FREQUENCY EFFECTS IN THE PROCESSING OF MORPHOLOGICALLY COMPLEX TURKISH WORDS

ORHAN BİLGİN

BOĞAZİÇİ UNIVERSITY

2016

FREQUENCY EFFECTS IN THE PROCESSING OF
MORPHOLOGICALLY COMPLEX TURKISH WORDS

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
Orhan Bilgin

Boğaziçi University
2016

Frequency Effects in the Processing of

Morphologically Complex Turkish Words

The thesis of Orhan Bilgin

has been approved by:

Assoc. Prof. Mine Nakipoğlu ____________________

(Thesis Advisor)

Assist. Prof. Arzucan Özgür ____________________

(Thesis Co-Advisor)

Prof. Aslı Göksel ____________________

Assist. Prof. Albert Ali Salah ____________________

Assist. Prof. Serkan Şener ____________________

(External Member)

September 2016

DECLARATION OF ORIGINALITY

I, Orhan Bilgin, 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 ....19 OCTOBER 2016....................

ABSTRACT

Frequency Effects in the Processing of Morphologically Complex Turkish Words

This is an empirical study that examines how morphologically simple and complex words in Turkish are represented in the brains of native speakers. Two experiments are reported that use various “frequency of occurrence” metrics as independent variables. The secondary findings of the study are that (a) frequency is an extremely complex concept, especially in the case of an agglutinating language like Turkish; (b) different frequency measures are highly correlated; (c) frequency distributions are uneven at several levels; (d) the overwhelming majority of grammatically possible forms are never used even in a large corpus; (e) in an agglutinating language like Turkish, morphology has a deep impact even at sub-lexical levels such as the distribution of letter-ngrams; (f) conducting psycholinguistics experiments online rather than in a laboratory environment is a feasible option; (g) letter shape does not have an effect on word recognition accuracy; (h) morphologically complex Turkish words are processed two times more slowly than simple words, suggesting that suffix sequences add a significant workload to the recognition process. The three main findings of the experiments, on the other hand, are that (a) more frequent simple words are processed faster than less frequent simple words, thus replicating a well-established finding in a typologically different language; (b) complex words are probably processed from left to right, and, most importantly, (c) the human brain can use suffix sequences to recognize complex words, thus suggesting that there exist mental representations for frequently occurring suffix sequences, probably in addition to mental representations for individual suffixes.

ÖZET

Biçimbilimsel Bakımdan Karmaşık Türkçe Kelimelerin İşlenmesinde

Frekans Etkileri

Bu çalışmada, biçimbilimsel açıdan basit ve karmaşık Türkçe kelimelerin, anadili Türkçe olan kişilerin beyninde ne şekilde kayıtlı olduğu, çeşitli sıklık ölçütlerinin bağımsız değişken olarak kullanıldığı iki çevrimiçi deney aracılığıyla incelenmiştir. Çalışmanın ikincil bulguları şunlardır: (a) özellikle Türkçe gibi bitişimli dillerde, "sıklık" son derece karmaşık bir kavramdır; (b) farklı sıklık ölçütleri arasında yüksek korelasyon mevcuttur; (c) sıklık dağılımları her düzeyde dengesizdir; (d) dilbilgisine uygun kelimelerin ezici çoğunluğu büyük bir derlemde dahi hiç kullanılmamaktadır; (e) Türkçe gibi bitişimli bir dilde, harf dizilerinin dağılımı gibi kelime-altı alanlar üzerinde dahi biçimbilimsel yapının yoğun etkileri gözlenmektedir; (f) ruhdilbilimi deneylerinin laboratuar ortamı yerine çevrimiçi ortamda uygulanması başarılı sonuçlar vermektedir; (g) harflerin büyük veya küçük olmasının, kelime tanımadaki başarı oranı üzerinde etkisi yoktur; (h) biçimbilimsel açıdan karmaşık Türkçe kelimeler, basit kelimelere kıyasla iki kat yavaş işlenmektedir, bu ise kelime sonlarındaki ek dizilerinin kelime işleme sürecine büyük bir yük getirdiğini göstermektedir. Çalışmanın üç temel bulgusu ise şunlardır: (a) çok çeşitli diller için uzun yıllardır bilindiği üzere, daha sık kullanılan basit kelimeler, daha az kullanılan basit kelimelerden daha hızlı işlenmektedir; (b) karmaşık kelimeler muhtemelen sağdan-sola veya tek parça olarak değil, soldan-sağa işlenmektedir, ve en önemlisi, (c) insan beyni karmaşık kelimeleri işlerken ek dizilerini kullanabilmektedir, bu ise, sık kullanılan eklerin yanısıra, sık kullanılan ek dizileri için de ayrı zihinsel kayıtlar bulunduğunu düşündürmektedir.

ACKNOWLEDGMENTS

A surprisingly large number of colleagues, academics, classmates, family members, friends, and anonymous participants have contributed to this thesis. A year ago, I had no idea that writing a master's thesis was such a collaborative effort. So, I would like to thank them all:

Esra Mungan, Aysu Mutlutürk, and Ali Tekcan from the Psychology Department of Boğaziçi University, Tınaz Ekim Aşıcı and Mustafa Baydoğan from the Industrial Engineering Department of the same university, Simay İkier from the Cognitive Science Department of Yeditepe University, Josh de Leeuw from the Cognitive Science Department of Indiana University, and Harald Baayen from Tübingen University have kindly responded to various appointment requests and quite lengthy e-mails, and have given valuable advice on experiment design, mathematical modelling, and statistical analysis. Elise Michon from the Cognitive Science Department of Boğaziçi University both listened to my informal progress reports more than once, and also pointed to a relevant paper that inspired the second experiment, which is the backbone of this study.

My colleagues Azize Akın, Hülya Doğramacılar, Şenol Karakaş, Umut Köse, Sultan Şahin, Güldalı Uğurlu, and Halil Ummak have been literally forced (by me) to participate in a series of pilot studies, and criticize anything related to experiment design, which has helped me avoid disasters more than once.

My father Ömer Bilgin, my mother Ülker Bilgin, my sister İdil Bilgin Sait, my brother-in-law Emre Sait, my 8- and 10-year-old nephews Arda Sait and Eren Sait, my aunt-colleagues Suna Banguoğlu and Çiğdem Aka, numerous high-school and university classmates, including Begüm Avar and Ozan Mahir Yıldırım, had to both participate in my pilot studies and listen to lots of whining for more than one-and-a-half years.

Halit Yurdal Tilki, Merve Tilki, as well as a large number of friends have helped me by sharing the links to my online experiments on the social media.

My committee members Arzucan Özgür, from the Computer Science Department of Boğaziçi University, and Serkan Şener, from the Cognitive Science Department of Yeditepe University have kindly accepted to be on my thesis committee. Arzucan Özgür moreover had to bear with me while I abandoned my original thesis subject, and shared with her the many details of the present one. Aslı Göksel from the Linguistics Department of Boğaziçi University, and Albert Ali Salah from the Computer Science Department of the same university have spent a lot of time and brain-power reading and commenting on many linguistic and mathematical aspects of this thesis. And of course, special thanks go to my advisor Mine Nakipoğlu from the Linguistics Department of Boğaziçi University, who has offered an entire one-semester course for the sole purpose of helping me and a few other classmates find a fruitful thesis subject. She also had to listen to my obsessive comments, tolerate my stubbornness, read unending e-mails, and extend deadlines quite a few times.

The online experiments designed for this thesis have been completed by more than 2,500 anonymous people, who have voluntarily spent valuable time in front of their computers, trying to decide if monstrous letter-sequences like *orangutanlaştırmamalıyız* are valid Turkish words, or just plain nonsense.

Finally, I would like to heartfully thank Evrem for listening to my complaints and unintelligible early-morning, mid-day and late-night mumblings for so many months, and also for agreeing to marry me around 40 hours before my thesis defense.

To my grandfather, for having instilled in me a love for words and languages. . .

TABLE OF CONTENTS

CHAPTER 1: INTRODUCTION 1

CHAPTER 2: LITERATURE REVIEW AND LINGUISTIC BACKGROUND 9

2.1 Morphological processing 10

2.2 Frequency effects 13

2.3 Lexical decision 22

2.4 Background on Turkish 24

2.5 Relevant work on Turkish 27

CHAPTER 3: PREPARATORY WORK 38

3.1 Design decisions 38

3.2 Notation, definitions and models 42

3.3 Measuring frequency 58

3.4 Processing corpus data 63

CHAPTER 4: METHODOLOGY AND THE EXPERIMENTS 79

4.1 Designing a platform for online experiments 79

4.2 Experiment 1 83

4.3 Results of Experiment 1 92

4.4 Experiment 2 101

4.5 Results of Experiment 2 120

CHAPTER 5: GENERAL DISCUSSION 123

5.1 Secondary findings 123

5.2 Main findings 130

CHAPTER 6: CONCLUSION AND FUTURE WORK 136

6.1 Continue quantification 136

6.2 Publish datasets 137

6.3 Analyze existing data 137
6.4 Additional experiments 138
APPENDIX A: SUFFIXES OF TURKISH 140
APPENDIX B: MOST FREQUENT LETTERS OF TURKISH ACCORDING TO BOUN CORPUS 143
APPENDIX C: MOST FREQUENT LETTER BIGRAMS OF TURKISH ACCORDING TO BOUN CORPUS 145
APPENDIX D: MOST FREQUENT LETTER TRIGRAMS OF TURKISH ACCORDING TO BOUN CORPUS 146
APPENDIX E: MOST FREQ. SUFFIXES OF TUR. ACC. TO BOUN CORPUS 147
APPENDIX F: MOST FREQUENT SUFFIX BIGRAMS OF TURKISH ACCORDING TO BOUN CORPUS 150
APPENDIX G: MOST FREQUENT SUFFIX TRIGRAMS OF TURKISH ACCORDING TO BOUN CORPUS 151
APPENDIX H: MOST FREQUENT NOMINAL SUFFIX TEMPLATES OF TURKISH ACCORDING TO BOUN CORPUS 152
APPENDIX I: NON-WORD STIMULI USED IN EXPERIMENT 1 156
APPENDIX J: REAL-WORD STIMULI USED IN EXPERIMENT 1 157
APPENDIX K: THREE SETS OF NOUN ROOTS USED IN EXPERIMENT 2 .. 158
APPENDIX L: SUFFIX TEMPLATES, WORD-FORMS AND NON-WORDS IN EXPERIMENT 2 159
APPENDIX M: INSTRUCTIONS USED IN EXPERIMENT 2 162
APPENDIX N: DESCRIPTIVE STATISTICS FOR EXPERIMENT 1 163
APPENDIX O: DESCRIPTIVE STATISTICS FOR EXPERIMENT 2 167
REFERENCES 173

LIST OF TABLES

Table 1. Four Voiceless Consonants and their Voiced Counterparts........................ 26

Table 2. Inflectional Bundles of the Stem göz ‘eye’.. 49

Table 3. Inflectional Bundles of the Stem ‘eye’ 50

Table 4. Derivational Bundles of the stem göz ‘eye’.. 52

Table 5. Inflectional Template Bundles.. 53

Table 6. Derivational Template Bundles 54

Table 7. Six Possible Interpretations of the String koyun.. 61

Table 8. Two Possible Interpretations of the Sequence *hüzün insanı derinleştirir* .. 62

Table 9. Aggregating the Tokens in BOUN Corpus.. 65

Table 10. Aggregating the Templates in the BOUN Corpus 66

Table 11. Correlation Between the Four Frequency Measures................................. 77

Table 12. Means and Standard Deviations of Independent and Control Variables .. 91

Table 13. Distribution of Subjects by Education Level (Experiment 1)................... 93

Table 14. Distribution of Subjects by Experiment Version (Experiment 1)............. 93

Table 15. Stimuli Removed from Experiment 1 Results due to Low Accuracy....... 94

Table 16. Number of Subjects Removed from Experiment 1, by Version 95

Table 17. Mean RT Values in the Five Conditions of Experiment 1........................ 95

Table 18. p-values for the Shapiro-Wilk Normality Tests.. 96

Table 19. Results of the Wilcoxon Test for Exp. 1 (HIGH vs. LOW)...................... 97

Table 20. Mean RT Differences and Effect Sizes for Exp. 1 (HIGH vs. LOW) 99

Table 21. Stems That Contain Agt, Ness, With and Without 105

Table 22. Correlation Btw. Independent Var. and Control Variables in Exp. 2 108

Table 23. Means and Std. Deviations in Experiment 2 (HIGH & LOW)............... 109

Table 24. Distribution of Subjects by Education Level (Experiment 2)................. 113

Table 25. Distribution of Subjects by Experiment Version (Experiment 2)........... 114

Table 26. Stimuli Removed from Experiment 2 Results due to Low Accuracy..... 114

Table 27. Number of Subjects Removed by Version (Experiment 2) 115

Table 28. Mean RT Values in the Twelve Conditions of Experiment 2................. 116

Table 29. p-values for the Shapiro-Wilk Normality Tests (Experiment 2)............. 117

Table 30. Results of the Wilcoxon Test for Experiment 2 (HIGH vs. LOW) 118

Table 31. Results of the Wilcoxon Test for Experiment 2 (PR vs. RP).................. 119

Table 32. Mean RT Differences and Effect Sizes for Exp. 2 (HIGH vs. LOW) 119

Table 33. Mean RT Differences and Effect Sizes for Exp. 2 (RP vs. PR).............. 120

Table 34. Global Average RTs and Accuracies in Exp. 1 and Exp. 2 129

LIST OF FIGURES

Figure 1. Vowel features 25

Figure 2. A suffix tree representing the accumulation of suffixes on a root 45

Figure 3. