the priming experiment.

The table below elucidates the significant correlations between the mean response times in the collocational priming experiment and the frequency values employed in the study.

Table 4. Correlation Analysis Results

	Mean Response Times
Collocation status	-.224*
Target word frequency	-.346**
t-score	-.334**
$\Delta P_{1\|2}$	-.248**
$\Delta P_{2\|1}$	-.199*
MI score	-.166*

**Correlation is significant at the .01 level
*Correlation is significant at the .05 level

It can be concluded based on the results of the correlation analysis that the mean response times of the lexical items seem to have significant inverse correlations with collocation status (*r*=-.224, *p*.05), target word frequency (*r*=-.346, *p*.01), *t*-score (*r*=-.334, *p*.01), $\Delta P_{1|2}$ (*r*=-.248, *p*.01), $\Delta P_{2|1}$ (*r*=-.199, *p*.05), and MI (*r*=-.166, *p*.05) scores in Turkish. To be more precise, the inverse relations considering the negative correlations between the mean response times of the lexical items show that as the frequency values increase, the mean response durations decrease. That is to say, frequency can be regarded as a medium that facilitates collocational processing. All the frequency values presented in the table indicated a moderate correlation strength, whereas the $\Delta P_{2|1}$, and MI value revealed a weak correlation.

The fact that there is a correlation between the mean response times and ΔP values in both directions is also worth underlining, which could mean that the effect of the prime word on the target is as important as the influence of the target word on the prime word; that is to say, the interaction of the lexical items in the mental lexicon may be bidirectional.

Another obvious negative correlation can be seen in the variable, collocation status. As the analysis in the first part of this research revealed, if the presented lexical combination was a collocation, it led to a faster response time and the correlation results show a similar trend. Though the results should be treated cautiously, the correlations could indicate a possible effect of frequency on collocational priming in Turkish. Further research is needed to make strong claims about the reasons for the priming effect.

In addition to the correlation analysis, which revealed some significant relationships between the mean response time and frequency values, a regression analysis was carried out in order to investigate the potential predictors of the mean response time in the priming experiment, which could yield some information regarding the partial effect of frequency on the priming effect and the processing of collocations in the mental lexicon.

The table below shows the regression analysis results of the monolingual collocational priming experiment.

Table 5. Regression Analysis Results

	B	***SE b***	**Beta**
Model			
Constant	619.920	14.545	
POS	14.387	6.814	.179*
Target word frequency	-23.308	7.554	-.285*
t-score	.643	1.004	-.085
MI score	3.832	3.144	.347
$\Delta P_{1\|2}$	-39.370	26.384	-217
$\Delta P_{2\|1}$	-23.963	25.286	-.131

Note for model: R=.543[a] and R^2=.2295 (p<.001)
* The significance level is p<.05

The results of the regression showed the predictors explained 22.9% of the variance (R^2=.229, F=4.76, p<.001) for the model. It was found that part of speech significantly predicted the mean response time in the collocational priming experiment (β=-.179, p=.05). In addition, target word frequency revealed itself as another significant indicator of mean response time (β=-.-285, p=.05). *t*-score can also be claimed to predict the mean response time in the priming experiment based on the regression results, but the p value does not allow to make strong claims.

Overall, it can be stated that part of speech and target word frequency appear to influence the mean response time more than other variables indicating collocational frequency. The effect of part of speech can be deduced based on the numbers in the previous analysis showing faster processing in V+N collocations and a more robust priming effect in V+N word combinations than ADJ+N collocations. Therefore, one can assert that part of speech, target word frequency, and *t*-score (though tentatively) plays a partial role in how collocations are processed and appears to have an impact on collocational priming. Unlike the correlation analysis, which revealed the frequency measures, *t*-score, ΔP, and MI having a significant correlation with the mean response times, the regression analysis didn't indicate a similar pattern for the predictors of mean response time in the priming experiment. This is likely to raise some issues regarding the claims made earlier about the priming effect; however, it must be underlined that the experiment was designed and the items were controlled in such a way that the participants saw different prime words but the same target words with regard to the collocational and non-collocational items, the mean response times of which were compared to find

proof for collocational priming. To be more precise, if the subjects saw the collocational item *derin uyku* – "deep sleep", the non-collocational item whose response time was taken into account in the analysis was *gizli uyku*-"secret sleep". That is to say, the target words were the same and the possible effect of the differing frequency between the words were eliminated.

The explanations so far have addressed the L1 Turkish subjects' performance in the priming study, the priming effects observed, and the relationship between the frequency values and the response times. The last section will deal with the interpretation of the findings of the collocational priming study, regression and correlation analyses and the research team will attempt to explain the issue of collocational priming in Turkish by referring to a mental lexicon model.

