The Processing of Speech Formulas on Turkish: A Masked Priming Study*

Dilek Göymen[1], Mehmet Aygüneş[2]

ORCID: [1]0000-0002-5507-0933, [2]0000-0002-0327-6905

[1]İstinye University, Foreign Languages Department, Zeytinburnu, İstanbul
[2]İstanbul University, Department of Linguistics, Fatih, İstanbul

[1]dilek.goymen@istinye.edu.tr, [2]mehmet.aygunes@istanbul.edu.tr

(Received 10 Haziran 2020; accepted 17 Aralık 2020)

ABSTRACT: Studies have indicated that formulaic sequences are processed significantly faster than newly created phrases; however, the source of this processing advantage has not been sufficiently investigated in the literature. The Holistic Approach justifies this processing advantage for formulaic sequences with the argument that they are processed and stored as single units without being decomposed into their constituents. On the contrary, Distributed Representation argues against holistic processing. It proposes instead that formulaic sequences are processed through their parts as in novel non-formulaic phrases. Their constituents form a mutual association in the sense that the mental activation of a component part activates the other, thus leading to faster processing. The present study reports findings from a masked priming experiment investigating Turkish speech formulas' online processing in native processing. Results show that speech formulas and their matched novel phrases are processed similarly, as evidenced by no significant difference in reaction times. These findings support Distributed Representation in the processing of formulaic sequences. Results also suggest that non-transparent formulas are processed more slowly than transparent ones.

Keywords: speech formulas, processing of formulaic sentences, semantic transparency, light verbs, masked priming

* This article was developed from the 1st author's master's thesis, which was conducted under the 2nd author's supervision. This project was also supported by the Scientific Research Projects (BAP) Coordination Unit of Istanbul University (Project ID: SDP-2016-21580). The authors wish to thank Prof. Dr. Ayşe Gürel and Assoc. Prof. Dr. Mehmet Gürlek for their valuable criticisms and suggestions. All errors belong to authors.

http://dx.doi.org/10.18492/dad.750788

Dilbilim Araştırmaları Dergisi, 2020/2, 207-230.

Türkçede Konuşma Kalıplarının İşlemlenmesi: Maskelenmiş Çağrıştırma Çalışması

Alanyazında birçok çalışma kalıp ifadelerin diğer yapılardan daha hızlı işlemlendiğini göstermesine rağmen, bu işlemleme avantajının kaynağı üzerine oldukça sınırlı sayıda çalışma bulunmaktadır. Bütüncül Yaklaşım, bu avantajı kalıp ifadelerin parçalarına ayrıştırılmadan bir bütün olarak işlemlenmesi ile açıklarken, kalıp ifadelerin diğer yapılar gibi parçaları aracılığıyla işlemlendiğini savunan Dağıtılmış Simgeleme, bunu ifadeleri oluşturan sözcüklerin zihinsel bir bağ geliştirerek birbirini etkinleştirmesi ile açıklamaktadır. Maskelenmiş çağrıştırma yöntemiyle Türkçe konuşma kalıplarının işlemlenmesinin incelendiği bu çalışmada, kalıp ifadelerin işlemlenmesinde bu görüşlerden hangisinin geçerli olduğu sorgulanmıştır. Tepki süreleri incelendiğinde, Dağıtılmış Simgeleme görüşünün öne sürdüğü gibi konuşma kalıpları ile kalıplaşmamış ifadeler arasında anlamlı bir fark tespit edilmemiştir. Ayrıca, anlamsal açıdan saydam olan konuşma kalıplarının saydam olmayanlara göre daha hızlı işlemlendiği tespit edilirken, katkısız eylem içermenin anlamlı fark yaratmadığı görülmüştür.

Anahtar sözcükler: konuşma kalıpları, kalıp ifadelerin işlemlenmesi, anlamsal saydamlık, katkısız eylemler, maskelenmiş çağrıştırma

1 Introduction

Formulaic sequences are frequently occurring, multi-worded sequences with specific functions or meanings such as idioms, proverbs, collocations, and speech formulas (Wray, 2002). Corpus-based studies have shown that formulaic language contributes substantially to written and spoken discourse (Erman & Warren, 2000). Formulaic sequences have also been approached from the perspective of language acquisition, and studies (Qin, 2004; Pérez-Llantada, 2014) find that the use of formulaic sequences by the second language (L2) speakers deviates from native speakers due to the influence of their native language. They use a lower number and variety of formulaic sequences than native speakers. Moreover, appropriate and frequent use of formulaic language is regarded as necessary to attain native-like language proficiency in second language acquisition (Wray, 2002; Ortaçtepe, 2013).

In recent years, studies also have investigated the processing and representation of formulaic sequences using behavioral and cognitive methodologies such as auditory lexical decision tasks (Sosa & MacFarlane, 2002; Kapatsinski & Radicke, 2009), self-paced reading paradigm (Conklin & Schmitt, 2008), grammatical judgment tests (Isobe, 2011), priming experiment (Cangır et al., 2017), eye-movement paradigms (Underwood et al., 2004; Siyanova-Chanturia et al., 2011; Arıca-Akkök & Uzun, 2018), and Event-Related Potentials (ERP) (Tremblay & Baayen, 2010; Siyanova-Chanturia et al.,

2017). These studies show that formulaic sequences are processed faster than the matched novel phrases.

There are three main views on the source of this advantage in the processing of formulaic sequences. The first of these views, the Holistic Approach, claims that formulaic sequences are processed and stored as a whole unit. Namely, it proposes that formulaic sequences are not decomposed into their parts. It suggests that this diminishes the mind's processing load, and thus these sequences are processed faster (Underwood et al., 2004; Tremblay & Baayen, 2010; Isobe, 2011). This Holistic Approach argument also supports dual-system models for language processing (Wray, 2008; Sinclair, 1991; Erman & Waren, 2000; Lamb, 1999), which argue that formulaic language is processed with a secondary system different from newly created utterances. These language processing models claim that there are two independent systems in language processing. One of them is responsible for the processing of grammatically generated new expression via decomposition. The other is responsible for processing formulaic sequences as a whole unit.

The Distributed Representation, which argues completely against the Holistic Approach, proposes that formulaic sequences are not processed as a whole unit, but rather the words that frequently occur together in these sequences develop a mental bond over time and thus mentally activate one another (Haselow, 2018; Siyanova-Chanturia, 2015). In other words, according to this approach, formulaic sequences are processed through their constituents.

Contrary to these two opposing views, the Continuum Processing Approach (Lamb, 1999; Kapatsinski & Radicke, 2009) claims that formulaic sequences' processing is a graded phenomenon. Lamb (1999) states that speakers combine words to construct phrases, and in this case, they are not formulaic. With subsequent use, they become lexicalized and remembered as a unit. Building connections between words to form formulaic phrases is gradual, and as a connection is used, it gets stronger. Namely, connections between words come in a continuum of degrees of strength, affecting how speakers process them.

In the literature, the number of studies investigating whether formulaic sequences are processed as a whole or decomposed into their constituents is quite limited. These studies also focus on only the Holistic Approach and Distributed Representation, not the Continuum Processing Approach. Namely, in the literature, the main discussion is made through the Holistic Approach and Distributed Representation. That is why this study did not aim to test the Continuum Approach, and we did not design the experiment set to test this, either.

1.1 Processing of Formulaic Language

Some studies on the processing of formulaic sequences associate these sequences' rapid processing with holistic processing, although they do not aim to directly question these hypotheses (Underwood et al., 2004; Isobe, 2011; Tremblay & Baayen, 2010).

In an eye-tracking study in which Underwood, Schmitt and Galpin (2004) examined the processing of the English formulaic sequences, 20 formulaic sequences and 20 matched novel expressions were presented to the participants in short stories, and the eye movements of the participants were recorded while reading the stories. The eye-tracking data analysis indicated that the participants made fewer and shorter fixations on the final words of the formulaic sequences; in other words, they read these sequences faster. They proposed that formulaic sequences were processed and represented in the lexicon as a whole since the earlier parts of these sequences allowed the participants to predict the final words.

In a grammatical judgement task, Isobe (2011) wanted Japanese English speakers to decide whether the word order of English formulaic and non-formulaic sequences was appropriate and recorded their reaction times. The analysis demonstrated that the participants responded to the formulaic sequences faster and with higher accuracy. These findings were interpreted as indicating that the formulaic sequences are processed and stored as single units.

Tremblay and Baayen (2010) investigated the processing of English four-word phrases via Event-Related Potentials (ERP). They presented six phrases to participants in random order and asked them to recall and type as many sequences as possible at the end of each trial. They reported that the formulaic four-word sequences, which had higher frequency and predictability, were recalled more than the non-formulaic sequences, and thus formulaic sequences were retrieved holistically.

Some studies on the processing of formulaic sequences (Cangır et al., 2017; Siyanova-Chanturia et al., 2017) support the Distributed Representation approach. Cangır et al. (2017) administered a lexical decision task using a masked priming paradigm to examine Turkish collocations' priming effect. They presented verbs and adjectives as primes and nouns as target items. The reaction time analysis revealed that nouns were processed faster when primed by their collocations in Turkish. They reported that its collocations were also mentally activated when a word was encountered, supporting the Distributed Representation. Likewise, in the study where the processing of English binominals was examined via ERP, they suggested that frequently occurring multi-word expressions eased the processing load and were previously activated in the mind of participants through their constituents (Siyanova-Chanturia et al., 2017).

Although these studies did not question whether formulaic sequences are processed as a single unit or decomposed into their constituents, they provided important data on formulaic language processing. In addition to these studies, there are three studies (Sosa & MacFarlane, 2002; Arnon & Cohen Priva, 2014; Kapatsinski & Radicke, 2009) that aim to reveal how formulaic sequences are processed by investigating the arguments of the Holistic Approach and Distributed Representation. In the first of these studies, Sosa and MacFarlane (2002) stated the reaction times given to the English word "of", which was a part of the formulaic sequence with the highest frequency, were significantly longer than those of lower frequency in auditory reaction time task. In other words, they reported that it took longer to perceive and respond to the particle "of", which is a part of formulaic sequences with high frequency. They interpreted this finding as follows; participants processed the formulaic sequences with high frequency as a single unit without decomposing them into constituents and did not access each component separately. However, since the participants were asked to respond to "of", namely as a part of these sequences, it required additional processing to decompose "unanalyzed single units" into constituents to access it, which led to longer reaction times. On the other hand, since the sequences with lower frequency were already processed by being decomposed into constituents, it was faster to access and respond to "of" without additional processing.

Kapatsinski and Radicke (2009) criticize this study (Sosa and MacFarlane, 2002) with the claim that the word "of" in high-frequency sequences is reduced and mostly articulated without the consonant in spoken English, which might make it take longer to be detected and responded to in auditory reaction time task. In other words, Kapatsinski and Radicke (2009) stated that the word "of" was not suitable for questioning these hypotheses, and they carried out a similar auditory reaction time task with the English word "up". Analysis of reaction times revealed that it took longer to detect and respond to the word "up" in the high-frequency and low-frequency sequences than in the medium-frequency ones. This result indicated that formulaic sequences with high frequency were processed and stored as a whole unit in the lexicon and supported Sosa and MacFarlane's (2002) findings. However, they emphasized that formulaic sequences were stored as a single unit in the lexicon only if they occur with a very high frequency. They also claimed that the component words in formulaic sequences fuse over time as they frequently occur together in an identical form, and their individual effects decrease in processing. In other words, this study supports Continuum Processing, primarily Cognitive Grammar (Lamb, 1999), by claiming that words frequently occurring together become formulaic over time.

Lastly, Arnon and Cohen Priva (2014) investigated the Holistic Approach's hypotheses by examining the effects of component word frequency on the phonetic duration of spontaneous speech sequences. They selected English formulaic sequences with different frequencies from a corpus. It contained

orthographical and phonetical transcriptions of phrases, and their phonetic durations were calculated. The study's findings demonstrated that constituent words' frequency affected the phonetic duration's length in the frequent formulaic sequences. Namely, although the frequent use of formulaic sequences increased the importance of multi-word information, it did not completely obliterate the effect of word information. They argued that constituent words were still accessed while formulaic sequences were processed as Distributed Representation suggests.

To sum up, processing studies (Underwood et al., 2004; Tremblay & Baayen, 2010; Isobe, 2011; Jeong & Jiang, 2019; Conklin & Schmitt, 2008; Carrol & Conklin, 2019; Siyanova-Chanturia et al., 2017; Carrol & Conklin, 2019; Jeong & Jiang, 2019) have indicated that formulaic sequences are processed faster than matched novel utterances. However, there is no consensus on the source of this processing advantage. It is impossible to form generalizable judgments since the number of studies questioning whether these sequences are processed as a single unit or decomposed into their parts is highly limited and have reported conflicting results. Therefore, it is frequently emphasized in the literature that new studies investigating this issue in different languages should be done because most studies have examined English formulaic sequences (Siyanova-Chantura & Martinez, 2014; Siyanova-Chanturia, 2015).

1.2 Factors in the Processing of Formulaic Sequences

It is suggested that factors such as frequency of occurrence, structural completeness, and semantic transparency play an essential role in processing formulaic sequences.

1.2.1 Frequency of occurrence

Frequently used multi-word phrases take a fixed form over time, so one of the conditions for accepting a multi-word phrase as a formulaic sequence is its frequent use. Usage-based theories on language also posit that a formulaic sequence frequency provides important information about its processing. The studies investigating this argument have suggested that sequences with a higher frequency are produced and processed faster (Arnon & Cohen Priva, 2013; Siyanova & Schmitt, 2008; Tremblay & Baayen, 2010).

Siyanova and Schmitt (2008) examined the processing of English adjective-noun collocations. 27 native speakers and 27 L2 speakers participated in the study. Participants were asked to rate the collocations based on their commonness in English. They were required to rate the collocations they saw on the screen as quickly as possible by pressing a button. RT analysis showed that native and L2 speakers responded more rapidly to frequent collocations than the

atypical ones. In other words, it was observed that all participants processed collocations with a higher frequency of occurrence faster than other expressions.

Tremblay and Baayen (2010) examined four-word English sequences processing from both a behavioral and an electrophysiological perspective. They presented sequences on the computer screen and asked participants to recall and type as many sequences as they could remember. Tremblay and Baayen (2010) argued that structures with a lower frequency of occurrence would attract more participants' attention and that more resources will be allocated to these sequences in working memory and predicted that these sequences would be recalled more. Contrary to these expectations, the results revealed that sequences with a higher frequency and, therefore, higher predictability were recalled more.

Arnon and Cohen Priva (2013) administered three different studies to examine the effect of multi-word frequency on phonetic duration. These studies revealed that durations are reduced in higher frequency sequences in both elicited and spontaneous speech. Therefore, they stated that sequences with higher frequency are produced and processed faster.

1.2.2 Structural completeness

While structurally complete formulaic sequences (e.g., *for example*) form a complete phrase, incomplete sequences (e.g., *the presence of a*) run across phrasal boundaries. Some studies (Jeong & Jiang, 2019; Tremblay & Baayen, 2010) suggest structurally complete formulaic sequences are processed as a whole without being decomposed into their constituents. Thus, they are processed faster.

Jeong and Jiang (2019) examined the processing of English complete and incomplete phrases by native and L2 speakers. They required participants to react to target words that are part of the phrases they saw on the screen and recorded their reaction times. RT analysis revealed that they responded to the target words, which are part of the structurally complete phrases, faster than the incomplete control items.

Similarly, Tremblay and Baayen (2010) examined the processing of 4-word expressions by native English speakers using electrophysiological and behavioral reaction time paradigms. They found that participants remembered 4-word structurally complete English expressions more than incomplete expressions. Therefore, they concluded that structurally complete phrases are processed faster, and this advantage is not seen in incomplete phrases.

To sum up, most of the studies conclude that the high frequency of occurrence and structural completeness give formulaic sequences processing advantages. For this reason, we selected Turkish speech formulas with high frequency and structural completeness for this study.

1.2.3 Semantic transparency

In addition to the frequency and structural completeness, semantic transparency is also a significant factor. Processing studies argue that non-transparent formulaic sequences are processed more analytically than transparent ones (Wood, 2015: 60). They claim that semantically transparent sequences are decomposed into their constituents while non-transparent figurative sequences are processed as a single unit without being decomposed. Although there is a consensus in the literature that the high frequency and structural completeness promote the rapid processing of formulaic sequences, the studies investigating the effect of semantic transparency have reported conflicting results. Some studies argued that semantically transparent and non-transparent sequences were processed similarly (Conklin & Schmitt, 2008; Siyanova-Chanturia et al., 2011). Others suggested that non-transparent metaphorical sequences were processed slower than transparent ones (Arıca-Akkök & Uzun, 2018).

Conklin and Schmitt (2008) investigated English idioms' processing by native and non-native speakers in a self-paced reading task. They found that both groups of participants processed idioms that were used idiomatically or literally faster than matched novel phrases. Namely, they concluded that semantic transparency did not affect the processing of formulaic sequences. In a similar study, Siyanova-Chanturia et al. (2011) examined English idioms' processing via the eye-tracking paradigm, and they obtained the same results from native speakers. However, they reported that non-native speakers processed metaphorical idioms slower than the literal ones.

On the other hand, Arıca-Akkök and Uzun (2018) questioned the Direct and Indirect Access Models in processing non-transparent Turkish phrases via an eye-tracking study. The Indirect Access Model posits that literal meaning is processed and analyzed first to access figurative meaning; thus, it takes longer to process non-transparent phrases. The Direct Access model argues against this argument and proposes that figurative meaning can be directly accessed without analyzing the phrases' literal meaning if sufficient contextual information is provided. The study found that participants made regressions when they met with figurative phrases, and this finding was considered to support the Indirect Access Model. In other words, it argued that non-transparent phrases were analyzed, and their literal meanings were activated first to access their figurative meanings, which led to the longer processing time for non-transparent phrases.

Therefore, studies have reported inconsistent results on whether semantic transparency, namely being semantically transparent or non-transparent, affects formulaic sequences processing.

1.2.4 Light verb constructions

Only one study examines the processing of Turkish formulaic sequences (Cangır et al. 2017). They examined the processing of Turkish collocations by native speakers. In this study, while creating an experiment list, they chose collocations without a case suffix with the concern that it might affect the study's result. In this study, with the same concern, we selected Turkish speech formulas without case suffixes. However, when we chose sequences without a case suffix, we found that a significant portion of the experiment list's verbs was light.

Although it is examined in the literature whether factors such as frequency of occurrence, structural completeness, and semantic transparency affect the processing of formulaic sequences, the presence of light verb constructions is not considered as a factor. A light verb (e.g., *et-, ol-,*) is a verb that does not contribute meaning to the phrases and cannot assign syntactic roles to words in sentences. They cannot stand in the sentence on their own and are mostly used in combination with nouns (e.g., yardım et, sakin ol), forming light verb constructions (LVC). In LVC, light verbs' primary role is to verbalize nouns and make it possible to add grammatical information such as tense, mood, and person (Akşehirli, 2013; Berk et al., 2018). For these reasons, in this study, it is considered that the presence of light verb constructions might affect the processing of Turkish speech formulas. Thus, we examined the presence of light verbs as an additional factor.

The present study aims to answer the following questions by conducting a lexical decision experiment using a masked priming paradigm. This study examined frequent Noun-Verb speech formulas, a common type of formulaic sequences in Turkish.

1. Are Turkish noun-verb speech formulas processed as a single unit or decomposed into their constituents? In other words, does their processing support the Holistic Approach or Distributed representation?
2. Does semantic transparency affect the processing of Turkish noun-verb speech formulas?
3. Does including light verb constructions affect the processing of Turkish noun-verb speech formulas?

2 Method

2.1 Participants

30 Turkish native speakers (19 female, mean age: 26.36, age range: 19-43; 11 male, mean age: 29.90, age range: 23-52) participated in the lexical decision experiment. None of the participants were bilingual, but they have an intermediate to advanced level of English. All the participants study at the

undergraduate level in various universities in Istanbul or have at least a bachelor's degree. All the participants had a normal or corrected vision, and they had never been diagnosed with any psychiatric disorders. Before the experiment, the participants read and signed the Informed Voluntary Consent Form approved by the Istanbul University Faculty of Humanities and Social Sciences Research Ethics Committee, stating that they voluntarily participated.

2.2 Materials

A total of 480 items were tested in the experiment. The test materials included 120 critical items and 360 fillers. Critical items consisted of speech formulas, non-formulaic sequences, and semantically anomalous expressions based on formulaicity to compare the processing of speech formulas and the matched non-formulaic sequences. In this way, we could investigate whether speech formulas were processed as a single unit or decomposed into their constituents. Speech formulas consisted of an equal number of semantically transparent and non-transparent sequences. We also balanced the test items in terms of the inclusion of light and non-light verbs because we aimed to investigate their effects on processing speech formulas. In line with these experimental conditions, two types of fillers were added for various purposes (see Table 1).

Table 1. Distribution of Items in the Experiment

Critical Items			N
Speech formulas	Transparent	Light verb	10
		Non-light verb	10
	Non-transparent	Light verb	10
		Non-light verb	10
Non-formulaic sequences			40
Anomalous expressions			40

Fillers			N
Binominals	Formulaic		40
	Non-formulaic		40
	Anomalous		40
Pseudowords			240
		Total	480

Therefore, the test materials were developed based on three variables: formulaicity, semantic transparency, and the inclusion of light verbs. Filler items were also added to the experiment for various purposes. The variables used in determining the sequences in the experiment are as follows:

2.2.1 Formulaicity

We examined various textbooks and dictionaries and formed a corpus of approximately 900 expressions to determine the 40 speech formulas included in the experiment. From this corpus, 40 formulas were selected based on various criteria and experiment variables such as semantic transparency, frequency, number of words, and syllables. All the formulas were noun-verb sequences, and the nouns consisted of two syllables without any affix. All the verbs were single syllable and inflected for second person singular and imperative mood (e.g., merak etme 'don't worry'). Additionally, we assured that all formulas had a Mutual Information (MI) score of at least 3.0 and a t-score of 2.0 based on the Turkish National Corpus (TNC) as a frequency measure (Aksan et al., 2012). In the literature, it is emphasized that for multi-word phrases to be accepted as formulaic sequences, they should have an MI score of at least 3.0 and a t-score of 2.0 (Hunston, 2002; Siyanova & Schmitt, 2008; Cangır et al., 2017). The mean MI score of 40 speech formulas in the test materials was 7.18 (range: 4.18-10.27), and the mean t-score was 6.54 (range: 2.82-16.98).

In addition to 40 speech formulas, we included 40 non-formulaic sequences with a low frequency and 40 semantically anomalous expressions in the experiment. In this way, we could make comparisons between the processing of formulaic and non-formulaic sequences. Non-formulaic sequences are grammatically generated novel expressions, and thus they are decomposed into their constituents while being processed. Therefore, these comparisons enabled us to investigate whether formulaic sequences were similarly decomposed into their components or processed as a single unit.

Non-formulaic sequences and semantically anomalous expressions were created by replacing the verbs in noun-verb sequences. While MI scores of non-formulaic sequences were below 3 and t-scores below 2, semantically anomalous expressions did not exist in the Turkish National Corpus (see Table 2).

Table 2. Experimental Conditions Based on Formulaicity

Experimental conditions	Prime	Target	N
Speech formulas	dikkat et	DİKKAT	40
Non-formulaic sequences	dikkat ver	DİKKAT	40
Semantically anomalous expressions	dikkat sor	DİKKAT	40
		Total	120

In the literature, it is suggested that keeping the target the same for different conditions and forming new primes for each condition is better (Jiang, 2012). That's why, in this masked priming experiment, we kept targets constant and replaced the verbs in primes. Furthermore, since we used the same target for three conditions, three counter-balanced lists were created so that a participant did not encounter the same target more than once in an experimental session. Thus, we divided test items into three counter-balanced lists. In each list, a target item appeared only once. Each list was applied to the participants in 3 different sessions at intervals of at least ten days.

2.2.2 Semantic transparency

Speech formulas consisted of 20 semantically transparent and 20 non-transparent formulaic sequences. The meaning of a transparent formula can be inferred from its constituents; namely, the meaning of the parts (e.g. soru 'question') was related to the meaning of the formula (e.g. soru sorma 'don't ask a question'). In non-transparent formulas (e.g. surat asma), constituents' meaning was not related to the overall meaning because they gained a metaphorical meaning.

To determine the semantic transparency of formulas, we modified and used the transparency judgment test developed by Uygun and Gürel (2017) in line with this study's purpose. We administered it to 52 Turkish speakers who did not participate in the priming experiment. They rated the semantic contribution of first and second constituents to the overall meaning on a 5-point scale (1: unrelated, 5: strongly related). The formulas with a mean score of 3.5 and above were accepted as transparent.

2.2.3 The inclusion of light verbs

Half of the 40 speech formulas included light verbs (e.g., et-, ol-), and 20 formulas contained non-light verbs (e.g., sor-, as-). Various criteria (Akşehirli, 2013; Berk et al., 2018) were taken into account in determining whether verbs in formulas were light or non-light. Speech formulas which included light or non-light verbs were as follows:

Table 3. Experimental Conditions Based on Semantic Transparency and Light Verb Constructions

Experimental Conditions		Example	N
Transparent	*Light verb*	merak etme	10
	Non-light verb	soru sorma	10
Non-transparent	*Light verb*	komik olma	10
	Non-light verb	surat asma	10
		Total	40

2.2.4 Fillers

We added two types of fillers in the experiment. The first of these, binominals (e.g., çatal kaşık 'fork spoon') were included in the test materials to prevent participants from developing a processing strategy during the experiment. Secondly, we developed pseudowords as a requirement of the lexical decision task. We considered that they complied with Turkish phonetics rules and had the same length as target words while forming these meaningless words. We did not analyze reaction times and responses to the fillers.

Binominals

We presented 40 speech formulas, 40 non-formulaic sequences, 40 semantically anomalous expressions to participants as prime stimuli (e.g., soru sorma), and the first words of these phrases as target stimuli (e.g., SORU). Although prime stimuli were shown only for 50 milliseconds, we aimed to prevent participants from unconsciously developing a strategy and focusing only on the first words. With this aim, we added 40 binominals to the test materials as fillers (e.g., çatal kaşık), and selected the second words of these expressions as target stimuli (e.g., KAŞIK). These binominals were compared in terms of length and frequency with other phrases. The same experimental conditions based on formulaicity were developed for these fillers (see Table 4). However, we excluded them from the analysis because they were not related to the study's research questions.

Table 4. Binominals as Fillers

Experimental conditions	Prime	Target	N
Formulaic	çatal kaşık	KAŞIK	40
Non-formulaic	bıçak kaşık	KAŞIK	40
Anomalous	yatak kaşık	KAŞIK	40
		Total	120

Pseudowords

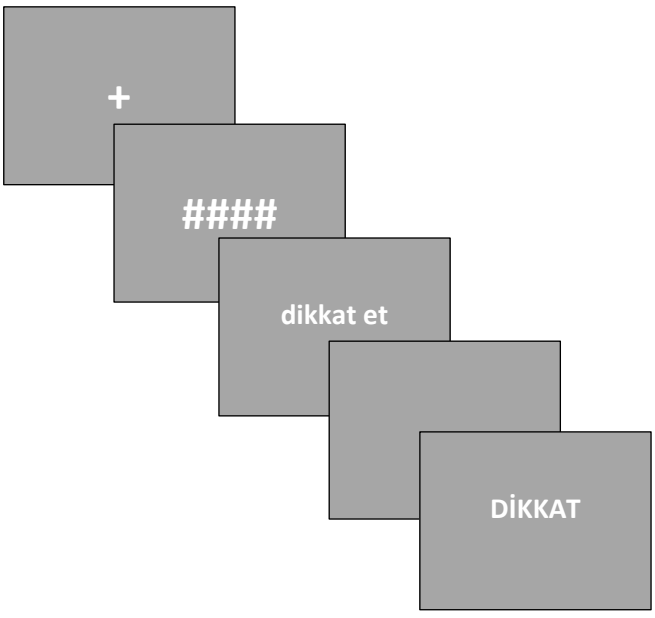
In the lexical decision task, the participants were asked to decide whether or not the string of letters on the screen was meaningful in Turkish. That's why we added meaningless pseudowords (e.g., BEREK) to the test materials, and most pseudowords in this experiment were developed using software called Wuggy (Keuleers & Brysbaert, 2010). To make the number of meaningful and meaningless targets equal, we included 240 pseudowords in the experiment. In half of the pseudoword phrases, we selected the second words as target stimuli with the concern that participants could develop processing strategies as in binominals.

2.3 Procedure

In the present study, we administered the lexical decision task using a masked priming paradigm (Forster & Davis, 1984). We carried out the experiments using E-prime 3.0 that records the reaction time and accuracy of the responses. For each trial, firstly, a fixation marker (*) was presented for 500 ms. and followed by a mask (#####) for 500 m. After the mask, phrases (e.g., dikkat et) were presented as prime for 50 ms, followed by a blank screen for 40 ms. Immediately after that, a constituent of phrases (e.g., DİKKAT) was presented as target items. A target stimulus remained on the screen until the participants responded to it (Figure 1). Participants were required to press "1" if the target word was meaningful in Turkish or if not, press "2". When they did not respond within 3000 ms., the program automatically moved to the next trial. After one trial finished, the subsequent trial was presented in the same order, but a 2000 ms. blank screen was inserted between each trial.

Before the actual trials, to familiarize participants with the experiment procedure, a practice of 30 trials was given.

Figure 1. The Presentation of Stimuli



We divided test items into three counter-balanced lists and administered them to the participants in three different sessions in at least 10-day intervals. Therefore, a participant did not encounter the same target item more than once in an experimental session. We provided these three lists to each participant in a random order, and each session lasted approximately 20 minutes.

2.4 Data Analysis

In this study, we performed six statistical analyzes to determine the effect of formulaicity, semantic transparency, and light verb inclusion and determine whether there was a difference between accuracy rates. We conducted statistics on the mean reaction times given to the target items. Only correct responses to target stimuli were included in reaction time analysis. Repeated measures ANOVA was conducted to compare the effect of formulaicity in speech formula, non-formulaic sequence, and semantically anomalous expression conditions. Significance levels of the F-ratios were adjusted with the Greenhouse–Geisser correction (Greenhouse-Geisser, 1959) for effects with more than 1 degree of freedom in the numerator, and the corrected (Bonferroni, 1936) p values were reported in post hoc comparisons. Paired-samples t-tests were conducted to compare the effect of transparency in transparent and non-transparent conditions and the effect of lightness in light verb and non-light verb inclusion conditions in speech formulas. A non-parametric Friedman test of differences among repeated measures was conducted for the accuracy of formulaicity. Non-parametric Wilcoxon signed-rank tests were conducted for accuracy of transparency and lightness in speech formulas.

3 Results

In the results, we reported that the study's reaction time and accuracy findings regarding the effect size, formulaicity, semantic transparency, and light verb inclusion.

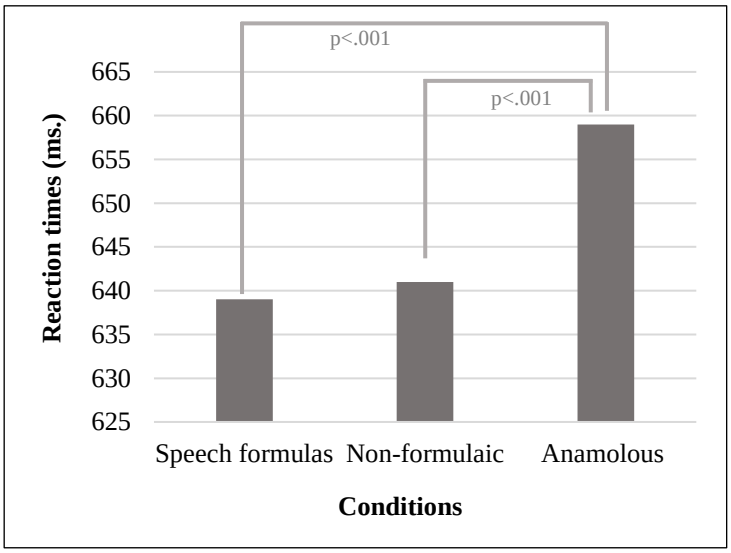
3.1 Effect Size

A statistical power analysis was performed for sample size estimation using G*Power3 (Faul, Erdfelder, Lang, & Buchner, 2007). This study's effect size was 0.229, considered very close to the medium using Cohen's (1988) criteria. With an alpha = .05, power (1-β) = 0.80, the projected sample size needed was 28 (N = 28) for this simplest within-group comparison. Thus, our proposed sample size of 30 was more than adequate for this study's main objective and allowed for expected attrition and our additional objectives of controlling for possible subgroup analysis.

3.2 Reaction Times (RTs)

There was a significant effect of formulaicity on RTs, F (2,58)= 8.542, MSE = 409.093, p <.001, ηp^2 =.228. Three paired samples t-tests were conducted to make post hoc comparisons between conditions. The first paired samples t-test indicated that there was no significant difference in the RTs for speech formulas (M = 640.92, SD = 103.29) and non-formulaic sequences (M = 641.55, SD = 92.05) conditions; t(29) = -0.132, p > .05. The second paired samples t-test indicated that there was a significant difference in the RTs for speech formulas and anomalous expression (M = 659.05, SD = 92.89) conditions; t(29) = -3.525, p < .001. A third paired samples t-test indicated a significant difference in the RTs for non-formulaic sequences and anomalous expression conditions, t(29) = -3.719, p < .001. These results suggested that formulaicity affected processing. Specifically, our results suggested that anomalous expressions' processing took longer than speech formulas and non-formulaic sequences. However, the results indicated no significant difference in the processing of speech formulas and non-formulaic sequences. Namely, there was no real difference in speech formulas and non-formulaic sequences (Figure 2).

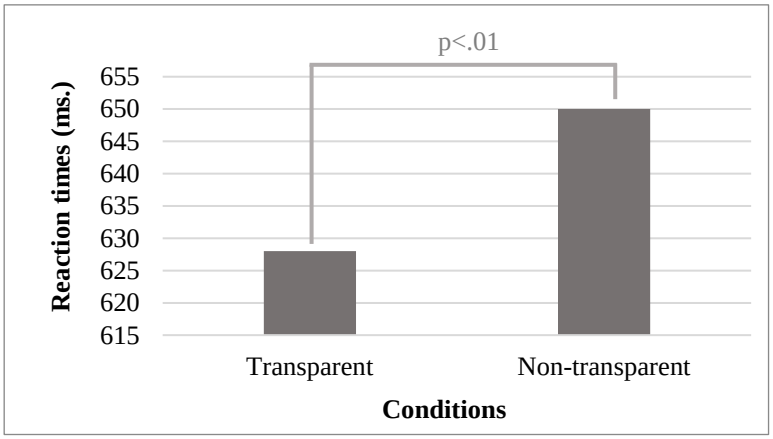
Figure 2. The Mean Reaction Times in Three Conditions for Formulaicity



There was a significant difference in the RTs for transparent (M = 630.94, SD = 97.76) and non-transparent (M = 650.90, SD = 111.05) conditions in speech formulas; t(29) = -3.299, p < .01. It revealed that participants responded to the

target words of transparent speech formulas significantly faster than those of non-transparent ones (Figure 3). This result indicated that participants processed semantically transparent formulas significantly faster than non-transparent formulas.

Figure 3. Effect of Semantic Transparency on Reaction Times



There was no significant difference in the RTs of light verb (M = 634.40, SD = 100.52) and non-light verb (M = 647.44, SD = 109.89) conditions in speech formulas; t (29) = -1.742, p >.05. In other words, the inclusion of light or non-light verbs did not affect the processing of speech formulas.

As a result, the first analysis investigated whether formulaicity had a significant effect on the processing of phrases and revealed such a pattern; anomalous expressions > speech formulas = non-formulaic sequences. Namely, it demonstrated no difference between the mean reaction times given to speech formulas and non-formulaic sequences. Still, anomalous expressions had a significantly longer mean reaction time than the other two conditions. The second analysis revealed that transparent formulas were processed faster than non-transparent sequences, whereas light verbs did not significantly affect the processing of speech formulas.

3.3 Accuracy of Responses

A Friedman's test indicated that there was no significant difference between accuracy rates of formulaicity, $\chi 2F(2) = .350$, $p > .05$. In other words, participants responded with similar accuracy rates to speech formulas (M = 39.70, SD = .595), non-formulaic sequences (M = 39.83, SD = .379), and anomalous expressions (M = 39.76, SD = .505). We compared the accuracy rates of

transparent (M = 19.80, SD = .484) and non-transparent (M = 19.90, SD = .305) formulas. A Wilcoxon signed-rank test revealed that there was no significant difference between transparent and non-transparent formulas, T = 15, z = -1.000, p > .05. Similarly, a Wilcoxon signed-rank test indicated that there was no significant difference between light verbs (M = 19.90, SD = .305) and non-light verbs (M = 19.80, SD = .484), T = 15, z = -1.000, p > .05. As a result, analyses revealed that participants responded with high accuracy rates to the phrases in all experimental conditions.

4 Discussion

In this study, we examined the processing of Turkish noun-verb speech formulas to reveal whether they are processed as a single unit or decomposed into their parts and whether semantic transparency and inclusion of light verbs affect these phrases' processing.

4.1 Holistic Approach vs. Distributes Relationship

Dual-system models for language processing (Wray, 2008; Sinclair, 1991; Erman & Waren, 2000) argue that formulaic sequences are processed with a secondary system different from grammatically generated novel utterances. Namely, these models claim that one of the two systems in language processing is responsible for the decomposition of newly created phrases, whereas the other is responsible for the processing of formulaic sequences without being decomposed into their parts. As these models suggest, the Holistic Approach proposes that formulaic sequences are processed without being decomposed and represented as a whole in the lexicon. Some studies in the literature (Underwood et al., 2004; Tremblay & Baayen, 2010; Isobe, 2011; Sosa & MacFarlane, 2002; Kapatsinski & Radicke, 2009) have supported this argument. However, there are also studies (Arnon & Cohen Priva, 2014; Cangır et al., 2017; Siyanova-Chanturia et al., 2017) supporting the Distributed Representation, which asserts that formulaic sequences are processed through their parts as non-formulaic novel utterances.

Findings obtained in this study revealed that formulaic sequences were processed in the same way as novel utterances. There was no significant difference between the reaction times of speech formulas and non-formulaic sequences. When the participants were presented with speech formulas or non-formulaic sequences as prime stimuli, the mean reaction time given to the following target word was similar. Therefore, this result supported the arguments of the Distributed Representation in the processing of formulaic sequences.

As the Holistic Approach suggests, if formulaic sequences were processed without being decomposed into their parts and represented as a whole in the lexicon, the mean reaction time given to the target words of speech formulas

would be longer than that of the non-formulaic sequences. The Holistic Approach argues that formulaic sequences are processed and stored as single units without access to their components. Thus, accessing part of a speech formula and responding to it would be expected to take longer. However, the findings of this study suggested otherwise.

The present study found no significant difference between the mean reaction times of speech formulas and non-formulaic sequences, which contradicts Sosa and MacFarlane (2002), and Kapatsinski and Radicke (2009) findings. These two studies stated that participants responded to "of" and "up," which were a part of formulaic sequences slower because formulaic sequences with high frequency were processed as a whole without being decomposed, and it required a secondary analysis to access their parts. Sosa and MacFarlane (2002) reported that the mean reaction time given to "of," which was a part of the formulaic sequences with high frequency, was longer than that of the non-formulaic sequences. This finding allegedly supported the arguments of the Holistic Approach. However, this study was criticized in that the word "of" was not suitable for the study and may have affected the findings because "of" undergoes a substantial articulatory reduction in spoken language, which may have made it difficult to detect "of" and take longer to respond to it.

Kapatsinski and Radicke (2009) carried out a similar study with "up" that shows less articulatory reduction. They found that participants responded to "up" in low-frequency and high-frequency phrases faster than in medium-frequency phrases. They interpreted this finding as phrases with high frequency were processed as a whole and supported the study of Sosa and MacFarlane (2002). However, participants responded to "up" in low-frequency phrases faster as they did in high-frequency phrases. This finding contradicts the argument of holistic processing.

On the other hand, some studies (Underwood et al., 2004; Isobe, 2011; Tremblay & Baayen, 2010), which claim to support the Holistic Approach in the processing of formulaic sequences, focused on the processing speed rather than the relation of formulaic sequences to their parts. In an eye-tracking study, Underwood et al. (2004) found that participants read formulaic sequences in short stories faster and made less frequent and shorter fixations on the last words of the sequences. Isobe (2011) found that participants rated the grammaticality of formulaic sequences more quickly. Tremblay and Baayen (2010) verified that participants remembered formulaic sequences more often than non-formulaic ones because they were more predictable. All these studies claimed to support the arguments of the Holistic Approach in the processing of formulaic language. However, reaching such a conclusion is impossible based only on processing speed without examining the relationship between formulaic sequences and their parts. Many studies have already demonstrated that formulaic sequences are processed significantly faster than novel utterances. To make inferences about

this advantage's source, researchers need to investigate how these sequences relate to their parts.

The present study aimed to examine the relation of formulaic sequences to their parts. That is why speech formulas and non-formulaic sequences were presented as prime stimuli for a very short time, and then participants were asked to react to the first words of these phrases as target items. When the reaction times were analyzed, there was no significant difference between the reaction times given to the target items when speech formulas or non-formulaic sequences were presented as prime. Therefore, this study found that participants processed speech formulas through their parts by decomposing them into constituents, as in non-formulaic sequences. If speech formulas were processed as a single unit without being decomposed into components, it would take longer to access and react to the target items that were a part of the formulas.

This finding supports the findings of Arnon and Cohen Priva (2014). They found that the frequency of words affected the phonetic duration's length in frequent formulaic sequences; thus, participants accessed components while processing formulaic sequences. Similarly, in the study where the processing of English binominals was examined via ERP, it was reported that frequent multi-word phrases were activated in the mind of participants through their parts (Siyanova-Chanturia et al., 2017). Lastly, Cangır et al. (2017) investigated the priming effect in Turkish collocations. They found that when a word was encountered, its collocations were also mentally activated. These findings also support that formulaic sequences are processed through their parts. Therefore, these studies support the arguments of Distributed representation in the processing of formulaic sequences.

To sum up, Distributed Representation posits that formulaic sequences are not processed as a single unit but instead processed through their parts. It argues that words frequently occurring together develop a mental bond over time and mentally activate one another faster (Haselow, 2018; Siyanova-Chanturia, 2015). This study's findings support this view, unlike the Holistic Approach. It revealed that participants accessed parts of Turkish speech formulas and responded to them as fast as in non-formulaic sequences. With this study's findings, it is impossible to discuss the Continuum Processing Approach because we did not design the experiment set to test it. There are also no other studies investigating the Continuum Processing Approach's arguments in processing formulaic sequences. However, studies on this approach are needed to reveal the source of formulaic sequences' processing advantage over novel utterances.

4.2 Semantic Transparency

Semantic transparency is a significant factor in the processing of formulaic sequences. It is argued that non-transparent sequences are processed as single units and faster than transparent ones (Wood, 2015). However, studies

investigating this argument reported conflicting results. Conklin and Schmitt (2008) found that native and non-native participants processed the literal meaning of English idioms as fast as the figurative meaning in short stories in a self-paced reading task. Namely, they stated that semantic transparency did not affect the processing of idioms. They argued that presenting idioms in the context of short stories made the literal meaning to be processed fast. Siyanova-Chanturia et al. (2011) carried out a similar study with an eye-tracking paradigm and found no significant difference in native speakers' processing of figurative and literal meaning. However, they reported that non-native speakers processed the figurative meaning of idioms slower than the literal meaning. They argued that non-native speakers processed figurative idioms slower because the link between figurative idioms' form and meaning was not strong for them.

This study found that participants responded to the target items faster, which were a part of transparent formulas. Non-transparent speech formulas may have processed slower because it is more challenging to access figurative meaning, as Siyanova-Chanturia et al. (2011) suggest. The eye-tracking study (Arıca-Akkök & Uzun, 2018) on the processing of Turkish metaphorical expressions supports this interpretation. They reported that participants made regressions when they met with figurative phrases. Thus they argued that non-transparent sequences were analyzed, and their literal meanings were activated first to access figurative meanings. In other words, the processing of non-transparent expressions was more difficult and slower. Therefore, in this study, it may have taken longer for participants to process non-transparent speech formulas and respond to these formulas' target words.

4.3 Inclusion of Light Verbs

Light verbs do not contribute much meaning to phrases and cannot assign syntactic roles to arguments in a sentence, and they are mostly used in combination with nouns (Akşehirli, 2013). That's why this study investigated whether the inclusion of light verbs affects Turkish speech formulas' processing. Reaction time analysis indicated that including light verbs did not lead to a significant difference in the processing of noun-verb speech formulas in Turkish. However, the absence of studies questioning how light verbs affect formulaic sequences processing prevents drawing a generalizable inference. Therefore, further studies in different languages are needed on this topic.

5 Conclusions

This study examined the processing of Turkish noun-verb speech formulas by conducting a lexical decision experiment using a masked priming paradigm. It aimed to investigate whether Turkish noun-verb speech formulas are processed as a single unit as the Holistic Approach suggests or decomposed and processed

through their parts as Distributed Representation argues. This study's findings indicated that there was no significant difference in the processing of speech formulas and non-formulaic sequences; thus, this supported the Distributed Representation's arguments. There was a significant difference when we examined the effect of semantic transparency in processing speech formulas. Participants responded to the target words, which were a part of transparent formulas faster. The slower processing of non-transparent speech formulas may have led to this difference. We also found that the inclusion of light verbs did not lead to a significant difference in the processing of speech formulas. Since there are no other studies in the literature investigating light verbs' effect in processing formulaic sequences, it is difficult to interpret this finding.

References

Aksan, Y., Aksan, M., Koltuksuz, A., Sezer, T., Mersinli, Ü., Demirhan, U. U., Yılmazer, H., Kurtoğlu, Ö., Atasoy, G., Öz, S., & Yıldız, İ. (2012). Construction of the Turkish National Corpus (TNC). In N. Calzolari, K. Choukri, T. Declerck et al. (Eds.), *Proceedings of the 12th International Conference on Language Resources and Evaluation* (pp. 3223-3227). İstanbul: LREC 2012.

Akşehirli, S. (2013). Türkçede et- katkısız eyleminin sözlüksel işlevleri. *Turkish Studies - International Periodicals for The Languages, Literature and History of Turkish or Turkic, 8*(9), 481-494.

Arıca-Akkök, E., & Uzun, İ. P. (2018). Metaphor processing in Turkish: An eye-movement study. *Mersin Üniversitesi Dil ve Edebiyat Dergisi, 15*(1), 105-124.

Arnon, I., & Cohen Priva, U. (2014). Time and again: The changing effect of word and multiword frequency on phonetic duration for highly frequent sequences. *The Mental Lexicon, 9*(3), 377–400.

Berk, G., Erden, B., & Güngör, T. (2018). Turkish verbal multiword expressions corpus, *26th Signal Processing and Communications Aplications Conference, 64*, 1-4.

Bonferroni, C. E. (1936). Teoria statistica delle classi e calcolo delle probabilita. *Pubblicazioni del R Istituto Superiore di Scienze Economiche e Commerciali di Firenze, 8*, 3–62.

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

Carrol, G., & Conklin, K. (2019). Is all formulaic language created equal? Unpacking the processing advantage for different types of formulaic sequences. *Language and Speech, 63*(1), 95–122.

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.

Conklin, K., & Schmitt, N. (2008). Formulaic sequences: Are they processed more quickly than nonformulaic language by native and nonnative speakers? *Applied Linguistics, 29*(1), 72–89.

Erman, B., & Warren, B. (2000). The idiom principle and the open choice principle. *Text, 20*(1), 29-62.

Faul, F., Erdfelder, E., Lang, AG. et al. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods, 39*, 175–191.

Forster, K. I., & Davis, C. (1984). Repetition priming and frequency attenuation in lexical access. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 10*, 680–698.

Greenhouse, S. W., & Geisser, S. (1959). On methods in the analysis of profile data. *Psychometrika, 24*, 95–112.

Haselow, A. (2018). Language change from a psycholinguistic perspective: The long-term effects of frequency on language processing. *Language Sciences, 68*, 56–77.

Hunston, S. (2002). *Corpora in applied linguistics*. Cambridge: Cambridge University Press.

Isobe, Y. (2011). Representation and processing of formulaic sequences in L2 mental lexicon: How do Japanese EFL learners process multi-word Expressions? *JACET Kansai Journal, 13*, 38–49.

Jeong, H., & Jiang, N. (2019). Representation and processing of lexical bundles: Evidence from word monitoring. *System,* 188-198.

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

Kapatsinski, V., & J. Radicke. (2009). Frequency and the emergence of prefabs: Evidence from monitoring. In R. Corrigan, E. Moravcsik, H. Ouali, & K. Wheatley (Eds). *Formulaic Language. Vol. II: Acquisition, loss, psychological reality, functional explanations* (pp. 499-520). Amsterdam: John Benjamins.

Keuleers, E., & Brysbaert, M. (2010). Wuggy: A multilingual pseudoword generator. *Behavior Research Methods, 42*(3), 627-633.

Lamb, S. (1999). *Pathways of the brain: The neurocognitive basis of language*. Amsterdam: Benjamins.

Ortaçtepe, D. (2013). Formulaic language and conceptual socialization: The route to becoming nativelike in L2. *System, 41*, 852–865.

Pérez-Llantada, C. (2014). Formulaic language in L1 and L2 expert academic writing: Convergent and divergent usage. *Journal of English for Academic Purposes, 14*, 84-94.

Qin, J. (2014). Use of formulaic bundles by non-native English graduate writers and published authors in applied linguistics. *System, 42*, 220-231.

Sinclair, J. (1991). *Corpus, concordance, collocation*. Oxford: Oxford University Press.

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

Siyanova-Chanturia, A., Conklin, K., & Schmitt, N. (2011). Adding more fuel to the fire: An eye-tracking study of idiom processing by native and non-native speakers. *Second Language Research, 27*(2), 251–272.

Siyanova-Chanturia, A., Conklin, K., Caffarra, S., Kaan, E., & Heuven, W. J. B. (2017). Representation and processing of multi-word expressions in the brain. *Brain and Language 175,* 111–122.

Siyanova-Chanturia, A., & Martinez, R. (2014), ' The Idiom Principle Revisited'. *Applied Linguistics, 36*(5): 549–69.

Siyanova-Chanturia, A. (2015). On the'holistic'nature of formulaic language. *Corpus Linguistics and Ling. Theory, 11*(2), 285–301.

Sosa, A. V., & J. MacFarlane. (2002). Evidence for frequency-based constituents in the mental lexicon: Collocations involving the word of. *Brain and Language 83*(2), 227–236.

Tremblay, A., & Baayen, R.H. (2010). Holistic processing of regular fourword sequences: A behavioral and ERP study of the effects of structure, frequency, and probability on immediate free recall. In D. Wood (Ed.), *Perspectives on formulaic language: Acquisition and communication* (pp. 151–173). London and New York: Continuum

Underwood, G., Schmitt, N., & Galpin, A. (2004). The eyes have it: An eye-movement study into the processing of formulaic sequences. In N. Schmitt (Ed.), *Formulaic sequences* (pp. 153-172). Philadelphia: John Benjamins.

Uygun, S., & Gürel, A. (2016). Processing morphology in L2 Turkish: The effects of morphological richness in the L1. In A. Gürel (Ed.), *Second Language Acquisition of Turkish* (pp. 251-279). Amsterdam: John Benjamins Publishing.

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

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

Wood, D. (2015). *Fundamentals of formulaic language: An introduction*. London: Bloomsbury.