THE EFFECT OF FREQUENCY ON THE COMPREHENSION OF MULTI-WORD PHRASES IN TURKISH

HÜSEYİN DEMİRTAŞ

BOĞAZİÇİ UNIVERSITY

2018

THE EFFECT OF FREQUENCY ON THE COMPREHENSION OF MULTI-WORD PHRASES IN TURKISH

Thesis submitted to the

Institute for Graduate Studies in Social Sciences

in partial fulfillment of the requirements for the degree of

Master of Arts

in

Cognitive Science

by

Hüseyin Demirtaş

Boğaziçi University

2018

The Effect of Frequency on the

Comprehension of Multi-Word Phrases in Turkish

The thesis of Hüseyin Demirtaş

has been approved by:

Assoc. Prof. Mine Nakipoğlu
(Thesis Advisor) _____

Assoc. Prof. Albert Ali Salah
(Thesis Co-Advisor) _____

Assoc. Prof. Arzucan Özgür _____

Assist. Prof. Pavel Logačev _____

Assoc. Prof. Mine Güven
(External Member) _____

March 2018

DECLARATION OF ORIGINALITY

I, Hüseyin Demirtaş, certify that

- I am the sole author of this thesis and that I have fully acknowledged and documented in my thesis all sources of ideas and words, including digital resources, which have been produced or published by another person or institution;
- this thesis contains no material that has been submitted or accepted for a degree or diploma in any other educational institution;
- this is a true copy of the thesis approved by my advisor and thesis committee at Boğaziçi University, including final revisions required by them.

Signature.....

Date .....04/05/2018.....

ABSTRACT

The Effect of Frequency on the Comprehension of Multi-Word Phrases in Turkish

The effect of frequency on multi-word phrases is a comparatively recent area of inquiry. The interaction between multi-word phrases and frequency effects and the existence of multi-word frequency effects can shed light on the debates about the nature of human mental lexicon. Words and Rules approach to mental lexicon argues that words and rules are stored separately in the lexicon while Emergentist models argue that frequency of exposure shapes the representation of language and there is not a very clear-cut distinction between words and rules. In this study, we explored the effects of multiword frequency by testing reaction times to a number of different types of two-and-three word phrases. In Experiment 1, we analyzed the RT (reaction time) differences between 15 pairs of adjective + noun combinations sharing the same adjective. In Experiment 2, we analyzed the RT values for the nouns in Experiment 1 to see the relationship between collocation frequency and word frequency and to make sure that the RT differences in Experiment 1 was not due to single-word frequency. Finally, in Experiment 3, we tested three different types of collocations. The results of these experiments showed that multi-word frequency exists independent of single word frequency. The results of this study provide evidence that multiword frequency exists, and thus supports the previous studies in the field with data from Turkish. Moreover, the data support Emergentist models and show that multiword frequency may provide a processing advantage in language comprehension.

ÖZET

Sıklığın Türkçedeki Çoklu Sözcük Gruplarının Anlaşılması Üzerindeki Etkisi

Sıklığın çok sözcüklü ifadeler üzerindeki etkisi nispeten güncel bir araştırma alanıdır. Sıklık ve çoklu sözcük grupları arasındaki ilişki ve çoklu sözcük grupları sıklık etkilerinin varlığı, insan sözcük dağarcığının doğası üzerine devam eden tartışmalara ışık tutabilir. Kelimeler ve Kurallar yaklaşımı, insan zihninde kelime ve kuralların ayrı depolandığını savunurken Belirimci modeller maruz kalma sıklığının dilin temsilini şekillendirdiğini ve kelime ve kurallar arasında net bir ayrım yapılamayacağını öne sürer. Bu çalışmada, çoklu sözcük grubu sıklık etkileri, bir dizi 2 ve 3 sözcüklü ifadenin tepki süreleri ölçülerek incelenmiştir. Deney 1'de, aynı sıfatı paylaşan 15 çift Sıfat + Ad çiftini analiz ettik. Hem Deney 1'de elde edilen sonuçların bu isimlerin sıklığı ile ilgili olmadığını göstermek hem de çoklu kelime grubu sıklığı ve kelime sıklığı arasındaki ilişkiyi incelemek için Deney 2'de, Deney 1'deki isimlerin tepki sürelerini test ettik. Son olarak Deney 3'te üç farklı ifade türünü test ettik. Bu testlerin sonuçları çoklu kelime grubu etkilerinin, kelime grubu etkilerinden bağımsız olarak var olduğunu gösterdi. Bu araştırmanın sonuçları çoklu kelime grubu sıklığı etkilerinin varlığını gösteren kanıtlar sunuyor. Dahası elde edilen veri Belirimci modelleri destekliyor ve çoklu kelime grubu sıklığı etkisinin dil algısında bir işlemlenme avantajı sağladığını gösteriyor.

ACKNOWLEDGEMENTS

I am deeply grateful to a number of people for their support and encouragement during the planning, execution and analysis of this study.

First, I would like to offer my heartfelt gratitude to Assoc. Prof. Mine Nakipoğlu who has been not only and advisor but also a role model and inspiration to me throughout this whole research process. Without her guidance, support, and her contagious love and passion for language, I would not be able to complete my study.

I would also like to thank Assoc. Prof. Albert Ali Salah for introducing me to the field of cognitive science and allowing me to pursue many great courses and projects when I served as a research assistant at the Cognitive Science MA program at Boğaziçi. I am grateful for his support, humor and energy.

Additionally, I would like to thank the members of my jury Assoc. Prof. Arzucan Özgür, Assoc Prof. Mine Güven and Assist. Prof. Pavel Logačev as well as all of my MA program professors who inspired me with their discussions.

I should definitely take a moment to thank my father who has not missed any single phone call to nag me about the progress of my thesis and my friend Talip Gülle with whom I have had many insightful discussions on the nature of language and my study.

Finally, I would like to thank my wife Elif for being the loving, caring and understanding spouse she is. Without her constant support, encouragement and help I would be far from making progress. I would also like to thank my baby girl Fatma Sare who is about to turn 4 months old for encouraging me to work harder with both her cries and smiles.

TABLE OF CONTENTS

CHAPTER 1: INTRODUCTION 1

CHAPTER 2: LITERATURE REVIEW 8

2.1 The role of frequency in the processing and production of language 8

2.2 Formulaic language 9

2.3 Emergentist Models vs. Word and Rules Model 11

2.4 Experimental data on multi-word frequency effects 13

CHAPTER 3: EXPERIMENT 1 & EXPERIMENT 2 19

3.1 Experiment 1 20

3.2 Experiment 2 29

CHAPTER 4: EXPERIMENT 3 37

4.1 Method 37

4.2 Results 44

4.3 Discussion 52

CHAPTER 5: GENERAL DISCUSSION 61

CHAPTER 6: FUTURE DIRECTIONS 72

APPENDIX A: ITEMS IN EXPERIMENT 1 77

APPENDIX B: ITEMS IN EXPERIMENT 2 79

APPENDIX C: ITEMS IN EXPERIMENT 3 81

REFERENCES 84

LIST OF TABLES

Table 1. Items in Experiment 1 22

Table 2. Mean RTs for Adjective + Noun Collocations 26

Table 3. Nouns in Experiment 2 32

Table 4. Mean RTs for Adjective + Noun Collocations vs Nouns 33

Table 5. Selected Items from Experiment 2 34

Table 6. Items in Experiment 3.1 41

Table 7. Items in Experiment 3.2 42

Table 8. Items in Experiment 3.3 43

Table 9. Comparison of Two and Three-word Collocations in Experiment 3.1 46

Table 10. Mean RTs and SDs for Two-word and Three-word Collocations 47

Table 11. Collocations with "ve" 49

Table 12. Mean RTs and SDs for Collocations with 've' and Their Reversed Forms 49

Table 13. RT for Two-word Collocations and Initial Nouns in Experiment 3.3 51

Table 14. Mean RTs and SDs for Two-word Collocations and Their Initial Words 51

Table 15. Comparison of Two and Three-word Phrases in Experiment 3.1 54

Table 16. Three-word Collocations and Word Frequencies 56

Table 17. Three-word Collocations and Word Multiplication 57

LIST OF FIGURES

Figure 1. Collocation frequency and RT 26

Figure 2. Mean RTs for adjective + noun collocations vs nouns 33

Figure 3. Comparison of two-word and three-word collocations 47

Figure 4. Three-word collocations and their reversed forms 49

Figure 5. Two-word collocations vs their initial words 52

CHAPTER 1

INTRODUCTION

One of the hallmarks of human mind is the ability to produce and comprehend language. Language is one of the few truly unique traits that make human beings special. Language is what enables human knowledge and culture to be transferred from generation to generation and thus leading to an accumulation of knowledge in various fields of human endeavor.

In its most rudimentary form, language can be considered as a system that consists of a set of rules and words i.e. grammar and vocabulary respectively. Grammar or the rules mainly determine the way words are combined whereas words or the vocabulary carries the core of the meaning that a language user expresses. In this sense, words play a crucial part in the storage and transfer of meaning and the role of words has been the object of study in many fields such as linguistics, psychology, computer science, philosophy and recent fields such as cognitive science. Each discipline analyzes the words within the dominant frameworks and paradigm of their respective fields. However, more holistic and comprehensive insights are more likely to be obtained if the approaches of these various fields are blended. Therefore, this study will analyze the frequency effects on the comprehension of multiword phrases, an important issue about word storage and comprehension, by referring to the literature and methods from various scientific disciplines.

The main goal of this study is to shed light on the effects of frequency in the processing of multiword phrases in Turkish. Frequency effects have been shown to affect language processing and comprehension in various studies. Most of these

studies have focused on single word tasks, that is, in many studies researchers investigated the effect of frequency between single word pairs or items (Rayner & Duffy, 1986; Monsell, 1991; Jurafsky, Bell, Gregory, and Raymond, 2001). The studies on single words have contributed tremendously to the understanding of the effects of frequency on language but they have largely been restricted to single word domain. Given the fact that language users produce utterances of varying lengths, some ranging from one word to ten to fifteen words, it seems logical to look at the area beyond single word units. Language is a complex system with dynamic and ever-changing parts thus to obtain a full picture of the underlying principles of language, it is a necessity to enlarge and broaden our range of experimental items.

In this study, we will analyze the effects of frequency on multiword phrases in Turkish to uncover whether such phrases are stored in the minds of Turkish speakers holistically. Bybee (2006) states that although idioms have received a lot of attention in the linguistic literature, this is not really the case for collocations or *prefabs* which she describes as conventionalized and predictable word sequences such as *prominent role*. Similarly, Erman and Warren (2000) state that prefabricated word sequences account for 55% of written and oral language. Given their ubiquity and importance in both the comprehension and production of language (Malkiel, 1959), we hope that our study will make useful contributions to the study of collocations.

Previous studies such as Alegre and Gordon (1999), and Sosa and MacFarlane (2002) have shown that the frequency with which language users encounter a word has a significant impact on their processing of that word as exemplified by reaction time studies. Higher frequency generally leads to a faster response in reaction time studies. If subjects are presented with two words that have

the same length and structure, the word that has a higher frequency will generally have a faster reaction time whereas the word that has a lower frequency will exhibit more response latency.

This finding has serious implications for the models of mental lexicon and language processing (Aitchison, 2012). The fact that language users are sensitive to frequency information suggests that frequency information about words is tracked and must be stored in the human brain. What is more, since language users' exposure to a certain word varies greatly over time, the storage of frequency information on words seems to be an ongoing and dynamic process in the human brain. This is a strong illustration of the effect of experience on the representation of language in the human brain. It seems that human brain takes statistics on the use of language without humans consciously noticing it. A person who has an affection for flowers and who spends most of their leisure time caring for different species of flowers will respond much more quickly to the names of flowers than a person who has little to no interest in and thus exposure to flowers. As this example demonstrates, the representation of words and ultimately language in our mind is not static but rather dynamic and ever changing.

Considering the fact that exposure to a single word increases or strengthens its representation in the human brain it follows that a similar mechanism should apply to multiword expressions if they frequently occur as a whole. Such an idea is both plausible and striking at the same time because this means human brain takes statistics at various levels, not just on a single word level but at levels far greater than a single word. This might seem like a lot of overhead and use of mental resources yet on the other hand, it might just be using a fraction of the immense capacity of the human brain to reduce demand in online processing.

The issue of mental statistics as the result of exposure is one of the pillars of many Emergentist models of language. Emergentist models of language such as exemplar-based models or usage-based models are based on the fact that language use shapes the representation of language. For instance, in an exemplar based model, the representation of the word "dog" is shaped and constantly updated by the language user's interaction with and exposure to that word (Bybee, 1998; Goldberg, 1995, 2006; Christiansen & Chater, 1999; Elman 1990). Emergentist models go beyond single word level in their evaluation of language. According to Emergentist models, language users are sensitive to frequency effects beyond single word level. Yet this puts Emergentist models such as those of Goldberg (2006) and Tomasello (2003) in conflict with the Words and Rules model since WR model posits that rules operate on word level.

It is necessary to analyze the effect of phrase frequency to shed light on the debate between Emergentists models and the Words and Rules approach. This is the first study, to our knowledge, to test multiword or phrase frequency effects on comprehension in Turkish. The results of this study will inform models of language processing concerning the storage of words. Our hypothesis is that phrase frequency effects are real and they lead to holistic storage of phrases in the mental lexicon and thus provide advantage in language processing by facilitating word recall and retrieval.

In this study, we test the existence and effect of multiword frequency on the processing of multiword phrases in Turkish. As will be laid out in the next chapter, the results of studies on multiword frequency have been somewhat mixed as the results of some experiments cannot solely be attributed to multi-word frequency due to their experimental setups (Conklin and Schmitt, 2012). We conjecture that it

would be beneficial to test multiword frequency effects with a wide range of multiword phrases, thus, to explore the issue; in this study, we have conducted three experiments. In Experiment 1, we compare adjective + noun type of collocations such as *ağır ceza* 'heavy penalty' and *ağır yara* 'deep wound' in a phrasal decision task. In this pair, for instance *ağır ceza* has high collocation frequency whereas *ağır yara* has a comparatively low collocation frequency. Hence through a comparison of high frequency adjective + noun pairs to low frequency adjective + noun pairs, Experiment 1 attempts to find out whether phrase level frequency effects exist. Furthermore, we will compare the reaction times to phrases with their corpus-identified phrase frequency (i.e. the frequency of *ağır ceza*) and the frequency of their constituent words. We predict that there will be significant reaction time differences in Experiment 1, between high frequency adjective + noun pairs and low frequency adjective + noun pairs, as we believe that human brain stores and keeps track of statistical information at multiword level.

To ensure that the findings from Experiment 1 are not simply due to single word frequency, in Experiment 2, we conduct a lexical decision task for the nouns in Experiment 1. Experiment 2 will enable us to analyze the relationship and interaction between single word frequency and multiword frequency.

In Experiment 3, we test different types of collocations to be able to evaluate phrase frequency effects in a wide variety of collocations. In Experiment 3.1, we compare some of the two-word collocations from Experiment 1 to their extended three-word forms such as comparing *hızlı karar* 'quick decision' to *hızlı karar vermek* 'to make a quick decision'. In Experiment 3.2, we analyze collocations that are formed with *ve* 'and' such as *gelin ve damat 'bride and groom'*, against their reversed forms such as *damat ve gelin 'groom and bride'*. Both the typical forms and

the reversed forms of these collocations are perfectly acceptable in Turkish, thus, any difference between reaction times of the typical and reversed forms can be attributed to phrase frequency effects as the frequencies of the words that make up these phrases are the same. We predict that there will be a significant difference between reaction times to the typical form of a collocation and its reversed form.

In Experiment 3.3, we compare the reaction time differences in a lexical decision task between a number of two-word collocations and their initial words such as comparing *hesap kitap* 'calculation book' to *hesap* 'calculation'. Under normal circumstances, one would expect that the reaction time values for the initial word of a collocation would be faster than the reaction time values for the whole collocations. However, if we find that the reaction times of the two-word collocations groups is lower or the same with the initial word group, then this will provide strong evidence for phrase frequency effects. We predict that there will not be a significant difference between the reaction times to the initial word of a two-word collocation and the reaction times to the two-word collocation.