4. Discussion

As Bybee (2005, p.112) states "words used together fuse together". In a similar vein, Hoey (2005) claims words are primed to co-occur and the activation of the node spreads to the collocate. This priming is asserted to be the basis of our creative language system. Investigating the reality of collocational priming in Turkish, the current study attempted to shed light on the effect of frequency on a possible priming effect in Turkish and approach the issue of mental lexicon organization from a syntagmatic perspective.

The first overall conclusion that can be drawn based on the results of the priming experiment, regression and correlation analyses is that collocational priming seems to exist in Turkish for ADJ+N and V+N (though regular word order is N+V in Turkish) collocations with no case marking and frequency has an important impact on the lexical processing. As stated earlier, the lexical items were presented in V+N for a specific reason and the fact that there was a priming effect despite the irregular word order in Turkish presented in the priming experiment could be ascribed to the flexibility of Turkish in word order, particularly in spoken production. In other words, as opposed to the strict word order in English for V+N collocations, Turkish language users tend to switch between the two word order (N+V vs. V+N) frequently, though the written form (N+V) is strictly followed. Therefore, the facilitation of processing in spite of the irregular word order presentation could stem from this informal use. Another explanation could be that collocational priming in Turkish is bidirectional based on the significant correlations between the mean response time and the ΔP values in both directions.

4.1. Regression Results

According to the results of the regression, two significant predictors of the mean response time in the experiment were part of speech and target word frequency. The priming experiment revealed that the subjects of the study responded considerably faster to the V+N lexical items compared to the collocations in ADJ+N and the results of the regression indicating part of speech as a significant indicator of response duration seem to be in line with that finding. Though both part of speech categories reflected significant priming effects, the gap between the mean response times of V+N collocations and non-collocations (23.2 milliseconds) is comparatively bigger than the difference between the corresponding mean response times of ADJ+N combinations (13.0 milliseconds), leading to an assumption that nouns are processed faster when they are primed by a verb rather than an adjective in the Turkish language.

There are some explanations in the literature regarding the faster response times of V+N collocational items than ADJ+N lexical combinations, though they are not conclusive and further evidence is needed. Approaching the issue from a generative perspective, Wolter and Gyllstad (2013) think that verbs are represented in higher nodes and as the head node in our internal grammar structure

mechanism, which could indicate they are processed first and faster than adjectives that are processed as an integral part of an adjectival phrase. This view has its roots in Generative Linguistics perspective and what the researchers assert is that this phenomenon could be also valid from a usage-based language approach. They further claim that faster V+N collocational processing is possibly due to the fact that verbs are entrenched as the most meaningful units of a constituent and because they are generally more concrete and salient, they bear stronger links with their neighbouring nouns.

Another issue that needs to be emphasized is that the current research selected the lexical items with a specific purpose in mind, which was a cross-linguistic collocational priming experiment as the following step. The collocations exploited in the monolingual priming experiment were chosen among the lexical members with no case marking in order to avoid any misleading results. For instance, during trimming and frequency measuring process, the verb in the collocation *karar vermek* – "make a decision" was not lemmatized, and so forms like *vermesi* (3rd person singular), *vermen* (2nd person singular), *vermeden* (without making), etc. were ignored, which could have made a difference in the processing durations and the fact that no inflected forms were used might have resulted in faster response times for the collocational items. However, as the adjectives are not inflected in Turkish, the same situation might not have been possible for ADJ+N collocations, which could have resulted in the different response times between the two groups of word pairs.

The second significant predictor of mean response time in the priming study was target word frequency and it needs further investigation. Although one may think that the effect of frequency of the target word on the lexical decision is an expected result, the fact that single word frequency is still playing a role while processing collocations, particularly when there is evidence that priming is occurring may mean more than the expected finding. To be more precise, it may mean that single word frequency is still helping with the processing of collocations as well as the collocational frequency. There is a common belief and empirical evidence that collocational items (formulaic phrases) in general are stored as chunks in the mental lexicon and when native speakers produce the language, they do not need to retrieve those lexical units separately because they are already activated as a whole, processed holistically and this is what facilitates spontaneous speech and how fluency is achieved (Schmitt, 2010). However, the results of this study show that not only the collocational frequency but also the frequency of the lexical items seem to be responsible for the speed of lexical processing. (Wray, 2012) summarizes some of the studies (e.g. Conklin & Schmitt, 2008) claiming a holistic storage of formulaic language. She questions the reasons of processing advantage and discusses the effect of repeated use on fused word strings before underlining the necessity to do interdisciplinary research for stronger evidence to answer all these questions.

4.2. Correlation Results

In addition to the regression results, the correlations computed to find possible relationships indicated that the mean response times and target word frequency as well as the association measures (*t*-score, ΔP in both directions, and MI) correlated negatively, which was interpreted as a clear indication that frequency is playing a critical role in how collocations are processed in Turkish. It may further be claimed that the more frequent a collocational item is, the stronger priming effect it has or in other words, the faster it is processed.

Something that needs attention is the fact that one of the association measures exploited in this study, MI value, did not reflect a strong correlation although it was significant, which was at odds with some other research (e.g. Wolter and Yamashita, 2017). This finding itself could mean that due to its possible flaws, which were discussed in some earlier research, MI value as a frequency dimension by itself is not good at predicting collocational processing speed or there is a weak relationship between

the MI value, which measures effect size and sensitive to low frequency words, and collocational priming on the whole and the processing speed in a lexical decision task investigating collocations, in particular. As previous research also states the MI value is prone to mislead research results aiming at frequency as the core investigation and should be supported by other association measures, such as *t*-score (prioritizes adjusted frequency), ΔP (prioritizes directionality), log dice (prioritizes exclusivity) etc. to get a clearer picture (Gablasova, Brenzina and Mcenery, 2017). Another reason why the MI value did not reflect strong correlations could be the nature of the preferred lexical items. The fact that they were very commonly used word combinations in everyday language and consisted of very high frequency lexical members could have resulted in low MI scores, which might not have reflected the psychological reality of the collocations in terms of the participants' own experiences.

One last thing to discuss for the correlation analysis is the fact that ΔP value in both directions revealed significant negative correlations, though the $\Delta P_{2|1}$ one is weak, which could indicate a bidirectional relationship between the members of the collocational items and the mean response times in the lexical decision task of the priming experiment. In other words, the higher the ΔP values of the collocations for either direction were, the faster the participants responded to the lexical items and a stronger priming effect was observed. To exemplify, the effect of the word *soğuk* – "cold" on the word *savaş* - "war" was as important for the processing durations as the effect of the word "war" on the word "cold" in ADJ+N combinations and the same influence can be seen in V+N combinations, such as *dikkat* – "attention" and *etmek* – "pay".

5. Conclusions

On the whole, the priming effect observed based on the findings of this study seems to be in accordance with Hoey's (2005) and Durrant and Doherty's (2010) claims about collocational priming underlining the importance of frequency in collocational processing. The fact that Hoey's findings are consolidated by means of a morphologically different language, Turkish, makes his remarks more reliable and generalizable. Further research taking case marking into account in Turkish is needed to draw stronger conclusions about agglutinative languages, though.

As to a mental lexicon model accounting for the collocational priming phenomenon as well as semantic, orthographic and phonological aspects of lexical processing, The Spreading Activation Model (Collins and Loftus, 1975) can be seen as the best fitting framework emphasizing the activation of semantically related nodes as well as collocational items when a certain word is seen or heard by a language user. To be more precise, when a prime is presented (e.g. *sağanak*-"heavy"), the activation spreads to its collocate (*yağmur*-"rain") and facilitates its processing as well as some semantically related items, such as "light", "weight" etc. This spreading activation could be influenced by the salience and frequency of those single lexical items in addition to their collocational association strength. Salience and frequency are two important aspects of lexical processing underlined by cognitive linguists (Tomasello, 2003) as they play an important role in how entrenched single words or word combinations are in the mental lexicon and how often language users are exposed to them in their everyday life.

Figure 2 shows a sample lexical organization network illustrating the spreading activation of semantically related and collocational items, which can be regarded as an extension to the Revised Spreading Activation Model by Bock and Levelt (1994). A similar cross-linguistic form of this model was proposed by Wolter and Yamashita (2014). Concepts are displayed in capital letters, whereas the lexical units are in small letters. Two-way arrows stand for possible bidirectional interaction and one-way arrows reflect the supposed direction of the lexical spreading. The activation of certain concepts

is assumed to trigger the lexical items related to that concept (semantic or collocational in this case) together with the corresponding conceptual domains. The activation seems to take place both at the syntagmatic level as well as paradigmatic level in the proposed lexical organization framework and the strength of the links between the lexical units appear to be influenced by the frequency of the lexical units and the collocations. This must be seen as one layer of the lexical activation and access procedure. Different layers including phonetics, morphology and orthography can be added; however, they are not the main focus of the current research and needs to be addressed in a separate study.

It should also be noted that the proposed framework is nothing more than an assumption based on the results of a single research study and more empirical studies are required for a generalizable and multi-layered depiction of the internal lexicon at the lexical activation and access level, in particular.

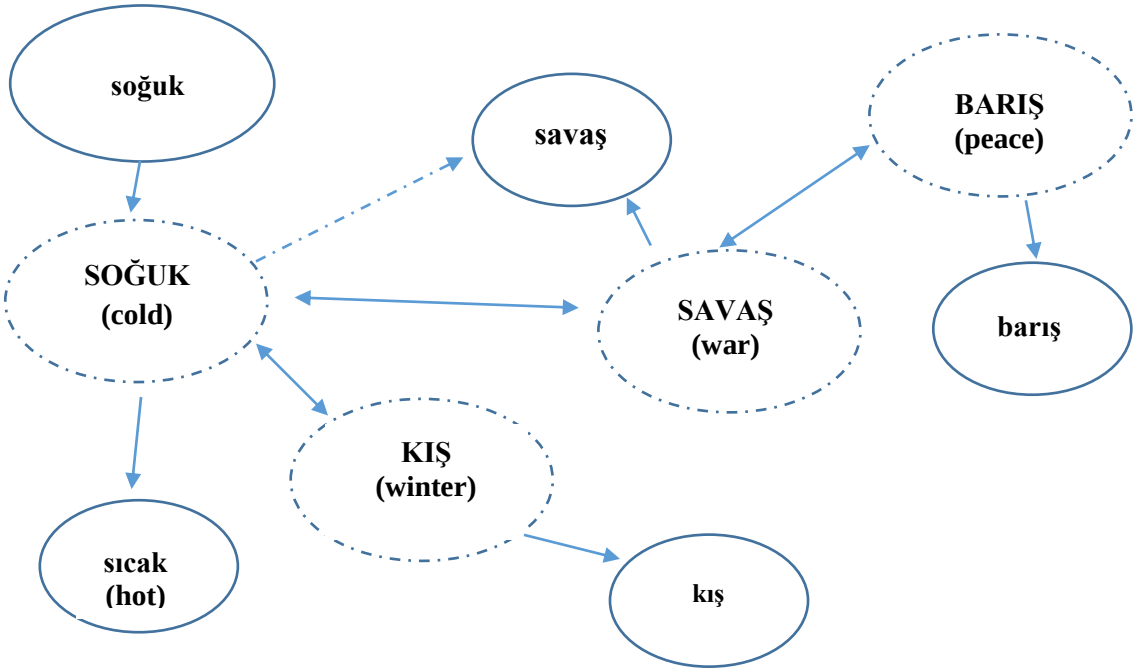


Figure 2. Proposed Lexical Organization in the Mental Lexicon

The framework proposed based on the assumptions of the current study needs further evidence to confirm collocational spreading activation by means of different cognitive methodologies, such as eye tracking (see Roberts and Siyanova-Chanturia, 2013; Carrol and Conklin, 2014 for a review on the use of eye-tracking to investigate lexical processing) and neuroimaging (see Henson, 2003 for a review of neuroimaging studies of priming). Until then, the idea of collocational spreading activation must be addressed tentatively. In addition, the issue of collocational priming, its psycholinguistic reality and its role in the organization of the internal lexicon, in particular needs further investigation from the glasses of morphologically different languages. This study focusing on the collocational priming and the effect of frequency on this phenomenon in Turkish could be regarded as a stepping-stone and aims to arouse more interest in lexical studies in Turkish.

6. Limitations and Suggestions for Further Research

Initially, it must be stated that the lack of lemmatization can be seen as a flaw of this study since the integration of all the inflections of a verb or a noun in Turkish could indicate a more thorough analysis of the situation and it should be applied in future work. For instance, a collocation in Turkish like *karar vermek*-"make a decision" can have many forms depending on the subject of the sentence, for instance. It may take the form *kararını vermek*-"make his decision", *karar vermesi*-"making a decision", which could make a difference in the processing times of the word pairs in a priming experiment. In addition, if all the lemmas of each word are taken into account while measuring frequency, it is likely to reflect the overall effect of frequency on processing times from a different angle. Furthermore, different forms of a word could prime different lexical items. To exemplify, if the bare form *okul*-"school" is used as a prime word, it is likely to prime a noun *önlüğü*-"uniform" in Turkish. However, if the inflected form *okula*-"to the school" is used, the verb *gitmek*-"go" seems more likely to be primed.

It should also be noted that lemmatization was omitted in this research mainly due to a lack of lemmatized search option in the Turkish National Corpus (TNC), which made the process of classification and integration of every inflected form challenging and time-consuming. The researcher had to make a decision owing to the time constraints.

Furthermore, some methodological extensions can be considered. For instance, a different SOA (Stimulus Onset Asynchrony) may indicate alternative results and the comparison between the priming experiments with different SOAs can suggest important interpretations for automatic and strategic priming paradigms and certain underpinnings are already existent in the priming literature. To be more precise, considering layout of this experiment, 50 milliseconds rather than 100 milliseconds could have made a difference in terms of the priming effect. It would have been possible to claim that even under masked priming conditions, which is claimed to occur in 50 milliseconds or less (Altarriba and Basnight-Brown, 2007) there was collocational priming in Turkish. In future research, the results of collocational priming experiments with both SOAs can be compared to analyse the possible difference and explore the influence of prime word duration in collocational priming, if any.

One of the extensions the writers of this research study are willing to make in their upcoming research is the inclusion of the lexical transparency into the regression model as a new and promising independent variable.

Acknowledgements

This research study was supported by TUBITAK 2214-A PhD Research Grant (Project ID-1059B141600068)

References

Aksan, Y. et al. (2012). *Construction of the Turkish National Corpus (TNC)*. In Proceedings of the Eight International Conference on Language Resources and Evaluation (LREC 2012). İstanbul. Turkiye. http://www.lrec-conf.org/proceedings/lrec2012/papers.html

Altarriba, J., & Basnight-Brown, D. M. (2007). Methodological considerations in performing semantic- and translation-priming experiments across languages. *Behavior Research Methods, 39*(1), 1–18. http://doi.org/10.3758/BF03192839

Barfield, A. & Gyllstad, H. (2009). Introduction: Researching L2 collocation knowledge and development. In A. Barfield & H. Gyllstad (Eds.), *Researching collocations in another language: Multiple interpretations* (pp. 1 – 18). Basingstoke, UK: Palgrave Macmillan.

Barlow M. and S. Kemmer (eds). (2000). *Usage- Based Models of Language*. CSLI Publications.

Benson, M., Benson, E., & Ilson, R. (1986). *Lexicographic description of English*. Amsterdam: John Benjamins.

Bock, K., & Levelt, W. (1994). *Language production: Grammatical encoding. In M. A. Gernsbacher (Ed.), Handbook of psycholinguistics*. (pp. 945-984). San Diego, CA US: Academic Press.

Bybee, J. (2005). *From usage to grammar: The mind's response to repetition*. Paper presented at the Linguistic Society of America, Oakland, CA.

Carrol, G., & Conklin, K. (2014). Eye-tracking multi-word units: some methodological questions. *Journal of Eye Movement Research, 7*(5), 1–11.

Chomsky, Noam (1956). "Three models for the description of language". *IRE Transactions on Information Theory* **2** (3): 113–124.

Collins, A. M., & Loftus, E. F. (1975). A spreading-activation theory of semantic processing. *Psychological Review*, 82(6), 407-428.

Conklin K. & Schmitt N. (2008) Formulaic sequences: Are they processed more quickly than nonformulaic language by native and nonnative speakers? *Applied Linguistics* 29: 72–89.

Cowie, A.P. (1994). Phraseology. In R. E. Asher (ed.), *The Encyclopedia of Language and Linguistics*, 3168-3171. Oxford: Pergamon.

Cruse, D.A. (2000) *Meaning in Language. An Introduction to Semantics and Pragmatics*. Oxford: Oxford University Press.

Davies, Mark. (2008-) *The Corpus of Contemporary American English: 520 million words, 1990-present*. Available online at http://corpus.byu.edu/coca/.

de Groot, A. M. B. (1984). Primed lexical decision: Combined effects of the proportion of related prime-target pairs and the stimulus-onset asynchrony of prime and target. *Quarterly Journal of Experimental Psychology, 36A,* 253–280.

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.

Durrant, P., & Schmitt, N. (2010). Adult learners' retention of collocations from exposure. *Second Language Research*, *26*(2), 163–188. http://doi.org/10.1177/0267658309349431

Durrant, P. (2014). Corpus frequency and second language learners' knowledge of collocations. *International Journal of Corpus Linguistics* 19(4). 443–477.

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

Ellis, N. C., & Simpson-Vlach, R. (2009). Formulaic language in native speakers: Triangulating psycholinguistics, corpus linguistics, and education. *Corpus Linguistics and Linguistic Theory*, *5*(1), 61–78.

Fan, M. (2009). An exploratory study of collocational use by ESL students - A task based approach. *System*, *37*(1), 110–123. http://doi.org/10.1016/j.system.2008.06.004

Firth, J. R. (1957). A synopsis of linguistic theory, 1930–1955. In F. R. Palmer (Ed.), *Selected papers of J.R. Firth*. London: Longmans.

Forster, K. I., & Forster, J. C. (2003). DMDX: A windows display program with millisecond accuracy. *Behavior Research Methods, Instruments, & Computers*, *35*, 116-124.

Gablasova, D., Brezina, V., & Mcenery, T. (2017). Collocations in Corpus-Based Language Learning Research : Identifying , Comparing , and Interpreting the Evidence. *Language Learning*, (January), 1–25.

Goldberg, A. (2006). *Constructions at work: The nature of generalization in language*. Oxford: Oxford University Press.

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.

Henriksen, B. (2013). Research on L2 learners ' collocational competence and development – a progress report. *EUROSLA Monographs Series*, *2*, 29–56.

Henson, R. N. A. (2003). Neuroimaging studies of priming. *Progress in Neurobiology*, *70*(1), 53–81.

Hoey, M. (2005). *Lexical Priming: A New Theory of Words and Language*. Routledge.

Howarth, P. (1998). Phraseology and second language proficiency. *Applied Linguistics*, 19(1), 24–44.

Hunston, S. & Francis, G. (2000). *Pattern Grammar. A Corpus-Driven Approach to the Lexical Grammar of English*. Amsterdam & Philadelphia: Benjamins.

Jiang, N. (2000). Lexical representation and development in a second language. *Applied Linguistics*, *21*(1), 47–77.

Jiang, N. (2012). *Conducting reaction time research in second language studies*. New York: Routledge.

Kjellmer, G. (1984). Some thoughts on collocational distinctiveness. In Aarts, J. and Meijs, W. (eds.), *Corpus Linguistics*, 163-171. Amsterdam: Rodopi.

Kjellmer, G. (1987). Aspects of English Collocations. In Meijs, W. (ed.), *Corpus Linguistics and Beyond*, 133-140. Amsterdam: Rodopi.

McKoon, G., & Ratcliff, R. (1992). Spreading activation versus compound cue accounts of priming: Mediated priming revisited. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 18(6), 1155-1172.

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

Roberts, L., & Siyanova-Chanturia, A. (2013). Using Eye-Tracking To Investigate Topics in L2 Acquisition and L2 Processing. *Studies in Second Language Acquisition*, *35*(2), 213–235.

Schmitt, N. (2010). *Researching Vocabulary: A Vocabulary Research Manual*. Palgrave Press.

Sinclair, J. M. (1987). Collocation: a progress report. In R. Steele & T. Threadgold (Eds.), *Language topics: Essays in honour of Michael Halliday* (Vol. 2, pp. 319-331). Amsterdam: John Benjamins.

Sinclair, J. (1991). *Corpus, Concordance, Collocation*. Oxford University Press.

Siyanova, A., & Schmitt, N. (2008). L2 learner production and processing of collocation: A multi-study perspective. *The Canadian Modern Language Review, 64*(3), 429-458.

Staples, S., Egbert, J., Biber, D., & McClair, A. (2013). Formulaic sequences and EAP writing development: Lexical bundles in the TOEFL iBT writing section. *Journal of English for Academic Purposes*, *12*(3), 214-225.

Tomasello, M. (2003). Constructing a language. Boston, MA: Harvard University Press.

Wolter, B. (2006). Lexical Network Structures and L2 Vocabulary Acquisition: The Role of L1 Lexical/Conceptual Knowledge. *Applied Linguistics*, *27*(4), 741–747.

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.

Wolter, B., & Gyllstad, H. (2013). Frequency of Input and L2 Collocational Processing. *Studies in Second Language Acquisition*, *35*, 451–482.

Wolter, B., & Yamashita, J. (2014). Processing collocations in a second language: A case of first language activation? *Applied Psycholinguistics*, *1*(29), 1–29.

Wolter, B., & Yamashita, J. (2017). Word Frequency, Collocational Frequency, L1 Congruency, and Proficiency in L2 Collocational Processing. *Studies in Second Language Acquisition*, 1–22. http://doi.org/10.1017/S0272263117000237

Wray, A. (2002). *Formulaic Language and the Lexicon*. Cambridge University Press.

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

Wray, A. (2012). What Do We (Think We) Know About Formulaic Language? An Evaluation of the Current State of Play. *Annual Review of Applied Linguistics*, *32*, 231–254.

Appendix A

A.1. VERB+NOUN

Turkish V+N Collocations	**English Translations**	**Turkish V+N Non-collocations**	**Direct English translations**
hata yapmak	make a mistake	*hata almak*	take mistake
izin vermek	give permission	*izin gitmek*	go permission
keyif almak	take pleasure	*keyif görmek*	see pleasure
huzur bulmak	find solace	*huzur bakmak*	look for solace
şefkat göstermek	show affection	*şefkat öğrenmek*	learn affection
nefes almak	take breath	*nefes yapmak*	make breath
çözüm bulmak	find a solution	*çözüm bilmek*	know solution
cinayet işlemek	commit murder	*cinayet bağırmak*	shout murder
öncelik vermek	give priority	*öncelik gitmek*	go priority
keşif yapmak	make a discovery	*keşif almak*	buy discovery
ipucu bulmak	find a clue	*ipucu bakmak*	look at clue
kalp kırmak	break heart	*kalp silmek*	erase heart
ateş açmak	open fire	*ateş tutmak*	keep fire
zafer kazanmak	win a victory	*zafer tutmak*	keep victory
zaman geçirmek	pass time	*zaman kurtarmak*	save time
karar vermek	make a decision	*karar gitmek*	go decision
dikkat etmek	pay attention	*dikkat yapmak*	make attention

şüphe uyandırmak	cast doubt	*şüphe kızdırmak*	annoy doubt
iflas etmek	go bankrupt	*iflas olmak*	be bankrupt
ara vermek	take a break	*ara görmek*	see break
ihtiyaç duymak	feel the need	*ihtiyaç sormak*	ask need
baskı yapmak	put pressure	*baskı etmek*	do pressure
kilo vermek	lose weight	*kilo görmek*	see weight
ziyaret etmek	pay a visit	*ziyaret olmak*	be visit
ışık tutmak	shed light	*ışık koymak*	put light
örnek olmak	set an example	*örnek etmek*	do example
sakal bırakmak	grow beard	*sakal görüşmek*	discuss beard
kaza yapmak	have an accident	*kaza etmek*	do accident
vurgu yapmak	place emphasis	*vurgu olmak*	be emphasis
sır saklamak	keep a secret	*sır götürmek*	get secret

A.2. ADJ+NOUN

Turkish ADJ+N Collocations	**English Translations**	**Turkish ADJ+N Non-collocations**	**Direct English translations**
derin uyku	deep sleep	*gizli uyku*	secret sleep
soğuk savaş	cold war	*uzak savaş*	far war
dış dünya	outside world	*geç dünya*	late world
kuvvetli delil	strong evidence	*şiddetli delil*	heavy evidence
çıplak göz	naked eye	*yapay göz*	artificial eye
sıcak karşılama	warm welcome	*mevcut karşılama*	current welcome
acı son	bitter end	*hoş son*	nice end
ateşli tartışma	heated debate	*şanslı tartışma*	lucky debate
zengin tarih	rich history	*sayılı tarih*	limited history
altın çağ	golden age	*kesin çağ*	certain age
orta sınıf	middle class	*ağır sınıf*	heavy class
karşıt görüş	opposing view	*neşeli görüş*	happy view
yüksek mahkeme	high court	*güzel mahkeme*	beautiful court
ölümsüz aşk	undying love	*çelimsiz aşk*	thin love
beyaz yalan	white lie	*siyah yalan*	black lie
açık fikir	open mind	*temel fikir*	basic mind
uzun vade	long run	*açık vade*	open run
sağanak yağmur	heavy rain	*gururlu yağmur*	proud rain
yoğun duman	thick smoke	*hızlı duman*	fast smoke
kabarık saç	wiry hair	*endişeli saç*	worried hair
keskin koku	strong smell	*parlak koku*	shiny smell
takma diş	false tooth	*sisli diş*	foggy tooth
koyu kahve	strong coffee	*adil kahve*	fair coffee
alkolsüz içki	soft drink	*renksiz içki*	colorless drink
itici güç	driving force	*nazik güç*	kind force
yüksek bina	tall building	*ciddi bina*	serious building
büyük başarı	high achievement	*doğru başarı*	correct achievement
sert düşüş	sharp fall	*ucuz düşüş*	cheap fall
köklü değişiklik	drastic change	*kızgın değişiklik*	annoyed change
tam yetki	free rein	*az yetki*	few rein

Appendix B

How t, MI, and Delta P scores are computed

The formula which COCA employed to compute the MI score indicating how strongly related word pairs are is as follows;

*"MI = log((AB * sizeCorpus)/(A * B * span))/log(2)"*

AB = frequency of collocations (eg. "heavy" used in front of the noun "rain")
sizeCorpus = how big the corpus is (# word)
A = frequency of node word (eg. "heavy")
B = frequency of collocate (eg. "rain")
span = span of words (note: 4 Left and 4 Right = 8 word span total was used)
log(2) = the log10 of the number 2

The calculation indicates that the bigger the MI value, the stronger the relationship between the lexical items. As stated earlier, word pairs with 3.0 or higher MI were accepted as valid and included in the study since 3.0 is claimed enough to state that a word pair does not co-occur randomly (Durrant and Doherty, 2010).

The other association measure is computed as follows:

$$t\text{-}score = \frac{O-E}{\sqrt{O}}$$

O: observed frequency of the collocation
E: expected frequency of the collocation

After the observed frequency is subtracted by the expected frequency, the result is divided by the standard deviation. Durrant and Doherty (2010) state 2.0 or higher *t* values show a statistically significant difference and is sufficient to claim that a word pair is a collocation.

Gries (2013) thinks that directional measures of collocational frequency have some drawbacks and as he claims ΔP succeeds in addressing these flaws by normalizing conditional probabilities, which makes ΔP a psychologically and psycholinguistically realistic measure.

The complementary association measure, ΔP included later in the study is computed as follows:

$\Delta P_{2|1}$ = p ($word_2$ | $word_1$ = present) − p ($word_2$ | $word_1$ = absent) = (a÷a+b) – (c÷ c + d)
$\Delta P_{1|2}$ = p ($word_1$ | $word_2$ = present) − p ($word_1$ | $word_2$ = absent) = (a÷a+c) – (b÷ b + d)

A sample calculation of ΔP is as follows:

Co-occurrence of the word "of course" in the spoken component of British National Corpus

	course: present	**course: absent**	**Totals**
of: present	5610	168.938	174.548
of: absent	2257	10.223.063	10.235.320
Totals	7867	10.402.001	10.409.898

$$\Delta P_{2|1} = p(course \mid word_2 = of) - p(course \mid word_2 \neq of) = \frac{5610}{174548} - \frac{2257}{10235320} \approx 0.032$$

$$\Delta P_{1|2} = p(of \mid word_2 = course) - p(of \mid word_2 \neq course) = \frac{5610}{7867} - \frac{168938}{10402001} \approx 0.697$$

The numbers could basically show that the word "course" is a better cue to "of" than vice versa. Given that each association measure discussed so far has its plus and minuses, the current research included MI, *t*, and ΔP values in the analysis to explore a possible frequency effect in collocational priming.

Eşdizimli kelimelerde öncelemenin Türkçe bağlamında incelenmesi

Öz

Tekdillilerin zihin sözlüğünün nasıl şekillendiğini açıklamaya çalışan bir çok teşebbüs olmuştur ve şimdiye kadar önerilen her bir model kelime işlemleme sürecinin farklı bir boyutunu ele almıştır. Bu modellerin ortak noktası, betimlemelerinin eşdizimli kelimeler gibi kelime gruplarını göz ardı etmeleri ve yaklaşımlarında paradigmatic ilişkilerin öne çıkmasıdır. Hoey (2005) tarafından ortaya atılan Kelimelerde Önceleme Teorisi, eşdizimli kelimelerin zihin sözlüğünde işlemlenmesine ve bu işlemlenmenin yaratıcı dil üretimimiz için olan önemine bilişsel ve psikodilbilimsel açıdan ışık tutmaya çalışmaktadır. Birçok psikodilbilimsel araştırma Hoey'in teorisini İngiliz dili bağlamında test etmiştir. Mevcut araştırma ise bu alanda yapılan çalışmaların kapsamını genişletmiş ve eşdizimli kelimelerde önceleme olgusunu Türk dili bağlamında incelemiştir. Ayrıca, sıklığın ve sözcük türünün tartışılan süreçteki muhtemel etkisi, önceleme deneyindeki sözcük karar verme süreleri ve söz konusu bağımsız değişkenler arasındaki ilişki incelenelerek mercek altına alınmıştır. Hoey'in iddialarını doğrular nitelikte olan bulgular, Türkçe anadil konuşucuları için önemli bir önceleme etkisini işaret etmektedir. Regresyon analizi göstermiştir ki, sıklık ve sözcük türü işlemleme süresinin önemli bir kestiricisidir. Son olarak, sözcük karar süreleri ile eş dizimli kelimelerde sıklık arasında güçlü bir korelasyon tespit edilmiştir. Araştırmanın bulgularına dayanarak mütevazi bir zihin sözlüğü modeli ortaya konmuştur.

Anahtar sözcükler: zihin sözlüğü; eşdizimli kelimelerde önceleme; sıklık

AUTHOR BIODATA

Hakan Cangır, PhD candidate, is a lecturer at Ankara University, School of Foreign Languages in Turkey. He teaches EFL courses at the preparatory school and pursues his PhD research at Hacettepe University, Linguistics Department. His research interests are lexical processing in L1 and L2, the organization of the mental lexicon and the use and building of corpora for educational purposes.

S. Nalan Büyükkantarcıoğlu, PhD, is a full professor at Hacettepe University, Department of English Linguistics. Her research mainly focuses on discourse analysis, cognitive linguistics, sociolinguistics and translation studies.

Philip Durrant, PhD, is a senior lecturer in language education at the University of Exeter. His research mainly focuses on corpus linguistics and phraseology.