We have selected a wide range of phrase types to obtain a broad picture on the interaction between phrase frequency and comprehension. By collecting data on adjective + noun collocations, single nouns, three word collocations with 've', as well as two-word idiomatic and non-idiomatic collocations and the extended phrases derived from adjective noun pairs, we cover a huge selection of phrase types and thus we expect our results to be widely applicable for many multi-word phrases. This study is also unique in that to our knowledge no other study analyzed multi-word frequency effects utilizing adjective + noun pairs such as *ağır ceza* "heavy penalty" and similarly no other study has analyzed their extended forms such as *ağır ceza hakimi* "high criminal court judge". We speculate that these additional types of

stimuli will provide further insights into the nature of phrase frequency effects. Another unique aspect of our study is that we not only compare the frequencies of single words and phrases but also the reaction times of both individual words and phrases. This helps us attribute processing advantages to phrase frequency more confidently as we eliminate single word reaction time (henceforth RT) as a confounding factor.

The thesis is structured as follows. Chapter 2 provides a survey of existing literature on frequency effects and models of lexical processing and representation. Chapter 3 provides a detailed look at Experiment 1 and Experiment 2, which explore frequency effects of Adjective + Nouns combinations and single nouns respectively. Chapter 4 introduces Experiment 3, which compares phrase frequency effects on 3 types of collocations i.e. extended forms of Adjective + Noun pairs such as *hızlı karar vermek* 'make a quick decision', collocations with *ve* 'and' such as *gelin ve damat* 'bride and groom' and common two-word collocations such as *çarşı pazar* "bazaar market". Chapter 5 provides a general discussion of all of the data and results from the experiments and analyzes their implications for a model of lexical processing. Chapter 6 concludes the study with future research directions that can complement further studies in the area of phrase frequency effects.

CHAPTER 2

LITERATURE REVIEW

There has been a lot of work on the effects of frequency on lexical processing especially at the single word level. Yet, the emphasis on multi-word phrases or structures remained quite limited until recently. The last decade has witnessed an increased interest on the effects of frequency on multi-word phrases. Arnon and Snider (2010) argue that language users are not only sensitive to single word frequencies but also to frequencies at different levels including two-word and three-word frequency. They also claim that frequency information is stored for multiword phrases in addition to single words. This is because the knowledge humans gather about language demonstrates itself at various levels, not just at a single word level, as the users of a language have intuitions about the use of words not only in isolation but also in groups. Therefore, linguistics models have to account for such types of information in order to capture the realities of the linguistic system.

2.1 The role of frequency in the processing and production of language

The role of frequency in the processing and production of language has been a prevalent topic. Several studies demonstrate the role of frequency on word recognition (Monsell, Doyle & Haggard, 1989; Monsell, 1991), sentence comprehension (Rayner & Duffy, 1986) and syntactic structures (Frazier & Fodor, 1978). Gürel (1999) shows that some frequent multimorphemic forms in Turkish probably induce a whole-form lexical access instead of a decomposition route because they demonstrate processing advantages and lower response latencies. This indicates that they are stored as a whole rather than being combined online. In a

comprehensive study on the nature of frequency effects on the processing of morphologically complex Turkish words, Bilgin (2016) argues that there might be mental representations of frequently recurring suffix sequences as well as individual suffixes. This idea is further supported by Durrant (2013) which examined the formulaic nature of complex inflectional patterns in Turkish. Durant (2013) argues that frequently occurring complex inflectional patters demonstrate qualities that are often associated with lexical bundles or formulaic language such as frequent co-occurrence of certain parts, fixed order and interaction between lexicon and grammar.

The role of frequency is not only limited to recognition; it can also be observed in production. For example, Jurafsky, Bell, Gregory, and Raymond (2001) show that the likelihood of the presence of a word in context is affected by the preceding word such as the word “sort” being succeeded by the word “of”. Diessel (2007) and Ellis (2002) provide detailed information on language acquisition, use, and diachronic change. These studies provide evidence that language users are sensitive to distributional information not only at word level but also at various syntactic and phrasal levels such as single word level and multi-word level. See Brysbaert, Mandera & Keuleers (2017) for an in-depth exploration of word frequency effects.

2.2 Formulaic language

Formulaic language is a type of language that makes extensive use of formulaic sequences. Biber (1999) recommended that for a two to four word sequence such as *in the case of the* or *do you want me to* to be considered formulaic, it should appear at least 10 times per million in a corpus. This number can act as a threshold value in

determining whether a phrase is formulaic. Various studies show that formulaic language is a common occurrence in language use. Sorhus (1977) states that one fifth of the items in her corpus of spontaneous Canadian English speech consisted of formulaic language such as *you know* or *that kind of thing*. There is a similar picture in written language as well. Howarth (1998) found that in around 40% of the cases formulaic expressions were present in an academic corpus.

How we store words in our mental lexicon is an open question. Whether words are stored in their base forms such as "car" instead of "cars" might depend on various factors such as the frequency of the base and the inflected forms. The role of frequency on storage is not entirely clear, either. How does a man who has several cars and who constantly talks about their cars store the word "car"? Does he store the base form 'car' or the plural form "cars" separately? Whether this will facilitate processing and provide an advantage for linguistic processing might inform models of language processing.

Formulaic language provides language users with a clear advantage. As Conklin and Schmitt (2012) notes, formulaic language helps language users to become more fluent as formulaic expressions decrease the cognitive load during language production. Conklin and Schmitt (2012) further argue that it makes sense for the brain to make use of the redundant long-term memory capacity in order to free up the limited working memory capacity. This might be the reason why multi-word frequency effects exist. The brain might simply be reducing its computational load by storing frequently recurring phrases. If this is the case, then frequency effects can be observed in phrases of various length as long as a phrase is recurrent enough.

2.3 Emergentist Models vs. Word and Rules Model

There is a debate on the nature of the lexicon in the field of linguistics. The main debate revolves around the unit of storage. Many early models propose a single word based storage and representation. Yet recent studies show that lexical storage is not exclusive to single word units. Studies in areas such as sentence processing, speech production and compound word processing using methods such as eye-tracking, lexical decision tasks provide evidence that multi-word sequences are stored in the lexicon. In what follows, I will lay out some of the prominent studies on the issue.

Frequency effects influence the theories and models about the representation and learning of language. Recent models place emphasis on experience in the creation of linguistic units. These models are generally termed as Emergentist models. Some prominent examples of these types of models where personal experience plays a crucial role in the creation, retention and processing of linguistic information are Usage-based Grammar (Bybee, 1998; Goldberg, 1995, 2006) and connectionist models of learning and processing by Christiansen & Chater (1999), Elman (1990), Christiansen (1994), McClelland and Kawamoto (1986) and Noelle and Cottrell (1995).

Usage based models (Bybee, 1998; Goldberg, 2006; Tomasello, 2003) and exemplar-based models (Abbot-Smith & Tomasello, 2006; Bod, 2006) state that the representation of language in the human mind is shaped by experience. Therefore, each encounter with a formulaic construction will shape its representation and storage. Such a view predicts a processing advantage for frequent structures over less frequent ones. More interestingly, such a model predicts that the representation of word and the formation of formulaic constructions is a dynamic and ever-changing process.

Similarly, as Conklin and Schmitt (2012) state, in a connectionist approach, units form relationships and do not exist solely in isolation. The frequency of occurrence of various patterns determines the activation and thus the strength of the connections in the mental lexicon. If two units recur frequently then they might form a relationship and activate each other. For instance, words such as “red” and “car” do not exist in isolation in the English language. They form phrases or sentences by combining with other words or structures. If they frequently occur together then this might strengthen the relation between these two words in the human mind and they might activate each other which means if one of them is uttered or seen it might prime the other one.

Emergentist approaches such as usage-based models and connectionist models are radically different from the Words and Rules approach (henceforth WR Model) to language as presented by Pinker (1991, 1999), and Pinker and Ullman (2002). In the WR Model, there is a clear separation between a mental lexicon and a mental grammar. In this system, lexicon consists of memorized word forms and the grammar includes rules and constraints employed to combine these words. These two components are thought to have different cognitive underpinnings.

In models favoring the WR Approach, frequency effects are expected to occur only in stored forms as the computed forms are claimed to be managed by different processes than stored forms. According to the Words and Rules Approach, multi-word phrases are not stored and thus computed and as a result, they are not expected to exhibit frequency effects (Arnon & Snider, 2010). Similarly, according to the Word and Rules Approach (Pinker & Ullman, 2002) non-compositional phrases are not expected to be represented in the lexicon. However, frequency threshold approaches such as those advocated by Goldberg (2006) and Wray (2002)

predict that phrases over a certain frequency threshold can be represented in the lexicon. Threshold approaches posit that there is a certain frequency threshold after which frequency effects can be observed. For instance, a certain word might be expected to have at least 20 occurrences per million in a corpus to exhibit frequency effects.

Threshold approach for single words and multiword phrases presents a number of problems. The main problem is finding the threshold values. It is difficult and arbitrary to separate different frequency bins or groups by selecting certain thresholds. However, Arnon and Snider (2010) argues for a continuous model in which there is no need for threshold values rather, the frequency effects operate on words and phrases on a continuum. For instance, in a continuous model, three given words which have 10, 20, 50 occurrences per million respectively might all exhibit frequency effects albeit at different levels of strength. Moreover, in their model the difference between high frequency and low frequency phrases is not in their storage but in their activation levels. Arnon and Snider's approach predicts that there will be frequency effects even for low frequency phrases but these effects will be lower compared to high frequency phrases. Such an approach seems compatible with the premises of usage-based and connectionist approaches, as both place strong emphasis on the importance of exposure on representation.

2.4 Experimental data on multi-word frequency effects

Numerous studies show that two-word frequency has an effect on processing. Word pronunciations are generally reduced if they are part of a frequent bigram (i.e. consecutive units such as letters and syllables or words) such as *kind of* or *sort of* (Bell et al., 2003; Gregory, Raymond, Bell, Fosler-Lussier, & Jurafsky, 1999;

Jurafsky et al., 2001). This kind of a reduction is an indicator that the brain might be merging these two words into a single unit and this is an indicator of the effect of phrase frequency on the representation of linguistic units. What is more, the reduced forms of these phrases can be treated as a single word by preschool children, as they have no orthographic experience to split a seemingly holistic unit into two separate parts.

In a different study, Alegre and Gordon (1999) show that word forms that are regularly inflected such as "walked" might behave like irregular forms. This means that they might be stored in the mental lexicon holistically rather than being derived from the base form by using a plural making rule online; thus, they have lower response latency. This finding demonstrates that word storage is shaped by frequency of occurrence. Inflected forms such as *walked, talked* are not expected to be stored under the Words and Rules approach. Yet, just as Alegre and Gordon (1999) do, various studies indicate that whole-word frequency effects can be observed in high-frequency inflected forms (Baayen, Dijkstra, & Schreuder, 1997; Baayen, Schreuder, deJong, & Krott, 2002; Stemberger & MacWhinney, 1988). This might be a sign that they are stored holistically instead of being computed online.

In another study, Bod (2001) analyzed the responses to three word SVO type sentences such as "I like it" and "I keep it". The items were matched for lexical frequency as well as surface structure and complexity; thus, the results could be attributed to phrase frequency. Participants responded faster to high frequency sentences such as "I like it". However, it might be the case that the difference was due to the reaction times of the second words and not due to the frequency of the phrase. In other words, the observed effect might be due to substring frequency but

not the phrase frequency. In order to eliminate this possibility we analyze the reaction times of both the phrases and the words in those phrases in our Experiments.

Sosa and MacFarlane (2002) argue that not only single words but also two word collocations and phrases are stored in the lexicon depending on the frequency of their co-occurrence and are accessed as a whole. They used a word-monitoring task to elicit reaction times to the English function word "of" in collocations with different levels of frequency. They measured the reaction times in the recognition of two-word combinations including "of" such as "kind of" and "sort of" in an auditory word monitoring experiment. Reaction times were faster for low frequency group whereas RT was slower in the high frequency group. What is more the number of correct responses were fewer in the high frequency group indicating extra cognitive load stemming from the decomposition of a wholly stored phrase. Since high frequency phrases were stored holistically in the minds of participants, the participants had to exert extra effort to decompose them. Thus, the response latencies were higher for the highly frequent phrases. These results indicate holistic processing in high frequency collocations and the latency is most probably due to the extra effort needed to decompose the holistically stored collocations.

Diessel (2007) argues that frequency strengthens linguistic representation, expectation and automatization of chunks. In a similar vein, Bannard and Matthews (2008) demonstrate that frequency of phrases is a predictor of how fast and accurate two and three-year-olds could repeat those phrases. They show that children store multi-word phrases as part of their linguistic knowledge. Their study employed a production task involving whole-form frequency manipulation with phrases. Their items such as *a drink of milk* and *a drink of tea* were matched on all frequency measures except phrase frequency. In this case, substring frequency is also matched

as tea is as frequent as milk. The differences in the performance of children are considered as evidence that phrase frequency is at work. Their results showed that children were comparatively faster at repeating high frequency phrases. This result indicated that high frequency phrases might be stored holistically and thus they are produced faster. This shows that frequently recurring phrases leave memory traces in the brain; and thus, they exhibit a processing advantage.

Arnon and Snider (2010) investigate the effects of frequency by testing four-word phrases that only differ in the final word such as “don’t have to wait” and “don’t have to worry”. These phrases are also matched for substring frequency. Thus, the results are expected to reflect the effects of phrase frequency. To test whether frequency effects move along a continuum or are governed by a certain threshold, they included phrases from different frequency ranges. Arnon and Snider (2010) also note that although reaction times decreased as collocation frequency increased, there was a deceleration in reaction times in the most frequent collocations. Similar to Sosa and MacFarlane (2002) they attribute this slow-down to the competition between parts and the whole because in the case of high frequency phrases that also include high frequency words, there is strong activation for both the part and the whole and thus this competition might increase response latencies.

Tremblay, Derwing, Libben, and Westbury (2011) analyzed the reaction times to sentences that involved lexical bundles such as “don’t worry about it”. The results showed that compared to control sentences, participants responded faster to the sentences that included lexical bundles. This indicates that lexical bundles are stored holistically and thus they exhibit a processing advantage. The result of this study showed that frequency of phrases was a strong predictor of reading speed. In a related study, Siyanova-Chanturia, Conklin, and van Heuven (2011) used eye

tracking to explore the processing of formulaic sequences by both native and non-native speakers of English using sentences that differed in the frequency of the embedded sequences. Sequences such as “bride and groom” were used both in normal and reversed form. Their results demonstrated that response latencies for normal forms of these formulaic sequences were shorter than their reversed forms. This clearly indicates that phrases frequency effect is a valid construct as both the reversed and the non-reversed forms had the same substring frequencies and only differed in their phrase frequency.

Shaoul and Westbury (2011) state that individual word frequencies and sub-component frequencies might be confounding factors in multi-word sequences. Hence, in the case of a 3-word sequence there are many frequency values to consider. Let’s assume that A, B, and C stand for the first, second and the third words. In this case, we can talk about the frequency of A, frequency of B, frequency of C as well as the frequency of AB, BC, and ABC. The availability of multiple substring frequencies makes it necessary to control for substring frequencies in order to make sure that the results obtained in an experiment can be attributed to the independent variable. Therefore, the study of Siyanova-Chanturia, Conklin, and van Heuven (2011) does a good job at eliminating these confounding factors.

As Arnon and Cohen Priva (2013) argue, the effect of multi-word frequency on linguistic production is in line with the basic premise of the Emergentist models since the effects are not only limited to individual words but larger chunks. The experiments presented so far support the Emergentist views that emphasize the effect of linguistic experience on the nature of linguistic representation and processing. The experiments show that multi-word frequency is a predictor of reaction times in frequently recurring phrases. Arnon and Snider (2010) argue that these effects are

continuous, meaning that the effects can be observed not only in high frequency phrases but also in mid and low frequency phrases albeit with a smaller effect on reaction times. To test whether phrase frequency effects operate on a continuum, we included phrases from various frequency ranges in our experiments.

The experiments given in this chapter provide evidence against the Word and Rules approach put forward by Pinker (1999). Word and Rules approach does not predict any effects at phrase level, as phrases are thought to be computed using words and rules during processing. However, the experiments cited in this chapter clearly indicate that frequency effects are prevalent beyond single word units. What is more, these experiments demonstrate that frequently recurring phrases facilitate language processing. Taken as a whole, these experiments provide compelling evidence in support of phrase frequency effects and Emergentist models of language. However, lack of data from languages other than English and lack of substring RT controls in previous studies makes our study a welcome addition to the literature of multi-word frequency effects with both substring RT controls and various types of collocations from Turkish.

In the next chapter, we will analyze phrase frequency effects in Experiment 1 by comparing high and low frequency Adjective + Noun pairs from various frequency ranges using a Phrasal Decision Task. We will also analyze the noun frequencies in the Adjective + Noun pairs in Experiment 2 in order to eliminate substring frequency as a confounding factor.

CHAPTER 3

EXPERIMENT 1 & EXPERIMENT 2

In Experiment 1, we have tested frequent adjective & noun pairs in Turkish. We have selected 15 adjectives and for each adjective, two or more nouns with varying levels of co-occurrence. For instance, pairs such as *ağır ceza* "heavy penalty" and *ağır yara* "deep wound" which differ from each other in respect of collocation frequency were chosen to see to what extent mental priming would be at issue. The adjectives were selected randomly from a list of common adjectives we compiled from online glosses. The main criterion was to select adjectives with varying degrees of frequency so that we could observe frequency effects over a wider continuum. Additionally, we tried to avoid the use of color adjectives as they might evoke strong imagery and thus distort the results. The list of adjectives and their corresponding nouns that were used in Experiment 1 are given in Table 1.

There were two main criteria in the selection of the adjectives used in Experiment 1. Since our goal is to inspect the frequency effects in multi-word phrases, we have chosen adjectives that are relatively frequent and common in both spoken and written Turkish. Secondly, we have avoided the use of color adjectives as we thought they might lead to issues of imageability. For each adjective, we selected two nouns that occurred with that adjective in written and spoken discourse. In order to select nouns, we first searched for an adjective in the TS Corpus (Sezer, 2017). TS Corpus is an online corpus of Turkish with more than 491 million tokens and includes part of speech tagging and morphological annotation. TS Corpus uses the Boun Corpus data developed at Boğaziçi University (Sak, Güngör & Saraçlar, 2010). It includes data from three major Turkish newspapers (184 million words) as well as

sampled collection of online Turkish texts (239 million words). After searching for a certain word, TS Corpus enables users to list frequently occurring collocations of that word. We used a case insensitive search thus the system returned all occurrences of a word regardless of capitalization. Additionally, our frequency counts included the frequency counts of the inflected form of the search query. Then, we analyzed the collocations of each adjective and selected nouns that had either a high frequency of co-occurrence with the adjective or low frequency of co-occurrence. After this step, we created a list of two nouns for each adjective, one with a high collocation frequency and one with a low collocation frequency.

3.1 Experiment 1

3.1.1 Methods

3.1.2 Participants

75 participants (41 males, M_{age} = 29.2, SD = 8.84; 34 females, M_{age} = 29.2, SD = 9.06, 1 other M_{age} = 24, SD = 0) participated in the study. All participants were native speakers of Turkish. Participation in the study was on a voluntary basis and participants did not receive any benefits or incentives for their participation. 70 participants were right-handed while five participants were left-handed. The majority of the participants were college students or graduates.

Initially, 80 participants started this experiment. However, either due to technical problems or personal preference 5 participants left the experiment unfinished. Thus, their data were discarded.

3.1.3 Design

The experiment used a phrasal decision task to elicit reaction times to a randomized list of Adjective + Noun pairs such as *kuru gürültü* 'much ado for nothing' or *ağır yara* 'deep wound' as well as pairs in the form of Adjective + nonce word such as *ağır kataku*. Dependent variable was RT and the independent variable was collocational frequency. The experiment used a between-subjects design meaning two separate groups were compared.

3.1.4 Materials

In Experiment 1, a Phrasal Decision Task was carried out. The items for the Experiment 1 were generated using TS Corpus. TS Corpus is a CQP-based online corpus of Turkish with more than 450 million words. CQP (Corpus Query Processor) is a flexible query processor which is a central part of the IMS Open Corpus Workbench (Evert, 2018). It allows complex queries based on word type, category and it also provides detailed statistical information for words and phrases.

We used TS Corpus to generate a list of Adjective + Noun type of Turkish collocations such as *hızlı karar* 'fast/ quick decision'. We were able to obtain frequency data both for the individual words in each collocation and the collocation itself. Thus, we generated 15 pairs of Adjective + Noun combinations. For each adjective, we generated two Adjective + Noun pairs and in each pair, one member (Adj+Noun) had a comparatively higher collocational frequency than the other. In this context, collocation frequency means the number of times a group of words occurred together in a million-word corpus. For instance, the collocation frequency of *ağır ceza* was 30.3, which meant *ağır ceza* had 30.3 occurrences out of 1 million-

word corpus. The experimental items of Experiment 1 and the collocational frequency of all the items are given in Table 1 below.

Table 1. Items in Experiment 1

	Freq	First Word	Second Word	Gloss	Collocation Frequency
1.	High	ağır	ceza	heavy penalty	30.3
	Low	ağır	yara	deep wound	0.75
2.	High	aşırı	hız	excessive speed	3.67
	Low	aşırı	kilo	excessive weight	2.58
3.	High	canlı	müzik	live music	2.64
	Low	canlı	bitki	live plant	0.07
4.	High	derin	nefes	deep breath	1.29
	Low	derin	üzüntü	deep sorrow	1.17
5.	High	dik	yamaç	steep slope	0.3
	Low	dik	yol	steep road	0.01
6.	High	doğal	yaşam	natural life	1.43
	Low	doğal	ekmek	natural bread	0.03
7.	High	düz	çizgi	straight line	0.42
	Low	düz	taban	flat feet	0.2
8.	High	hızlı	tren	high speed train	4.95
	Low	hızlı	karar	quick decision	1.02
9.	High	kötü	niyet	bad intention	7.47
	Low	kötü	görüş	bad view	0.01
10.	High	kuru	gıda	dry food	0.59
	Low	kuru	gürültü	much ado about nothing	0.48
11.	High	mutlu	son	happy end	7.77
	Low	mutlu	çift	happy couple	0.21
12.	High	parlak	fikir	bright idea	0.51
	Low	parlak	ayna	bright mirror	0.01
13.	High	sert	söz	harsh word	0.43
	Low	sert	plastik	rigid plastic	0.11
14.	High	temiz	hava	clean air	2.12
	Low	temiz	iş	clean job	0.12
15.	High	yüksek	lisans	master’s degree	20.76
	Low	yüksek	gerilim	high voltage	2.07

Our goal was to choose Adjective + Noun pairs with varying degrees of frequency. The highest collocation frequency was 30.3 per million while the lowest collocation frequency was 0.01 per million. Since we wanted to explore whether frequency effects operate on a threshold model or on a continuous basis, we selected our pairs at various levels of frequency. Thus if we observed an effect of collocation frequency we could make the argument that these effects could be observed over a continuum rather than after specific frequency thresholds.

In the next step, we generated 15 pairs of Adjective + Nonce combinations such as *hızlı gubaro* to serve as fillers in the experiment.

3.1.5 Procedure

The experiment was prepared and carried out online using the behavioral research application *Gorilla* (Behavioural Science In The Cloud, 2018). *Gorilla* allows millisecond precision in online experiments. All the data was collected online using the *Gorilla* experiment tool interface. The participants were requested to fill out a consent form before taking part in the experiment.

The experiment can only be run on desktop computers. Participants were first sent information about the experiment and how to use the system. Participants were also instructed to go to a quiet room, put away their devices and close other browser tabs in order to fully focus on the experiment. After they understood and agreed to the terms of the experiment, they started the experiment by clicking on their special participation URL.

In the Phrasal Decision Task, subjects were asked to decide whether a phrase they saw on the screen was acceptable in Turkish or not. If they thought that it was an acceptable phrase they were required to press the E (EVET ‘yes’) button and if

they thought the phrase was not acceptable in Turkish they were required to press the H (HAYIR ‘no’) button as quickly as possible. Participants first read the instructions and completed a trial run before moving on to the experimental items. The subjects initially saw eight phrases on the screen one by one to familiarize themselves with the experiment setup. Four of the practice items were acceptable Turkish phrases while the remaining four were not acceptable in Turkish. Those eight phrases were excluded from data analysis as they were meant to be a warm up practice for participants.

After the practice run, participants continued the task with the experimental items. Participants were presented with an asterisk (*) for 1000 milliseconds before and after each phrase was shown. The items were presented in a randomized order for all participants. There was also a 1000 millisecond pause before and after each phrase was shown. In order to move to the next item, participants were required to provide a keyboard response by either pressing E or H which meant “acceptable” and “not acceptable” respectively. The experiment took around 7-10 minutes.

Participants were not given any feedback during the experiment. Questions were untimed which meant subjects could take time to think as much as they wanted. Reaction times were measured in milliseconds. RT values were extracted from the time between the onset of each stimuli and a participant’s keyboard response to that stimuli.

3.1.6 Results

In this section, we discuss the results of Experiment 1. Let us first report on the statistical analyses of the data. First, we lay out the results of a two tailed unpaired t test performed on the mean reaction times for low frequency items and high frequency items. This task was carried out in order to explore the role of

collocational frequency on RT of Adjective + Noun combinations. We also provide detailed tables on the RTs for low and high frequency items in Experiment 1 in the Appendix. These data allow us to analyze the relationship between collocation frequency and RT.

Overall, more than 95% of responses to test items were accurate. Rosner's Extreme Studentized Deviate Test for multiple outliers (Rosner, 1975) was used at 0.5 significance level to eliminate outliers for each item. There were five outliers on average for each response. All of these outliers were discarded from data analysis.

Recall that our hypothesis was that two word adjective + noun collocations with high frequency would be stored as a whole in the mental lexicon and thus the reaction times for these items would be shorter compared to low frequency collocations. As Table 2 reveals the mean reaction times for the high frequency collocations – as indicated in bold – were much shorter compared to low frequency collocations.

An unpaired t-test was conducted to compare RT scores for the high frequency and low frequency collocation conditions. There was a significant difference in the scores for high frequency collocation (M=693, SD=41.29) and low frequency collocation (M=758, SD=54.94) conditions; t(28)=3.66, p=0.001. Results reveal that there is a significant difference between the mean reaction times of collocations in the high frequency group and the collocations in the low frequency group. Figure 1 shows the interaction between collocation frequency and RT.

Table 2. Mean RTs for Adjective + Noun Collocations

Group	Mean	SD
High Frequency Adjective + Noun Collocation e.g. ağır ceza 'heavy penalty'	693 milliseconds	41.29
Low Frequency Adjective + Noun Collocation e.g. ağır yara 'deep wound'	758 milliseconds	54.94

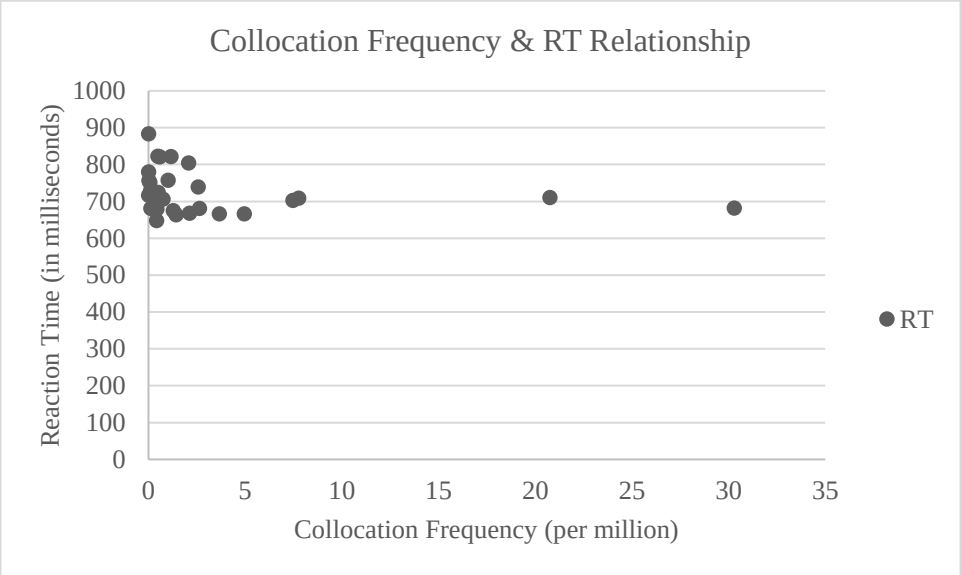


Figure 1. Collocation frequency and RT

It is interesting to note that the corpus used in this study is partly biased towards non-fiction as most of the text it includes comes from online written materials such as websites and blogs.

3.1.7 Discussion

The results indicate that there is an effect of collocation frequency on reaction times as shown by the t-test. For each adjective + noun pair, we looked at how fast subjects responded to the more frequent adjective + noun combinations compared to the less frequent ones. For instance we compared the reaction times for *ağır ceza* and *ağır*

yara whose collocation frequencies are 30.3 and 0.75 per million respectively. Similarly we looked at *aşırı hız* and *aşırı kilo* whose collocation frequencies stand at 3.67 and 2.58 per million respectively. In all the 15 Adjective + Noun pairs, the Adjective + Noun combination with a higher collocation frequency had a shorter RT.

The results show that collocation frequency affects RT at various levels. Collocation frequency seems to play a role both at 0.51 per million level as well as 30.3 per million level. This is an indicator that phrase frequency effects do not operate on a threshold basis. If this were the case, phrase frequency effects would not be observed for collocations at various levels of collocation frequency. The data clearly shows that collocation frequency effects operate on a continuum and can be observed at different levels of collocation frequency. The comparison of *ağır ceza* and *ağır yara* reveals that the RT for *ağır ceza* was 682 milliseconds while the RT for *ağır yara* was 707 milliseconds. The per million collocation frequencies of the phrases stood at 30.3 and 0.75 respectively. At the opposite end of the continuum, *parlak fikir* 'bright idea' had an RT of 725 milliseconds whereas the RT for *parlak ayna* 'bright/ shiny mirror' was 884 milliseconds. Their per million collocation frequencies are 0.51 and 0.01 respectively. The comparison of *temiz hava* and *temiz iş* with collocational frequencies of 2.12, and 0.12 respectively reveals that their RT times followed a similar trajectory with 668 and 681 milliseconds.

The results of Experiment 1 proved to be in parallel with our initial assumptions. However, Experiment 1 posed two challenges. First, although there is a significant difference between high and low frequency collocations, some of this difference might also be attributed to single word frequency of individual items rather than the overall collocational frequency, hence it is necessary to explore whether these effects were due to single word frequency. Secondly, a closer look at

the items in this experiment, reveals an interesting point. In designing the experiment, we had intuited that some of the collocations in Experiment 1, in fact, may evoke longer collocations. For instance, *hızlı karar* 'quick decision' may evoke *hızlı karar vermek* 'make a quick decision'. Similarly, *derin nefes* 'deep breath' may evoke *derin nefes almak* 'take a deep breath'. This observation raises an interesting question. What is the nature of the relationship between 2-word frequency (e.g. *hızlı karar*) and 3-word frequency (e.g. *hızlı karar vermek*)? It might be that three-word frequency overrides two-word frequency or alternatively three-word frequency might either hinder or complement two-word frequency.

We conjectured that all of these instances require further examination as there are probably other factors interacting with collocational frequency. This issue brings forth the question of whether there is any reaction time difference between phrases such as *derin nefes* 'deep breath' and *derin nefes almak* 'to take a deep breath' or whether a wider or higher level of phrase frequency overrides the expected collocational frequency effects in cases like *derin nefes* or *hızlı karar*.

We devised two more experiments in order to address these issues. In Experiment 2, we carried out a simple Lexical Decision Task to find out the reaction times for all the nouns in the adjective + noun combinations in Experiment 1 in order to find out the relationship between single word frequency and collocation frequency. Hence, in what follows we will first present Experiment 2.

3.2 Experiment 2

3.2.1 Method

3.2.2 Participants

37 participants (15 males, M_{age} = 31.86, *SD* = 12.17; 21 females, M_{age} = 28.42, *SD* = 8.34, 1 other M_{age} = 24, *SD* = 0) participated in the study. All participants were native speakers of Turkish. Participation in the study was on a voluntary basis and participants did not receive any benefits or incentives for their participation. 33 participants were right-handed while four participants were left-handed. The majority of the participants were college students or graduates.

40 participants started this experiment. However, either due to technical problems or personal preference 3 participants left the experiment unfinished. Thus, their data were discarded.

3.2.3 Design

In Experiment 2, we used a lexical decision task to elicit reaction times to the randomized list of nouns from Experiment 1 as well as nonce words such as *gutaba*. Dependent variable was RT and the independent variable was word frequency. The Experiment used a between-subjects design meaning two separate groups were compared.

3.4.3 Materials

In Experiment 2, a lexical decision task was implemented. The items for Experiment 2 were the nouns that were used in the Adjective + Noun combinations in

Experiment 1. We used TS Corpus again. Nonce words were generated based on a Japanese word list which included words resembling Turkish syllable structure.

3.2.4 Procedure

The procedure of Experiment 2 was the same as that of Experiment 1. The experiment was prepared and carried out online using the behavioral research application *Gorilla* (*Behavioural Science In The Cloud*, 2018). All the data were collected online using the *Gorilla* experiment tool interface. All participants were presented with the consent form before taking part in the experiment. The rest of the procedure and data processing followed the step in Experiment 1.

3.2.5 Results

In this part, we report on the statistical analyses of the data in Experiment 2. First, we report the mean reaction times for all of the words. Then we state the results of a two tailed unpaired t-test performed on the mean reaction times of single words and collocations. This task was carried out in order to explore the role of single word frequency on the RT of Adjective + Noun combinations. We also provide detailed tables on the RTs for single words and collocations in Experiment 1. These data allow us to analyze the relationship between collocation frequency and single word frequency.

Overall, more than 95% of responses to test items were accurate. Rosner's Extreme Studentized Deviate test for multiple outliers (Rosner, 1975) was used at 0.5 significance level to eliminate outliers for each item. There were six outliers on average for each response. All of these outliers were discarded from data analysis.

Our hypothesis was that collocations with high frequency would be stored as a whole in the mental lexicon and thus the reaction times for these items would be shorter compared to low frequency collocations. Experiment 1 showed this to be true. However, in order to ensure that this effect stems from collocation frequency and not from single word frequency, we need to show that collocation frequency and word frequency are different entities. We accomplished this by comparing the reaction times to Adjective + Noun combinations and the reaction times to the nouns in these combinations.

In Experiment 2, we carried out a simple lexical decision task to extract the reaction times for the nouns in Experiment 1. Our aim was to analyze the effect of single word RT on collocation RT. We also analyzed TS Corpus to find out per million frequencies of all nouns in Experiment 1. This kind of data gives us a broad and encompassing picture of all the variables at hand. As a result of these data points, we were able to analyze the interaction between single word RT, collocation RT, single word frequency and collocation frequency. This is essential to the aim of this study, as we want to demonstrate that frequency effects exist not only at single word level but also at multi-word level. We also want to illustrate that collocation frequency can override single word frequency. If this can be exemplified, it provides data in favor of collocation frequency.

In Experiment 2, we have first analyzed the reaction times and frequencies of all nouns in Experiment 1. This data provides a good starting point, however, for it to be more valuable we have decided to combine the data from Experiment 1 and Experiment 2 in a single table as in Table 3 below. Table 4 provides the mean RTs for adjective + noun collocations and nouns while Figure 2 visualizes the interaction between mean RTs for adjective + noun collocations and nouns.

Table 3. Nouns in Experiment 2

	Adj	Noun	Noun Freq	Noun RT	Coll Freq	Coll RT
1	ağır	ceza	650.15	601	30.3	682
	ağır	yara	107.44	616	0.75	707
2	aşırı	hız	497.7	583	3.67	667
	aşırı	kilo	116.55	618	2.58	740
3	canlı	müzik	302.38	601	2.64	681
	canlı	bitki	61.83	624	0.07	752
4	derin	nefes	65.08	586	1.29	675
	derin	üzüntü	39.51	586	1.17	822
5	dik	yamaç	9.73	650	0.3	707
	dik	yol	1252.57	541	0.01	717
6	doğal	yaşam	326.75	623	1.43	664
	doğal	ekmek	69.77	611	0.03	757
7	düz	çizgi	117.55	573	0.42	648
	düz	taban	71.61	588	0.2	714
8	hızlı	tren	61.3	577	4.95	667
	hızlı	karar	1369.78	640	1.02	758
9	kötü	niyet	127.97	557	7.47	703
	kötü	görüş	654.68	609	0.01	780
10	kuru	gıda	105.73	632	0.59	821
	kuru	gürültü	28.55	614	0.48	823
11	mutlu	son	1932.3	594	7.77	709
	mutlu	çift	121.19	618	0.21	713
12	parlak	fikir	197.78	635	0.51	725
	parlak	ayna	38.02	631	0.01	884
13	sert	söz	1026.54	615	0.43	680
	sert	plastik	26.76	614	0.11	729
14	temiz	hava	388.15	622	2.12	668
	temiz	iş	1893.08	607	0.12	681
15	yüksek	lisans	95.1	661	20.76	711
	yüksek	gerilim	50.27	692	2.07	805

Table 4. Mean RTs for Adjective + Noun Collocations vs Nouns

Group	Mean	SD
Adjective + Noun Collocation e.g. ağır yara 'heavy wound	726 milliseconds	57,08
Noun e.g. yara 'wound'	610 milliseconds	29,95

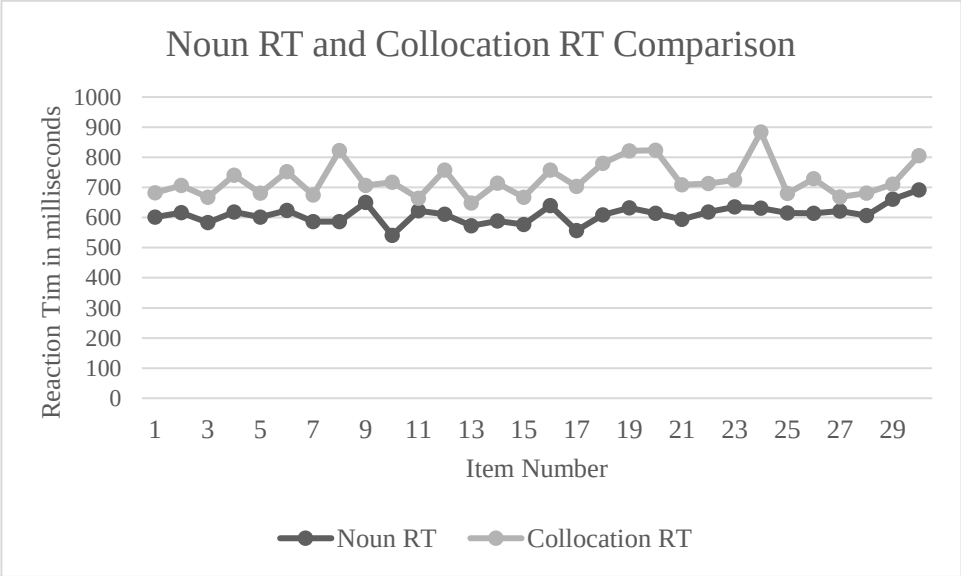


Figure 2. Mean RTs for adjective + noun collocations vs nouns

An unpaired t-test was conducted to compare RT scores for the single word and collocation conditions. There was a significant difference in the scores for single word (M=610.63, SD=30.46) and collocation (M=726.33, SD=58.06) conditions; t(58)=9.66, p=0.0001. The results reveal that there is a significant difference between the mean reaction times of single word frequencies and mean reaction times of their corresponding collocations.

3.2.6 Discussion

One of the important questions from Experiment 1 was whether the difference in RT was simply due to the frequency of the noun in the Adjective + Noun combination.

In fact, this is a very valid concern. One can argue that since both *ağır ceza* and *ağır yara* have *ağır* at the beginning, the real differentiator might be the noun part not the frequency of the whole collocation. This question is difficult to answer based on the *ağır ceza* and *ağır yara* case because the collocation frequency of these items are 30.3 and 0.75 respectively. Similarly, their noun frequencies follow a similar trend with 650.15 and 107.44 respectively. In order to demonstrate that collocation frequency really exists, we need collocation pairs in which despite the higher noun frequency of one of the pairs, the collocation with the higher collocation frequency receives lower RTs.

A look at the data reveals four such pairs. These are

- *Dik yol / dik yamaç:* steep road / steep slope
- *Hızlı tren / hızlı karar:* high speed train / quick decision
- *Kötü niyet / kötü görüş:* bad intention / bad view
- *Temiz hava / temiz iş:* clean air / clean job

Reaction time and frequency values for these pairs are given in Table 5 below.

Table 5. Selected Items from Experiment 2

Adjective	Noun	Noun Frequency	Noun RT	Collocation Frequency	Collocation RT
dik	yamaç	9.73	650	0.3	707
dik	yol	1252.57	541	0.01	717
hızlı	tren	61.3	577	4.95	667
hızlı	karar	1369.78	640	1.02	758
kötü	niyet	127.97	557	7.47	703
kötü	görüş	654.68	609	0.01	780
temiz	hava	388.15	622	2.12	668
temiz	iş	1893.08	607	0.12	681

In the case of *dik yamaç* and *dik yol*, the word *yol* has a much higher frequency, 1252.57 compared to 9.73 for *yamaç*. *Yol* is a much shorter word consisting of only three sounds. What is more, *yol* has an RT of 541 milliseconds while *yamaç* has an RT of 650 milliseconds. A comparison of these two words shows that *yol* is a clear winner. However, when we look at these two words in the collocations such as *dik yol* and *dik yamaç*, an interesting picture emerges. *Dik yol* has an RT of 717 milliseconds while *dik yamaç* has an RT of 707 milliseconds. Normally one would expect *dik yol* to have a much shorter RT than *dik yamaç* but this is not the case. Yet, let alone having a comparatively shorter RT, *dik yol* has in fact a longer RT than *dik yamaç*. This pair makes a strong case for collocation frequency because when we compare both nouns in isolation, single word frequency determines which of them will have a shorter RT but when the same nouns become a part of a collocation then the collocation frequency determines which of them will have a lower RT.

A similar case is observed in the collocation of *temiz hava* 'clean air' and *temiz iş* 'clean job'. *Hava* has a word frequency of 388.12 while *iş* has a word frequency of 1893.08. Their single word RTs also follow a similar trend. *Hava* has a longer RT at 622 milliseconds whereas the RT for *iş* stands at 607 milliseconds. However when both words are combined with *temiz* to form a collocation, things start to change. The RT for *temiz hava* is 668 milliseconds whereas the RT for *temiz iş* stands at 681 milliseconds. Again, in this case, too, collocation frequency overrides single word frequency otherwise; we would expect the RT for *temiz iş* to be shorter than the RT for *temiz hava*.

The ensuing chapter will present Experiment 3, where we have explored an extended list of two and three-word collocations with different structures such as

collocations with *ve* ‘and’ and their reversed forms and the comparison of two-word collocations to their initial words. The aim of Experiment 3 is to inspect the effect of collocation frequency on different types of collocations such as the ones formed with “*ve*” ‘and’ as in *gelin ve damat* ‘bride and groom” and *hesap kitap* ‘financial arrangement’ and to explore the reasons why some of the items in Experiment 1 did not behave as expected.

In the next chapter, we will explore three different types of phrases in order to obtain a wide range of data on phrase frequency effects. We will compare collocations formed with *ve* ‘and’ such as *gelin ve damat* ‘bride and groom’, common two word collocations such as *çarşı pazar* ‘bazaar market’ and extended forms of Adjective + Noun collocations such as *hızlı karar vermek* ‘to make a quick decision’.

CHAPTER 4

EXPERIMENT 3

In Experiment 3, we explore two additional collocation types to obtain a more comprehensive data set between phrase types and phrase frequency effects. We test collocations formed with *ve* 'and' against their reversed forms and then we test two-word idiomatic collocations against their initial words. We expect to see phrase frequency effects for the collocation conditions. For instance, we expect that the RTs will be significantly faster for the regular forms of collocations compared to their reversed forms although they are both plausible and share the same words. Similarly, we expect that when compared to their initial words, two word idiomatic collocations will provide a processing advantage and there will not be a significant difference between the single word and two-word collocation conditions.

4.1 Method

4.1.1 Participants

There were two groups in Experiment 3. Group A had 40 participants. However, either due to technical problems or personal preference 6 participants left the experiment unfinished. Thus, their were was discarded. 34 participants (16 males, M_{age} = 23, SD = 9.06; 18 females, M_{age} = 32.38, SD = 10.50) participated in the study. All participants were native speakers of Turkish. Participation in the study was on a voluntary basis and participants did not receive any benefits or incentives for their participation. 28 participants were right-handed while 6 participants were left-handed.

Group B also had 40 participants. Yet, either due to technical problems or personal preference 3 participants left the experiment unfinished. Thus, their data was discarded. 37 participants (19 males, M_{age} = 29, *SD* = 9.81; 16 females, M_{age} = 29.87, *SD* = 8.75, 2 others, M_{age} = 21.5, *SD* = 3.5) participated in the study. All participants were native speakers of Turkish. Participation in the study was on a voluntary basis and participants did not receive any benefits or incentives for their participation. 34 participants were right-handed while three participants were left-handed.

4.1.2 Design

In Experiment 3, we have used a Lexical Decision Task and a Phrasal Decision Task to elicit reaction times to a randomized list of nouns and collocations as well as nonce words. Dependent variable was RT and the independent variable was word frequency or collocation frequency. Experiment 3 employed a between-subjects design meaning two separate groups were compared.

4.1.3 Materials

There were three sets of items in Experiment 3. We used the TS Corpus, which is an online corpus of Turkish with more than 450 million words to extract frequency data for each noun and collocation. Nonce words were generated based on a Japanese word list.

4.1.3.1 Three-word collocations

The first set consisted of 3-word phrases such as *hızlı karar vermek* 'to make a quick decision'. This set was created in order to compare the frequency effects and RT

differences between two and three-word pairs. All the 3-word pairs were in fact extended versions of some of the selected 2-word pairs from Experiment 1. The list of items is given below in Table 6. The reason why we decided to test extended forms of the items in Experiment 1 is that some of the items in Experiment 1 may have evoked longer phrases in the minds of native speakers of Turkish. For instance *hızlı karar* 'quick decision' is a very common collocation but upon reading the first two words one also remembers *hızlı karar vermek* 'to make a quick decision' which is also a very common phrase. Thus, in examples such as *hızlı karar* 'quick decision', *derin üzüntü* 'deep sorrow', *derin nefes* 'deep breath, the adjectives in the Adj+ Noun collocations appear to prime adverbial uses, hence evoking verbs such as vermek 'give' for *hızlı karar*, duymak 'feel' for *derin üzüntü*, etc. Similarly examples such as *ağır ceza* or *doğal yaşam* may evoke further nouns where they function as parts of compounds such as *ağır ceza hakimi*/ *mahkemesi* 'high criminal judge/court, *doğal yaşam alanı* 'habitat'. Instances such as these prompted us to explore these extended collocations. This is important because it can inform us about the existence of phrase frequency effects beyond a two-word limit which can either inhibit or strengthen phrase frequency effects for shorter phrases and it can also inform us about potential conflicts between a two-word collocation and a three-word collocation starting with the same two words. To our knowledge, we are the first in the literature to explore the idea that XY can evoke XYZ.

4.1.3.2 Collocations with *ve* and their reversed forms

The second set of items consisted of collocations that included the conjunction *ve* 'and' in Turkish as in *kedi ve köpek 'cat and dog'*. In the second set, collocations and their reversed forms such as '*köpek ve kedi*' were compared with regards to their

collocation frequency and RT. The items used in this Experiment are given in Table 7 below. The structure of these collocations is Noun + ve + Noun. All of these collocations are compositional in nature, which means that their meaning can be obtained from their individual parts.

4.1.3.3 Two-word collocations and their initial words

The last set of items consisted of a list of 2-word collocations and these items were contrasted with the first words of the collocations in terms of RTs. For instance, *çarşı pazar* 'bazaar market' was compared with *çarşı* 'bazaar' in terms of its RT and frequency. The list of items is given below in Table 8. The reason why we have chosen collocations such as the ones in Experiment 3.3 is that they provide a different type of collocation to extend our data set. The collocations in Experiment 1 are Adjective + Noun combinations, the collocations in Experiment 3.1 are three-word collocations derived from the two-word collocation list in Experiment 1. The collocations in Experiment 3.2 are collocations in the form of Noun + *ve* + Noun.

We divided the list of all items into two groups. Each group had items from Set 1, Set 2 and Set 3. The items were given to two different groups in a combined Lexical Decision and Phrasal Decision Task.

Table 6. Items in Experiment 3.1

Adj + N	Frequency	Adj + Noun RT	3 Words	Gloss	Frequency
ağır ceza	30.3	682	ağır ceza hakimi	high criminal (court) judge	0.02
derin nefes	1.29	675	derin nefes almak	to take a deep breath	1.08
derin üzüntü	1.17	822	derin üzüntü duymak	to feel deep sorrow	0.5
doğal yaşam	1.43	664	doğal yaşam alanı	natural habitat	0.08
hızlı karar	1.02	758	hızlı karar vermek	to make a quick decision	0.33
kuru gürültü	0.48	823	kuru gürültü yapmak	to make a lot of nonsense noise	0.02
yüksek gerilim	2.07	805	yüksek gerilim hattı	high tension line	0.39
yüksek lisans	20.76	711	yüksek lisans öğrencisi	master’s student	0.69

Table 7. Items in Experiment 3.2

	Items	Gloss	Frequency Of the 1st Word per million	Frequency Of the 2nd Word per million	Frequency Of the Collocation per million
1	ağız ve diş	mouth and tooth	18.15	45.59	1.45
	diş ve ağız	tooth and mouth			0.11
2	gelin ve damat	bride and groom	37.79	11.45	0.37
	damat ve gelin	groom and bride			0.03
3	güzel ve çirkin	beautiful and ugly	616	21	0.25
	çirkin ve güzel	ugly and beautiful			0.1
4	kedi ve köpek	cat and dog	20	22	0.22
	köpek ve kedi	dog and cat			0.03
5	köprü ve otoyol	bridge and highway	30	7	0.83
	otoyol ve köprü	highway and bridge			0.26
6	suç ve ceza	crime and punishment	126	271	1.61
	ceza ve suç	punishment and crime			0.01
7	çevre ve şehircilik	environment and urban planning	114	2	0.2
	şehircilik ve çevre	urban planning and environment			0.1

Table 8. Items in Experiment 3.3

	Items	Gloss	Frequency Of the 1st Word per million	Frequency Of the 2nd Word per million	Frequency Of the Collocation per million
1	çarşı pazar	bazaar market	8.3	165.47	0.45
	çarşı	bazaar			
2	ipsiz sapsız	footloose	0.16	0.19	0.05
	ipsiz	vagabond			
3	dolu dizgin	full speed	102.25	0.85	0.74
	dolu	full			
4	hesap kitap	financial arrangement	72.64	143.6	0.88
	hesap	calculation			
5	hal hatır	State	34.32	1.98	0.49
	hal	mood			
6	iş güç	job power	420.67	135.97	0.57
	iş	job			
7	çanak çömlek	pottery	6.5	1.1	0.37
	çanak	bowl			
8	iğne iplik	needle yarn	7.3	3.69	0.02
	iğne	needle			

4.1.4 Procedure

The experiment was prepared and carried out online using the behavioral research application *Gorilla* again.

The experiment could only be run on desktop computers. Participants were first informed about the experiment and the functioning of the *Gorilla* experimental software. All participants were required to complete the experiment in a quiet room and close tabs in their browser that were not related to the experiment. Then, they started the experiment by clicking on their special participation URL.

In the experiment, subjects were asked to decide whether a word or phrase they saw on the screen was acceptable in Turkish or not. If they thought that it was acceptable they were required to press the E button and if they thought it was not acceptable in Turkish they were required to press the H button as quickly as possible.

Participants first read the instructions and completed a trial run before moving on to the real items. The subjects initially saw eight words on the screen one by one to familiarize themselves with the experiment setup. Four of these practice items were acceptable Turkish words while the remaining four items were not acceptable in Turkish. Those eight words were excluded from data analysis as they were meant to be a warm up practice for participants.

After the practice run, participants continued the task with real items. Participants were presented with an asterisk (*) for 1000 milliseconds before and after each word or phrase was shown. The order of items was presented randomized for all participants. There was also a 1000-millisecond pause before and after each phrase was shown. In order to move to the next item, participants were required to provide a keyboard response by either pressing E or H which meant “acceptable” and “not acceptable” respectively. The experiment took around 7-10 minutes.

Participants were not given any feedback during the experiment. Questions were untimed, which meant subjects could take as much as time as they wanted to decide on an answer. Reaction times were measured in milliseconds. RT values were extracted from the time between the onset of each stimuli and a participant’s keyboard response to that stimuli.

4.2 Results

In this part, we report on the statistical analyses of the data in Experiment 3 regarding three different sets of experimental items.

Overall, more than 95% of responses to test items were accurate. Rosner's Extreme Studentized Deviate test for multiple outliers (Rosner, 1975) was used at 0.5

significance level to eliminate outliers for each item. There were six outliers on average for each response. All of these outliers were discarded from data analysis.

4.2.1 Three-word collocations

In this part, we report on the statistical analyses of the data in Experiment 3 regarding 3-word collocations. First, we report the mean reaction times for all of the 3-word collocations. Then we state the results of a two tailed unpaired t test performed on the mean reaction times of 2-word collocations from Experiment 1 and 3-word collocations in Experiment 3. This task was carried out in order to explore the role of 3-word frequency on the RT of Adjective + Noun combinations. We also provide detailed tables on the RTs for 2-word and 3-word phrases in Table 9. This data allows us to analyze collocation frequency at different levels. Table 10 provides mean RTs and SDs for two-word and three-word collocations while Figure 2 visualizes the RT differences between two-word and three-word collocations.

An unpaired t-test was conducted to compare RT scores for the Adjective + Noun and 3-word collocation conditions. There was not a significant difference in the scores for Adjective + Noun (M=742.50, SD=67.95 and 3-word collocation (M=753.75, SD=48.98) conditions; t(14)=0.38, p=0.71. This means that there is statistically no significant difference between the reaction times of 2-word collocations and the 3-word collocations that were tested in this study. This means that the RT values are almost the same for both the 2-word and the 3-word collocation groups despite an additional word in the 3-word group.

In Experiment 1, we have observed that a number of collocations such as *hızlı karar* evoked longer collocations such as *hızlı karar vermek*. In the first part of Experiment 3, we analyzed extended forms of such collocations from Experiment 1.

We added verbs or nouns to some of these selected collocations from Experiment 1. The list of items is given in the table below. We wanted to see how the RTs of the extended forms would compare to the RTs of the collocations from Experiment 1. This serves two purposes. The first is to see if there are frequency effects at 3-word level. Secondly, it helps us to further our understanding of the interaction between frequency effects at single word, two-word and three-word levels.

Table 9. Comparison of Two and Three-word Collocations in Experiment 3.1

Adj + N	Frequency	Adj + N RT	3 Words	Frequency	3 Word RT
ağır ceza	30.3	682	ağır ceza hakimi	0.02	684
derin nefes	1.29	675	derin nefes almak	1.08	733
derin üzüntü	1.17	822	derin üzüntü duymak	0.5	836
doğal yaşam	1.43	664	doğal yaşam alanı	0.08	744
hızlı karar	1.02	758	hızlı karar vermek	0.33	709
kuru gürültü	0.48	823	kuru gürültü yapmak	0.02	764
yüksek gerilim	2.07	805	yüksek gerilim hattı	0.39	804
yüksek lisans	20.76	711	yüksek lisans öğrencisi	0.69	756

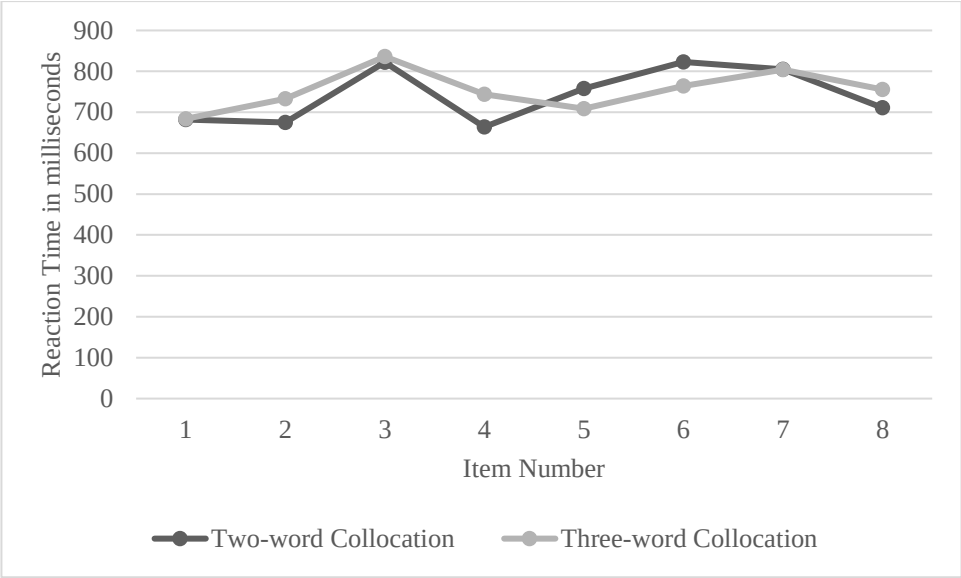


Figure 3. Comparison of two-word and three-word collocations

Table 10. Mean RTs and SDs for Two-word and Three-word Collocations

Group	Mean	SD
Two-Word Collocation e.g. hızlı karar 'quick decision'	742 milliseconds	63,55
Three-Word Collocation e.g. hızlı karar vermek 'to make a quick decision'	753 milliseconds	45,81

The analysis of the results provides some striking figures. The RT for *hızlı karar* was 758 milliseconds. Normally one would expect the RT for *hızlı karar vermek*, which is a longer phrase to be comparatively longer. Yet, surprisingly the RT for *hızlı karar vermek* was 709 milliseconds. *Kuru gürültü* and *kuru gürültü yapmak* follow a similar trend with RTs of 823 milliseconds and 764 milliseconds respectively. Finally the RTs for *ağır ceza, derin üzüntü* and *yüksek gerilim* are almost the same as their extended counterparts *ağır ceza hakimi, derin üzüntü duymak* and *yüksek gerilim hattı*.

4.2.2 Collocations with *ve* and their reversed forms

In this part, we report on the statistical analyses of the data in Experiment 3 regarding collocations and their reversed forms. First, we report the mean reaction times for all of the collocations in Set 2. Then we state the results of a two tailed unpaired t test performed on the mean reaction times of 2-word collocations and their reversed forms. This task was carried out in order to explore the role of collocation frequency in lexical access. We also provide detailed tables on the RTs of collocations in Set 2. This data allows us to analyze collocation frequency from different perspectives. The details are given in Table 11 below. The items are similar to the ones in Siyanova-Chanturia, Conklin, and van Heuven (2011). Table 12 provides mean RTs and SDs for collocations with 've' and their reversed forms while Figure 4 visualizes the RT differences between three-word collocations and their reversed forms.

An unpaired t-test was conducted to compare RT scores for the high frequency collocation and low frequency collocation conditions. There was a significant difference in the scores for high frequency collocation (M=646.14, SD=38.39) and low frequency collocation (M=705.29, SD=53.28) condition; $t(12)=2.38$, $p=0.0346$. This shows that word order and in this case, collocation frequency played a significant role in response latencies. Since both the typical and the reversed form of the collocations in the second set were acceptable and had the same word frequencies, any change in RT can be attributed to collocation frequency.

Table 11. Collocations with "*ve*"

Items	RT	Frequency Of the 1st Word per million	Frequency Of the 2nd Word per million	Frequency Of the Collocation per million
ağız ve diş	607	18.15	45.59	1.45
diş ve ağız	710	45.59	18.15	0.11
gelin ve damat	661	37.79	11.45	0.37
damat ve gelin	712	11.45	37.79	0.03
güzel ve çirkin	614	616	21	0.25
çirkin ve güzel	664	21	616	0.1
kedi ve köpek	637	20	22	0.22
köpek ve kedi	653	22	20	0.03
köprü ve otoyol	723	30	7	0.83
otoyol ve köprü	753	7	30	0.26
suç ve ceza	641	126	271	1.61
ceza ve suç	653	271	126	0.01
çevre ve şehircilik	640	114	2	0.2
şehircilik ve çevre	792	2	114	0.1

Table 12. Mean RTs and SDs for Collocations with *'ve'* and Their Reversed Forms

Group	Mean	SD
Collocations with 've' e.g. gelin ve damat 'bride and groom'	646 milliseconds	35,53
Reversed Collocations e.g. damat ve gelin 'groom and bride'	705 milliseconds	49,33

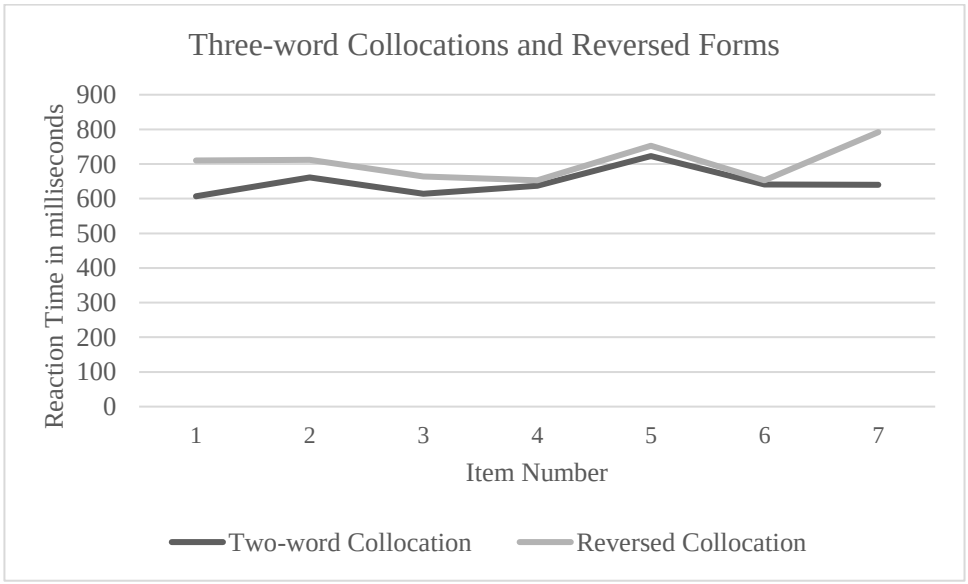


Figure 4. Three-word collocations and their reversed forms

4.2.3 Two-word collocations and their initial words

In this part, we report on the statistical analyses of the data in Experiment 3 regarding RT for collocations and their first words. First, we report the mean reaction times for all of the collocations and their first words in Set 3. Then we state the results of a two tailed unpaired t test performed on the mean reaction times of collocations and their first words. This task was carried out in order to explore the role of collocation frequency on RTs. We also provide detailed tables on the RTs collocations in Set 3. These data enable us to analyze collocation frequency from different perspectives.

An unpaired t-test was conducted to compare RT scores for single word and collocation conditions. There was not a significant difference in the scores for single word (M=665.25, SD=64.47) and collocation (M=707, SD=54.15) condition; $t(14)=1.40$, $p= 0.18$. This means that there is no statistically significant difference between the reaction times for collocations and their first words although collocations were much longer. This is an indicator that collocation frequency exists. Otherwise, there would be a significant difference between the RTs of collocations and their first words. A complete list of all items in Experiment 3.3 and their RTs are reported in Table 13. Table 14 provides information about mean RTs and SDs for two-word collocations and their initial words while Figure 5 visualizes the RT differences between two-word collocations and their initial words.

Table 13. RT for Two-word Collocations and Initial Nouns in Experiment 3.3

Items	RT	Frequency Of the 1st Word per million	Frequency Of the 2nd Word per million	Frequency Of the Collocation per million
çarşı pazar	600	8.3	165.47	0.45
çarşı	665	165.47		
ipsiz sapsız	768	0.16	0.19	0.05
ipsiz	780	0.19		
dolu dizgin	699	102.25	0.85	0.74
dolu	696	0.85		
hesap kitap	704	72.64	143.6	0.88
hesap	608	143.6		
hal hatır	779	34.32	1.98	0.49
hal	728	1.98		
iş güç	705	420.67	135.97	0.57
iş	618	135.97		
çanak çömlek	703	6.5	1.1	0.37
çanak	623	1.1		
iğne iplik	698	7.3	3.69	0.02
iğne	604	3.69		

Table 14. Mean RTs and SDs for Two-word Collocations and Their Initial Words

Group	Mean	SD
Two-word collocation e.g. çarşı Pazar 'market bazaar'	707 milliseconds	50,65
Initial noun e.g. çarşı 'market'	665 milliseconds	60,30

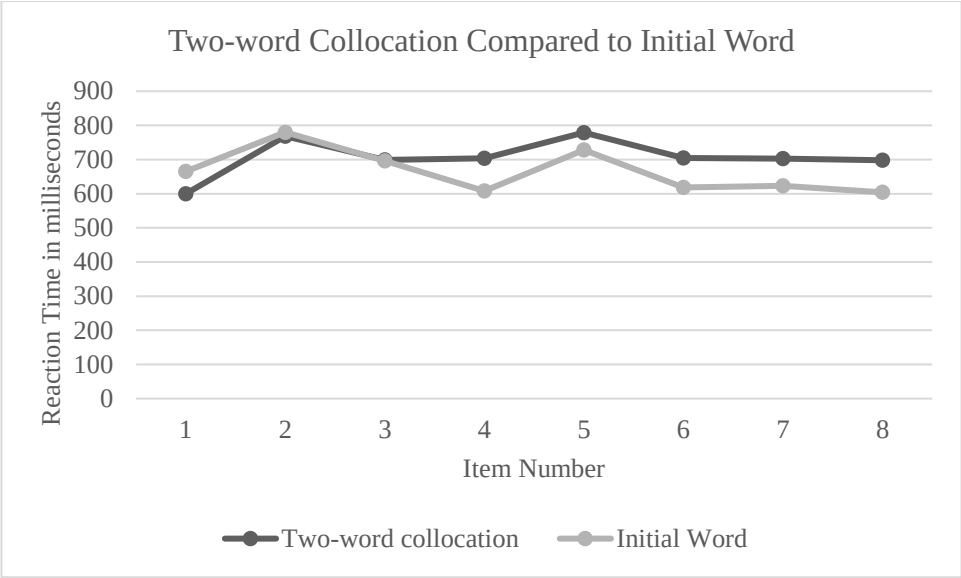


Figure 5. Two-word collocations vs their initial words

4.3 Discussion

4.3.1 Three-word collocations

It should be noted that the discussion relating to the three-word collocations we claim to have been evoked by their two-word counterparts is speculative and these findings should be enriched with further data. These results were surprising for us because although we expected phrase frequency effects we did not expect any of the extended collocations to have a faster or similar reaction time to 2-word collocations. To our surprise, 3 out of 5 collocations exhibited a faster reaction time than their corresponding two-word counterparts. For instance, the RT for *kuru gürültü* was 823 milliseconds while the RT for *kuru gürültü yapmak*, which is the extended and thus longer form, was 764 milliseconds. This finding is remarkable in that it demonstrates that there is much more to frequency effects than just single word frequency and phrase frequency effects can be observed in phrases of various lengths. This finding is also supported by Experiment 3 where we compare the RTs of 2-word collocations to the RTs of their initial words.

The results in this part require a more in depth look. There are two ways to approach these results. If we only consider phrase frequencies, we see that the phrase frequencies of some of the 3-word items are much lower than the phrase frequencies of their 2-word counterparts. For instance, the phrase frequency of *hızlı karar* is 1.02 whereas the phrase frequency of *hızlı karar vermek* is just 0.33 per million. This might mean that a longer phrase needs a lower frequency for it to exhibit phrase level frequency effects and similarly a shorter phrase will require a higher frequency to exhibit phrase frequency effects. By a longer phrase, we mean a phrase that has more words in it. In other words, it might be claimed that there is an inverse relationship between the number of words in a phrase and the required number of occurrences in a corpus in order to obtain frequency effects. This is an idea that definitely needs further analysis and tests.

Another way to approach this data is to look at the structures of the phrases. For instance, when we compare *hızlı karar* and *hızlı karar vermek*, are we really comparing the two or are we comparing *hızlı karar vermek* with *hızlı* (adverb) + (*karar vermek*) because *karar vermek* is also an acceptable and common collocation in Turkish. If we look at the items this way, the list turns into:

- *Hızlı + karar vermek*
- *Kuru + gürültü yapmak*
- *Yüksek + gerilim hattı*
- *Ağır + ceza hakimi*
- *Derin + üzüntü duymak*

Such an approach brings a completely new level to this analysis. In order to get a broader picture, we extracted the phrase frequency of the last parts and the second words in each pair. The details are given in Table 15 below.

Table 15. Comparison of Two and Three-word Phrases in Experiment 3.1

Word	Freq	RT	N+V	Freq	Adj + N	Freq	RT	3 Words	Freq	RT
gürültü	28	614	gürültü yapmak	1.34	kuru gürültü	0.48	823	kuru gürültü yapmak	0.02	764
karar	1369	640	karar vermek	235	hızlı karar	1.02	758	hızlı karar vermek	0.33	709
gerilim	50	692	gerilim hattı	0.97	yüksek gerilim	2.07	805	yüksek gerilim hattı	0.39	804
ceza	650	601	ceza hakimi	1.16	ağır ceza	30.3	682	ağır ceza hakimi	0.02	684
üzüntü	39	586	üzüntü duymak	6.82	derin üzüntü	1.17	822	derin üzüntü duymak	0.5	836

One of the trends that can be elicited from this table is that ending a phrase with a verb seems to decrease its response latency. This might be due to the fact that verbs are extremely common and have very high frequencies. Thus, they might provide a processing advantage. This is an idea that needs further exploration. Although we have not carried out an RT experiment solely on words, a frequency search on TS Corpus for the verbs *yapmak, vermek* and *duymak* provides the following statistics.

- Yap-: 6706 per million
- Ver-: 3986 per million
- Duy-: 441 per million

This is very interesting because it seems that the RT difference between the Adjective + Noun type of collocations and their extended counterparts is shaped by the per-million frequency of these verbs. For instance, normally one might expect the response latency to be shorter for *kuru gürültü* in comparison to *kuru gürültü yapmak*. Yet the RT for *kuru gürültü yapmak* is 59 milliseconds lower than *kuru gürültü*. The RT for *gürültü* is 614 milliseconds, the RT for *kuru gürültü* is 823 milliseconds and the RT for *kuru gürültü yapmak* is 764 milliseconds. This might be due to the fact that *yapmak* 'to do' is an extremely frequent verb in Turkish and thus it might have a facilitatory effect on the processing of *kuru gürültü yapmak*. It might also mean that the brain does not read the whole word sequence with full attention and skips from the beginning of the phrase to the end. Such kind of a paradigm has been observed in eye tracking studies (Siyanova-Chanturia, Conklin, and van Heuven, 2011). This means that the participant knows that the phrase he or she is processing is a holistic chunk and thus this provides a processing advantage. If the phrase were not stored in the mental lexicon as a unit then it would not be processed so fast.

Some might argue that the reaction time of a phrase is tied to the sum of the frequencies of its words. If this were the case, then we would not be able to support phrase frequency effects. In order to test this idea, we analyzed seven phrases from the experiment. We analyzed frequency counts for all of their words and then summed up single word frequencies for each phrase to get a total frequency number for each phrase. Then we compared these numbers to the RT values for each phrase. The details are given in Table 16 below. If the sum of single word frequencies were the determining factor in the RTs of these phrases, then *yüksek gerilim hattı* and *kuru gürültü yapmak* should have the lowest RT values as these phrases would take more

time to respond to under such an assumption. Yet, the RT values of these two phrases are even higher than *doğal yaşam alanı* and *yüksek lisans öğrencisi* whose sum of word frequencies are 1171 and 1205 respectively. This really makes a strong case in showing that although single word frequency plays a role in the processing of multi-word phrases, it is not the main factor. There are frequency effects beyond single word level and these frequency effects interact with each other at various levels.

Table 16. Three-word Collocations and Word Frequencies

Phrase	1st word	2nd word	3rd word	Sum of word frequencies	RT
hızlı karar vermek	185	1369	3986	5540	709
derin nefes almak	145	65	5005	5215	733
doğal yaşam alanı	162	326	683	1171	744
yüksek lisans öğrencisi	602	95	508	1205	756
kuru gürültü yapmak	86	28	6706	6820	764
yüksek gerilim hattı	602	50	171	823	804
derin üzüntü duymak	145	39	441	625	836

The results showed that the sum of individual word frequencies for the given phrases did not correlate with their reaction time values. However, we decided to explore a different approach. Instead of summing up individual word frequencies in a phrase, we multiplied the frequencies of the words in a given phrase. The results are presented in Table 17. The analysis of the data provides an unexpected correlation. In this case, there is a perfect correlation between the multiplication of word frequencies of a phrase and its RT. The higher the end number for multiplication is, the lower the RT is.

Table 17. Three-word Collocations and Word Multiplication

Phrase	1st word	2nd word	3rd word	W1*W2*W3	Sum of Word Frequencies	RT
hızlı karar vermek	185	1369	3986	1009514290	5540	709
derin nefes almak	145	65	5005	47172125	5215	733
doğal yaşam alanı	162	326	683	36070596	1171	744
yüksek lisans öğrencisi	602	95	508	29052520	1205	756
kuru gürültü yapmak	86	28	6706	16148048	6820	764
yüksek gerilim hattı	602	50	171	5147100	823	804
derin üzüntü duymak	145	39	441	2493855	625	836

4.3.2 Collocations with *ve* their reversed forms

In Experiment 3, we tested seven pairs of three-word collocations all involving "*ve*" in the middle. "*ve*" means "and" in Turkish and the collocations that we tested were pairs such as "*gelin ve damat*" which translates into English as "bride and groom". In order to test whether collocations are stored as a whole and accessed holistically we compared collocations with their reversed forms. For instance, we compared "*gelin ve damat*" against "*damat ve gelin*". Both phrases are acceptable in Turkish. However, the first one is more common and thus we expected a processing advantage and proof of holistic storage in the mental lexicon.

Given the fact that both the normal and reversed forms include the same words and thus have the same word frequencies and given the fact that both the typical and the reversed forms are acceptable in Turkish one should not expect to find any difference in reaction times to these collocations if human brain only takes statistics on single words but not multi-word units. However, if human brain takes statistics on word order and frequency of co-occurrence, i.e. if the human brain takes

multi-word statistics, then there should be significant RT differences between the typical forms and reversed forms of the collocations in this task.

The results demonstrate that human brain takes multi-word statistics and multi-word phrases are stored in the mental lexicon. The RT for "*ağız ve diş*" was 607 milliseconds whereas the RT for its reversed form "*diş ve ağız*" was 710 milliseconds. Similarly, the RT for "*gelin ve damat*" stood at 661 milliseconds while the RT for "*damat ve gelin*" stayed at 712 milliseconds. These results indicate that the representation of words in the mental lexicon does not only depend on the frequency of their isolated forms but also on the combinations and co-occurrences they have with other words. The co-occurrence patterns of words and the frequency with which they co-occur play a key role in their representation, storage and processing in the mental lexicon.

4.3.3 Two-word collocations and their initial words

In the third set of Experiment 3, we compared the reaction times for eight collocations that consisted of two words and the reaction time to their first words to see whether collocations would have lower response latencies compared to just their first word. We also analyzed the frequency of the first word, frequency of the second word and the frequency of the whole collocation in TS Corpus. Let us assume that we have word A and word B. Under normal circumstances, one might expect that the reaction time to a given word A would be definitely much faster than a combination of word A and word B. In simple terms RT (Word A) < RT (Word A & Word B). Such an assumption would hold under the condition that frequency effects only apply at a single word level not at multiword level. However, any evidence showing that language users respond to Word A & Word B faster than Word A would mean that

the combination of Word A & Word B are stored as a single combined unit in the human brain. It would also mean that the brain takes statistics not just at a single word level but also across multiple words.

The result of Experiment 3.3 show that the average reaction time for "*çarşı pazar*" was 600 milliseconds whereas the average reaction time for "*çarşı*" was 600 milliseconds. This is surprising given the fact that "*çarşı pazar*" is much longer in length. "*Çarşı pazar*" is a clear example of the fact that some multi word units are stored holistically and not just as individual words in the mental lexicon. A similar case can be made for "*dolu dizgin*" vs "*dolu*" whose RT difference is almost negligible at 699 milliseconds and 696 milliseconds respectively.

The pairs "*çarşı pazar*" and "*hesap kitap*" resemble each other in many ways. Both pairs consist of words that can be used independently. In both pairs the per million frequency of the second word is much higher than the first word, with 165.47 and 143.6 respectively. Interestingly, both of these collocations can be followed by a verb; and thus, create a three-word phrase. In the case of "*çarşı pazar*", the verb is "*dolaşmak*" and in the case of "*hesap kitap*", the verb is "yapmak". The frequency counts from TS Corpus reveals that the occurrence of "*çarşı pazar dolaşmak*" is 0.07 per million, whereas the occurrence of "*hesap kitap yapmak*" is 0.27 per million. It might be the case that when a subject encounters either "*hesap kitap*" or "*çarşı pazar*" they unconsciously activate the three-word forms. In that case, there is either a competition between the three-word form and the two-word form or the three-word phrase is evoked when the two-word phrase is seen. In both cases, the three-word phrase would probably increase the reaction time for the two-word collocation. In such a case, the RT to *hesap kitap* is expected to increase and the results show that this is the case. However, in order to offer a more comprehensive analysis of and

explanation for the relationship between two-word phrases and their three-word versions further studies need to be conducted. The data available regarding this issue are limited.

In the next chapter, we will provide a general discussion of all of the results and data from our experiments by referring to the existing literature in the field of phrase frequency effects. The contribution of this study to the literature, the insights from our wide stimuli set and the implications of our experimental results for model of lexical processing will be discussed in the next chapter.

CHAPTER 5

GENERAL DISCUSSION

The effects of frequency in language processing have been the subject of various studies. It seems obvious that the frequency with which a word occurs plays a role in how that word is processed and stored during language use. Though the effects of frequency on words in isolation have received a lot of attention, its effects on word combinations consisting of two or more words have not received the same level of attention and inquiry. However, this has started to change in the last decades with studies being conducted on the effects of multi-word frequency.

The effects of frequency on single words have been proven many times. Although it is called single frequency, one might assume that this effect should also apply to multiword phrases because instead of considering a single word as a single word, one can consider it as a single unit. Technically speaking, the orthographic difference we make between a single word and multiword phrase might not exist in the human brain if the parts of a multi-word phrase act as one. A single word can be a single unit. Similarly, a phrase or a multiword combination can also act as a single unit and if a multi-word phrase acts as a single unit, it should demonstrate the effects of frequency just as a single word does, namely it should be stored in the lexicon, and it should have a shorter reaction time as frequent exposure to a word strengthens its representation in the mental lexicon.

Many of the previous studies have shown this to be the case (Alegre and Gordon, 1999; Bod, 2001, Sosa and MacFarlane, 2002). Yet the results of some of those studies have been contested (Conklin and Schmitt, 2012). Moreover, many of the previous studies have been carried out on European languages. Thus, it is

definitely useful to test multiword frequency effects on a language which does not belong to the family of Indo-European languages. In this case, our language of choice was Turkish. Due to its agglutinative nature, it is different from many European languages except Finnish. Thus, Turkish proves to be a good choice in that it allows the exploration of frequency effects in a different linguistic environment.

Our goal in this study was to explore the effects of multiword frequency at various levels and to compare it to single word frequency effects to gain a better and more thorough understanding of the relationship between these concepts. We conducted three experiments and we tested four types of multiword phrases. In Experiment 1, we tested Adjective + Noun type of combinations. In Experiment 2, we tested the single word frequencies of the nouns tested in Experiment 1 to see if there was an overlap or conflict between single and multiword frequency effects. In Experiment 3, we tested three different types of multiword phrases. We first tested extended versions of some of the items in Experiment 1 by adding one extra word to those items. We, then, tested collocations formed by the use of "*ve*" such as *gelin ve damat,* 'bride and groom'. Finally, we tested some common two-word collocations against their first words such as comparing *çanak çömlek* to *çanak*. It is interesting to note that none of the previous studies surveyed included Adjective + Noun type of collocations. Thus, in addition to testing some of the typical collocation or multi-word pairs, we also tested a comparatively novel phrase type and obtained useful insights in our study.

All of these different types of collocations and comparisons with the single word counterparts allowed us to obtain a much broader perspective with regards to single word and multi-word frequency effects. By comparing single word frequency effects to multiword frequency effects as well as analyzing their interaction, we were

able to show that the results obtained were attributable to multiword frequency effects. What is more, to our knowledge, we are the first to use substring RT to confirm phrase frequency effects. The studies in the literature consider phrase and substring frequencies but only test the phrase RT with regard to phrase frequency. However, such an approach ignores the effect of single word RT, which might underlie some of the phrase frequency effects.

The results of Experiment 1 showed that there was a significant difference between the mean reaction times of low frequency adjective + noun pairs and high frequency adjective + noun pairs such as *parlak ayna* and *parlak fikir*. The difference in the RTs were in parallel with the phrase frequency values of the pairs. When comparing pairs that shared the same adjective, the pair with the higher phrase frequency had a lower response latency than the pair with the lower frequency. This was a good starting point in addressing our main question: do multiword frequency effects exist?

The answer seemed to be a clear affirmative, yet there was a confounding factor. One could argue that the difference between the reaction times of *parlak ayna* and *parlak fikir* was not due to the effect of phrase frequency but it was merely a reflection of single word frequency. Although we believed that pairs such *parlak fikir* acted as a unit and thus demonstrated phrase frequency effects, it was necessary to demonstrate that these effects were not due to single word frequency effects. One could argue that *parlak fikir* was responded to faster than *parlak ayna* not because of its phrase frequency because *fikir* is a much more common word than *ayna*. In order to disprove this idea, we needed two things. First, we needed to obtain the frequency and RT values for each noun in the Adjective + Noun pairs in Experiment 1 and secondly we needed to find cases where phrase frequency effects overrode single

word frequency. This means a case in which the Adjective + Noun combination with a higher phrase frequency but low noun frequency scored a lower RT despite starting with the same adjective and having a high frequency noun. In fact, we found four such cases.

For example, when comparing *dik yamaç* and *dik yol*, we noticed that although *yol* has a frequency of 1252.57 per million compared to 9.73 per million for *yamaç*, the RT for *dik yamaç* is shorter at 707 milliseconds compared to 717 milliseconds for *dik yol*. If single word frequency were the overriding or the main factor, determining reaction time, then we would expect the RT for *dik yol* to be shorter than *dik yamaç*. However, one can argue that single word frequency interacts with phrase frequency because the RT difference between *yol* and *yamaç* was 109 milliseconds with *yol* having the shorter RT, yet the RT difference between *dik yamaç* and *dik yol* was just 10 milliseconds. It might be that single word frequency of the word *yol* accelerated the processing of *dik yol* although it has a lower phrase frequency. A similar case is observed in the case of *temiz hava* and *temiz iş* where despite having higher single word frequency, the RT for *temiz iş* is longer than the RT for *temiz hava*, which has a higher phrase frequency.

These results indicate that the RT differences between the low frequency collocation pairs and the high frequency collocation groups were not due to single word frequency but due to phrase frequency. Furthermore, the effect of phrase frequency indicates that phrases are stored holistically in the mental lexicon and the brain takes statistics on multiword units.

Experiment 3.1 compared phrases such as *hızlı karar* and *kuru gürültü* with their extended versions such as *hızlı karar vermek* and *kuru gürültü yapmak*. It is important to note that these extended versions were also common

phrases/collocations in Turkish. The reason why we embarked on such a comparison was to investigate whether we would observe effects of phrase frequency in phrases consisting of three words since most of our experimental items were two-word phrases. Additionally, we suspected that some of the collocation in Experiment 1 evoked longer collocations such as *kuru gürültü* evoking *kuru gürültü yapmak*. We wanted to see if this was the case and to our knowledge, we are the first to explore such an idea.

The results showed that there was not a significant difference between the mean reaction times for two-word phrases and their corresponding three-word counterparts. This was a rather striking result because the addition of a full extra word seemed to have almost no effect on mean RTs. What is more surprising is the fact that the RTs for some three-word phrases were much shorter than their two-word counterparts. For instance, the RT for *hızlı karar* was 758 milliseconds while the RT for *hızlı karar vermek* was 709 milliseconds. Similarly, the RT for *kuru gürültü* was 823 milliseconds while the RT for *kuru gürültü yapmak* was 764 milliseconds. If a three-word phrase has a lower response latency than a two-word phrase despite an additional word and the processing cost it will bring, that means the three-word phrase is acting as a unit and not just as the sum of separate words. This point indicates the existence of frequency effects beyond single word and two-word levels. However, we have to be cautious about what these results suggest as not all items behaved similarly.

In Experiment 3.2, we analyzed the effects and existence of phrase frequency by comparing the reaction times to common collocations formed with the use of "*ve*" and their reversed forms. For instance, *ağız ve diş* was compared to *diş ve ağız*. Both phrases are perfectly acceptable in Turkish and the main difference is their phrase

frequency. If phrase frequency did not exist and the only factor was single word frequency, then the RT for *diş ve ağız* would be the same as *ağız ve diş* since A + B and B + A would be expected to have the same result. Yet, the result showed that there was a significant difference between the mean reaction times of the typical forms of phrases and their reversed forms. For instance, while the RT for *ağız ve diş* was 607 milliseconds, the RT for *diş ve ağız* stood at 710 milliseconds.

In Experiment 3.3, we compared some common two-word collocations to their initial words. For instance, we compared the RT for *hesap kitap* with the RT for *hesap*. Normally one would expect that the RTs for the one-word condition would be much lower than the RTs for the two-word collocation. Yet again, commonsense proved to be wrong. We were not able to find a significant difference between the mean RTs of the single word group and the two-word collocations. In fact, some of the two-word collocation such as *çarşı pazar* and *ipsiz sapsız* had lower response latencies than their one-word counterparts. Such a case cannot be explained by single word frequency effect. This was another interesting finding in favor of phrase frequency effects.

These findings might be better described as unit frequency effect rather than phrase frequency effects. What we mean by unit frequency is a combination of word frequency and phrase frequency. Instead of using two separate constructs to explain frequency effects at various word lengths we might adopt a unitary term that encompasses all frequency effects as long as the given sequence of letters act as a unit. One of the benefits of using the term unit frequency effect as opposed to single word frequency or phrase frequency is that it offers a concept that can capture frequency effects at various word levels. It is also a more compatible term for a continuous model of frequency as opposed to a threshold model.

Taken together, these results provide evidence that frequency effects go well beyond single word level and can be observed in two or three-word phrases. This has serious implications for the models of mental lexicon and language processing. First, it clearly shows that human brain pays attention to frequency and takes statistics at various levels constantly and unconsciously. Otherwise, we would not expect to see such differences and effects regarding single words and phrases. Secondly, it seems that human brain stores frequent phrases, collocations or constructs holistically in the memory. If this were not the case, we would not be able to observe phrase frequency effects because there would not be any processing advantage for frequent phrases if they were not stored in the memory.

These findings are clearly pertinent to the debate on the nature of the mental lexicon. What is and is not stored in the mental lexicon is one of the main questions of lexical models. Word and Rules model (Pinker & Ullman, 2002) argues that words and rules are stored separately in the mental lexicon. Under such a model, the word "flowers" would be stored separately as the word "flower" and plural rule and during production human brain would apply the plural rule to the word "flower" to turn it into "flowers". However, such an argument is not compatible with our findings. Our results show that frequency of co-occurrence leads to a processing advantage and this is achieved by storing multiword phrases as a unit or chunk in the memory. That means if the word "flowers" is much more common than the word "flower", that means there is a high probability that it will be stored holistically, not as a word + rule.

There are many studies in the literature that demonstrate the effects and existence of phrase frequency as a valid construct. Alegre and Gordon (1999) show that regularly inflected word forms such as "walked" might behave like irregular

forms and thus this might be an indicator that they are stored as a whole or a chunk instead of just the base form. Bod (2001) analyzed the responses to three word SVO type sentences such as “I like it” and “I keep it” by matching all items for lexical frequency as well as surface structure and complexity. Thus, the results could be attributed to phrase frequency. The results showed that participants responded faster to high frequency sentences. Sosa and MacFarlane (2002) used a word-monitoring task to measure reaction times to the English function word “of” in collocations with different levels of frequency. Reaction times were faster for the low frequency group whereas RT values were slower in the high frequency group. The results indicated that that higher frequency words were stored as units and thus were harder to decompose.

Similarly, Bannard and Matthews (2008) show that children store multi-word phrases. Their study employed a production task involving whole-form frequency manipulation with phrases. Their items such as *a drink of milk* and *a drink of tea* were matched on all frequency measures except phrase frequency. Their results showed that children were comparatively faster at repeating high frequency phrases. In a similar study, Tremblay, Derwing, Libben, and Westbury (2011) analyzed the reaction times to sentences that involved lexical bundles such as “don’t worry”. The results showed that sentences that included lexical bundles had faster reaction times compared to control sentences. In another study, Siyanova-Chanturia, Conklin, and van Heuven (2011) utilized eye tracking to explore the processing of formulaic sequences by both native and non-native speakers of English using sentences that differed only in the frequency of the embedded sequences. Sequences such as “bride and groom” were used both in normal and reversed form and typical forms had significantly lower response latencies compared to reversed forms. In a related study,

Tremblay, Derwing, Libben, and Westbury (2011) carried out a self-paced reading experiment using items such as '*in the middle of the*' and '*in the front of the*'. This made it different from many other studies that used items which only differed in their final word. The result of this study showed that frequency of phrases was a strong predictor of reading speed.

The results of these studies indicate holistic processing and storage in high frequency collocations. Moreover, in many cases, processing advantages for high frequency phrases are observed. If phrase level frequency effects did not exist, then such results would not be obtained. The results from all of these experiments along with the results of our study provide strong evidence in support of multi-word or phrase frequency effects. These results are not in line with the Words and Rules approach (Pinker, 1999) which argues that words are stored in isolation.

Our findings lend support to Emergentist models of language and lexicon in which language use shapes language representation (Bybee, 1998; Goldberg, 1995, 2006; Christiansen & Chater, 1999; Elman 1990). In such models, frequency and exposure play a significant role in the representation of words and language in the human brain. Thus, we believe that Emergentist models represent the frequency effects of single words and phrases more accurately and our study provides evidence that phrase frequency effects play a significant role in language comprehension. The results also indicate that mental lexicon is not fixed but rather dynamic. The fact that frequency affects processing requires frequency information to be constantly monitored and updated in the mental lexicon and that requires a dynamic structure. Thus, as the Emergentist models put forward, exposure to language shapes the representation of language.

Another important point is that models of language processing and mental lexicon need to integrate phrase frequency effects because as shown by our study and various other studies, there is more to frequency effects than just single word frequency. Experimental data show that phrase frequency is a significant construct in language processing and thus, it should be integrated into existing models. An additional point to consider is whether frequency effects are continuous or whether they follow certain thresholds. Our data showed frequency effects in phrases with various degrees of frequency; however, our item set did not include a systematic list of continuous frequency data for us to analyze. Thus, we hope to enrich this study with future studies with a larger and more standardized set of phrases.

Furthermore, Emergentist models (Goldberg, 1995, 2006; Christiansen & Chater, 1999; Elman 1990) assume a continuous, not a binary, approach when it comes to frequency effects. Thus, our suggestion to use the term unit frequency instead of single word or phrase frequency provides a more unifying approach. It is becoming increasingly clear that whether we notice it or not, frequency effects play a key role in our processing of language not just in English but also in Turkish and many other languages. We should also analyze the implications of these data on lexical representation and linguistic models. Language users are obviously sensitive to distributional data. According to Word and Rules models, frequency effects should only be observed in words not in longer phrases (Pinker & Ullman, 2002). Thus, Word and Rules approach is not in line with the findings that these studies report.

In conclusion, the results of our experiments indicate that frequency effects are not limited to single words. They can be observed in multiword phrases and phrase frequency effects can either complement or inhibit single word frequency

effects or vice versa. The findings from our study indicate that there is a high probability that frequent phrases are stored as a single unit or chunk and this provides frequent phrases with a processing advantage. Moreover, this makes sense from a computational perspective as it might be much easier for the human brain to store and retrieve frequent multiword phrases when needed than combining individual words in each case to form these phrases online. This reduces the demand on cognitive capacity and leads to quick and automatic processing. Such an approach is in line with the premises of Emergentist models, which place a heavy emphasis on the role of experience and exposure to language on the representation of language. The results also have important implications for the models of mental lexicon. Considering the results of this study and many other previous studies, it seems necessary for models of lexicon to integrate phrase frequency effects as one of the main factors in lexical storage and production.

In the final chapter, we will provide future research directions that can complement and further the data and insights obtained in our study. Various research paths and their respective contributions to the studies on phrase frequency effects and models of lexical processing will be addressed.

CHAPTER 6

FUTURE DIRECTIONS

Our main aim in this study was to demonstrate that frequency effects exist beyond single word level. In order to show this, we compared single words to two-word and three-word phrases using lexical and phrasal decision tasks. The results we obtained provided insights into the nature of phrase frequency effects and contributed to the scientific literature on multi-word frequency effects by providing data on a language such as Turkish. However, we consider our contribution as just the starting point of a promising path on the discovery of the mental representation of language. There are a number of ways in which our study can be extended and complemented in order to arrive at a more comprehensive view of frequency effects.

First, our study relied on corpus data to obtain frequency counts. The main issue with relying on corpus data with frequency counts is that corpus data cannot be fully representative of a language. Thus, there is always the possibility that there might be some sort of discrepancy between corpus data and real life language use as corpus data contains samples from a language. TS Corpus, the corpus we used in this study, is heavily based on written materials collected from online sources. Thus, it is more biased towards written language. Paetzold and Specia (2016) demonstrate that frequency norms they extract from the SubIMDB corpus they compiled by using English movie subtitles are more effective in predicting variables such as lexical decision times, simplicity and familiarity compared to more established corpora such as Kucera-Francis, HAL and SUBTLEXus. In a related study, Herdağdelen and Marelli (2016) show that corpora compiled from social media sites such as Facebook and Twitter are better predictors of lexical decision reaction times for both British

and American English. They argue that this superior performance is not simply due to the size of the corpora but due to better representation of language in various categories. Yet, many previous studies in this field used written corpora to calculate word frequencies and thus our approach is line with previous studies.

Moreover, it might be interesting to analyze the relationship between transitional probabilities and phrase frequency effects. This study focused on the effects of phrase frequency; however, some of these effects could also be explained by transitional probabilities. Yet as Janssen and Barber (2012) argue, our results do not show a transitional probability effect that is independent of the phrase frequency effects observed. In many cases, transitional probabilities and phrase frequency effects seem to work hand in hand. Moreover, Janssen and Barber (2012) state that the frequency of a phrase and the transitional probability between individual words do not always correlate especially for phrases that are longer than two words.

An additional point of inquiry could be the use of additional experimental methods and setups to validate and delve into phrase frequency effects. The use of eye tracking experiments can provide many additional insights by providing data in a natural environment and in context as their unobtrusive nature allow for a more natural data collection. The use of eye tracking can also lead to exploration of frequency effects on much longer phrases in context. Moreover, use of an auditory task would also bring an additional dimension to the analysis of this issue. Our experimental items can be tested in an auditory phrasal decision task. The use of auditory tasks can both help confirm the results of visual tasks and explore any differences between auditory and visual medium on phrase frequency effects. A more complex but definitely worthwhile step would be to examine phrase frequency effects through EEG (Electroencephalography) or fMRI (Functional magnetic

resonance imaging) tests to explore the effects of frequency on language comprehension and production in the human brain. Such approaches can help researchers examine the neural correlates of phrase frequency effects. For instance, if frequently used collocations light up the areas in brain that are mostly associated with storage rather than the application of grammatical rules, then it would provide strong evidence in support of storage of frequent phrases in the mental lexicon. Moreover, the use of brain imaging techniques might also uncover effects of correlations that are not readily observable in a behavioral psycholinguistic experiment.

Another important direction concerns the further exploration of single-word frequency and phrase frequency. Data from some of our experimental items indicated that single word frequency could either boost or inhibit phrase frequency. Although it was a point worth considering, we did not pursue this point thoroughly as it was not the main question of our study. Thus, a more in-depth study on the interaction between single word frequency and phrase frequency can better inform the models of lexical access and storage.

We tested more than 30 pairs of collocations and they provided us with valuable insights. However, we believe it is just the beginning as the Turkish language is rich and complex in its use of phrases and collocations. Thus, there is definitely a need to explore more pairs at different ranges of frequency. Most of our comparisons were binary comparisons between high frequency and low frequency collocations. Thus, it will definitely be useful to analyze words and phrases from various levels of frequency to observe the effects of frequency on a continuum.

Our study can be further improved in a few ways. First, since this is the first study on multi-word frequency effects in Turkish, it would be a great idea to replicate these experiments with a much wider participant pool to see if any considerable changes take place. Secondly, although the number of left-handed participants was comparatively very small, in a future study their data should be treated separately to see whether frequency effects behave differently in the case of left-handed speakers. Thirdly, it would be very beneficial to test a much longer and broader phrase set to observe frequency effects on word sequences longer than three words and further explore the nature of interaction between single word frequency, substring frequencies and whole-phrase frequency and to see how substring frequency can evoke or activate a longer phrase frequency. Finally, we were not fully able to explain why some two-word collocations had lower response latencies compared to their initial word, while many others did not exhibit such behavior.

Finally, whether frequency effects operate on a continuum or a threshold basis can be tested more rigorously. Our results with phrases from various frequency ranges supported a continuous approach, as there was a facilitatory effect of phrase frequency over different frequency ranges. However, increasing the number of phrases and the frequency ranges tested can offer more definitive data on whether frequency effects operate on a threshold or continuum model. This can also provide more extensive data on the interaction between response latencies and phrase frequency. While testing phrase frequency effects it is also essential to consider the interaction of phrase frequency with word and substring frequency as well as the effect of orthography and syllable count on frequency effects. As mentioned earlier, employing unit frequency measures as opposed to word or phrase frequency might offer a more holistic approach to frequency effects in general.

Our study has been a productive and rewarding journey for us to uncover and reaffirm some of the key insights on the nature of the models of lexical storage, access and language processing in general as well as processing advantages and response latencies. We are grateful for the work of researchers in the field of frequency effects whose insights and guidance helped us to deal with such complex issues and we hope that our study will provide practical and useful data and insights for future studies into the substance of one of the most unique human traits: the gift of language.

APPENDIX A

ITEMS IN EXPERIMENT 1

Items in Experiment 1

1 ağır ceza: heavy penalty

1 ağır yara: deep wound

2 aşırı hız: excessive speed

2 aşırı kilo: excessive weight

3 canlı müzik: live music

3 canlı bitki: live plant

4 derin nefes: deep breath

4 derin üzüntü: deep sorrow

5 dik yamaç: steep slope

5 dik yol: steep road

6 doğal yaşam: natural life

6 doğal ekmek: natural bread

7 düz çizgi: straight line

7 düz taban: flat feet

8 hızlı tren: high speed train

8 hızlı karar: quick decision

9 kötü niyet: bad intention

9 kötü görüş: bad view

10 kuru gıda: dry food

10 kuru gürültü: much ado about nothing

11 mutlu son: happy end

11 mutlu çift: happy couple

12 parlak fikir: bright idea

12 parlak ayna: bright mirror

13 sert söz: harsh word

13 sert plastik: rigid plastic

14 temiz hava: clean air

14 temiz iş: clean job

15 yüksek lisans: Master’s degree

15 yüksek gerilim: high voltage

Fillers in Experiment 1

ağır yohakiyu	doğal soyo	mutlu doroba
ağır memumigu	doğal zahake	mutlu fuman
aşırı remipa	düz yuzonize	parlak nibonmu
aşırı teme	düz beringu	parlak yupihonon
canlı yomozon	hızlı bitarumi	sert tadaren
canlı topyunso	hızlı zonsoyo	sert metanbenya
derin kosimun	kötü nekiyon	temiz nuzanbe
derin kazeba	kötü keyan	temiz ade
dik mikanre	kuru koze	yüksek mekenpeo
dik poke	kuru bonyoni	yüksek onoko

APPENDIX B

ITEMS IN EXPERIMENT 2

Items in Experiment 2

1. Ceza: punishment/penalty
2. Yara: wound
3. Hız: speed
4. Kilo: kilo
5. Müzik: music
6. Bitki: plant
7. Nefes: breath
8. Üzüntü: sorrow
9. Yamaç: slope
10. Yol: road
11. Yaşam: life
12. Ekmek: bread
13. Çizgi: line
14. Taban: base/floor
15. Tren: train
16. Karar: decision
17. Niyet: intention
18. Görüş: view
19. Gıda: food
20. Gürültü: noise
21. Son: end
22. Çift: couple
23. Fikir: idea
24. Ayna: mirror
25. Söz: word
26. Plastik: plastic
27. Hava: air
28. İş: job
29. Lisans: license/degree
30. Gerilim: tension/voltage

Fillers in Experiment 2

ade	fugade	kazeba
agome	fuka	keranpi
beke	gehes	keyan
bitaru	gimozen	kodah
biyufu	gopo	kotonmo
buman	goranin	koze
dao	haba	kuge
doanu	hibi	makununi
echin	ichi	meken
epiho	kapodo	memu

APPENDIX C

ITEMS IN EXPERIMENT 3

Test Items for Group 1

1. ağır ceza hakimi: high criminal court judge
2. ağız ve diş: mouth and tooth
3. çanak çömlek: pottery
4. çarşı pazar: bazaar market
5. çevre ve şehircilik: environment and urban planning
6. ceza ve suç: punishment and crime
7. doğal yaşam alanı: natural habitat
8. dolu dizgin: full speed
9. gelin ve damat: bride and groom
10. güzel ve çirkin: beautiful and ugly
11. hal: manner/attitude/mood
12. hesap: calculation
13. hızlı karar vermek: to make a quick decision
14. iğne: needle
15. ipsiz sapsız: footloose
16. iş: job
17. köpek ve kedi: dog and cat
18. otoyol ve köprü: higway and bridge
19. yüksek gerilim hattı: high tension line

Test Items for Group 2

1. çanak: bowl
2. çarşı: bazaar
3. çirkin ve güzel: ugly and beautfiul
4. damat ve gelin: groom and bride
5. derin nefes almak: to take a deep breath
6. derin üzüntü duymak: to feel deep sorrow
7. diş ve ağız: tooth and mouth
8. dolu: full
9. hal hatır: state
10. hesap kitap: financial arrangement
11. iğne iplik: needle yarn
12. iş güç: job power
13. kedi ve köpek: cat and dog
14. köprü ve otoyol: bridge and highway
15. kuru gürültü yapmak: to make a lot of nonsense noise
16. şehircilik ve çevre: urban planning and environment
17. suç ve ceza: crime and punishment
18. yüksek lisans öğrencisi: Master’s student
19. ipsiz sapsız: footlose

Fillers in Experiment 3

ade kebeu

agome keguyo

azerun keyan

beke kigenpa

biyufu kosimun

buman kozemera

bumon kuge

doru mekenpeo

echin memumigu

epiho meronan

fuka mikanre

fuman mubineno

fuyupya mugon

hahimo nisazo

hasuga norun

hozemi nuryu

ichi nusho

kapodo nuzanbe

kazeba obebu

REFERENCES

Abbot-Smith, K., & Tomasello, M. (2006). Exemplar-learning and schematization in a usage based account of syntactic acquisition. *The Linguistic Review*, *23*(3), 275–290.

Aitchison, J. (2012). *Words in the mind*. Hoboken, NJ: John Wiley & Sons.

Alegre, M., & Gordon, P. (1999). Frequency effects and the representational status of regular inflections. *Journal of Memory and Language, 40*(1), 41–61.

Arnon, I., & Snider, N. (2010). More than words: Frequency effects for multi-word phrases. *Journal of Memory and Language*, *62*(1), 67–82.

Arnon, I., & Cohen Priva, U. (2013). More than words: The effect of multi-word frequency and constituency on phonetic duration. *Language and Speech*, *56*(3), 349-371.

Baayen, R. H., Dijkstra, T., & Schreuder, R. (1997). Singulars and plurals in Dutch: Evidence for a parallel dual-route model. *Journal of Memory and Language*, *37*(1), 94–117.

Baayen, H., Schreuder, R., deJong, N., & Krott, A. (2002). Dutch inflection: The rules that prove the exception. In S. Nooteboom, F. Weerman, & F. Wijnen (Eds.), *Storage and computation in the language faculty* (pp. 61–92). Boston: Kluwer.

Bannard, C., & Matthews, D. (2008). Stored word sequences in language learning: The effect of familiarity on children's repetition of four-word combinations. *Psychological Science, 19* (3), 241–248.

Behavioural Science In The Cloud. (2018). *Gorilla*. Retrieved 1 December 2017, from https://gorilla.sc/

Bell, A., Jurafsky, D., Fosler-Lussier, E., Girand, C., Gregory, M., & Gildea, D. (2003). Effects of disfluencies, predictability, and utterance position on word form variation in English conversation. *Journal of the Acoustical Society of America*, *113*(2), 1001–1024.

Biber, D. (1999). Lexical bundles in conversation and academic prose. *Language & Computers, 26(1)*, 181–190.

Bilgin, O. (2016). *Frequency effects in the processing of morphologically complex Turkish words.* (Unpublished master's thesis). Boğaziçi University, Istanbul, Turkey.

Bod, R. (2001). Sentence memory: Storage vs. computation of frequent sentences. Paper presented at CUNY Conference on Human Sentence Processing, University of Pennsylvania, Philadelphia, PA.

Bod, R. (2006). Exemplar-based syntax: How to get productivity from exemplars. *Linguistic Review*, *23*(3), 291–320.

Brysbaert, M., Mandera, P., & Keuleers, E. (2017). The word frequency effect in word processing: An updated review. *Current Directions In Psychological Science*, *27*(1), 45-50. http://dx.doi.org/10.1177/0963721417727521

Bybee, J. (1998). *The emergent lexicon* (Vol. 34, pp. 421–435). Chicago: Chicago Linguistic Society.

Bybee, J. (2006). From usage to grammar: The mind's response to repetition *Language*, *82*(4), 711–733

Bybee, J., & Scheibman, J. (1999). The effect of usage on degrees of constituency: The reduction of don't in English. *Linguistics*, *37*(4), 575–596.

Christiansen, M. H. (1994). *Infinite languages, finite minds: Connectionism, learning, and linguistic structure*. Unpublished doctoral dissertation, University of Edinburgh.

Christiansen, M. H., & Chater, N. (1999). Toward a connectionist model of recursion in human linguistic performance. *Cognitive Science*, *23*(2), 157–205.

Conklin, K., & Schmitt, N. (2012). The processing of formulaic language. *Annual Review of Applied Linguistics*, *32*, 45-61.

Diessel, H. (2007). Frequency effects in language acquisition, language use, and diachronic change. *New Ideas in Psychology*, *25*(2), 104–123.

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

Ellis, N. (2002). Frequency effects in language processing. *Studies in Second Language Acquisition*, *24*(2), 143–188.

Elman, J. L. (1990). Finding structure in time. *Cognitive Science*, *14*(2), 179–211.

Erman, B., & Warren, B. (2000). The idiom principle and the open choice principle. *Text - Interdisciplinary Journal for the Study of Discourse*, *20*(1), 36-37. http://dx.doi.org/10.1515/text.1.2000.20.1.29

Evert, S. (2018). *The IMS Open Corpus Workbench*. *Cwb.sourceforge.net*. Retrieved 21 March 2018, from http://cwb.sourceforge.net/

Frazier, L., & Fodor, J. D. (1978). The sausage machine: A new two-stage parsing model. *Cognition, 6*(4), 291–325.

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

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

Gregory, M., Raymond, W. D., Bell, A., Fosler-Lussier, E., & Jurafsky, D. (1999). The effects of collocational strength and contextual predictability in lexical production (Vol. 35, pp. 151–166). Chicago Linguistic Society.

Gürel, A. (1999). Decomposition: To what extent? The case of Turkish. *Brain and Language*, 68(1-2), pp.218-224.

Herdağdelen, A., & Marelli, M. (2016). Social media and language processing: How Facebook and Twitter provide the best frequency estimates for studying word recognition. *Cognitive Science, 41*(4), 976-995. http://dx.doi.org/10.1111/cogs.12392

Howarth, P. (1998). The phraseology of learners' academic writing. In A. Cowie (Ed.), *Phraseology: Theory, analysis, and applications* (pp. 161–186). Oxford, UK: Oxford University Press.

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

Jurafsky, D., Bell, A., Gregory, M., & Raymond, W. (2001). Probabilistic relations between words: Evidence from reduction in lexical production. In J. Bybee & P. Hopper (Eds.*), Frequency and the emergence of linguistic structure* (pp. 229–254). Amsterdam: John Benjamins.

Malkiel, Y. (1959). Studies in irreversible binomials. *Lingua, 8*, 113-160. http://dx.doi.org/10.1016/0024-3841(59)90018-x

McClelland, J. L., & Kawamoto, A. H. (1986). Mechanisms of sentence processing: Assigning roles to constituents of sentences. In J. L. McClelland, D. E. Rumelhart, & the PDP Research Group (Eds.), *Parallel distributed processing: Explorations in the microstructure of cognition*. Volume 2: Psychological and biological models (pp. 272–325). Cambridge, MA: MIT Press.

Monsell, S. (1991). The nature and locus of word frequency effects in reading. In D. Besner & G. W. Humphreys (Eds.), *Basic processes in reading: Visual word recognition* (pp. 148–197). Hillsdale, NJ: Erlbaum.

Monsell, S., Doyle, M., & Haggard, P. (1989). Effects of frequency on visual word recognition tasks: Where are they? *Journal of Experimental Psychology: General, 118*(1), 43-71. http://dx.doi.org/10.1037//0096-3445.118.1.43

Noelle, D. C., & Cottrell, G. W. (1995). A connectionist model of instruction following. In Proceedings of the 17th annual conference of the Cognitive Science Society (pp. 369–374). Hillsdale, NJ: Lawrence Erlbaum Associates.

Paetzold, G., & Specia, L. (2016). Collecting and exploring everyday language for predicting psycholinguistic properties of words. In *Proceedings of the 26th COLING*, pp. 1669–1679.

Pinker, S., & Ullman, M. (2002). The past and future of the past tense. *Trends in Cognitive Sciences*, *6*(11), 456–463.

Pinker, S. (1991). Rules of language. *Science, 253*(5019), 530–535.

Pinker, S. (1999). *Words and rules: The ingredients of language*. Oxford: Basic Books.

Rayner, K., & Duffy, S. A. (1986). Lexical complexity and fixation times in reading: Effects of word frequency, verb complexity, and lexical ambiguity. *Memory & Cognition*, *14*(3), 191–201.

Rosner, B. (1975). On the Detection of Many Outliers. *Technometrics, 17*(2), 221. http://dx.doi.org/10.2307/1268354

Sak, H., Güngör, T., & Saraçlar, M. (2010). Resources for Turkish morphological processing. *Language Resources and Evaluation, 45*(2), 249-261. http://dx.doi.org/10.1007/s10579-010-9128-6

Sezer, T. (2017). TS Corpus Project: An online Turkish Dictionary and TS DIY Corpus. *European Journal of Language and Literature, 9*(1), 18-24.

Shaoul, C. & Westbury, C. (2011). Formulaic sequences: Do they exist and do they matter? *Methodological and Analytic Frontiers in Lexical Research (Part II). Special Issue of The Mental Lexicon, 6*(1), 171–196.

Siyanova-Chanturia, A., Conklin, K., & van Heuven, W. (2011). Seeing a phrase "time and again" matters: The role of phrasal frequency in the processing of multiword sequences. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 37*(3), 776-784. doi: 10.1037/a0022531

Sorhus, H. (1977). To hear ourselves—Implications for teaching English as a second language. *English Language Teaching Journal, 31*(3), 211–221.

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

Stemberger, J. P., & MacWhinney, B. (1988). Are inflected forms stored in the lexicon? In M. Hammond & M. Noonan (Eds.), *Theoretical morphology: Approaches in modern linguistics* (pp. 101–116). San Diego, CA: Academic Press.

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

Tremblay, A., Derwing, B., Libben, G., & Westbury, C. (2011). Processing advantages of lexical bundles: Evidence from self-paced reading and sentence recall tasks. *Language Learning*, *61*(2), 569–613.

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: Acquisition, processing, and use* (pp. 153–172). Amsterdam: John Benjamins.

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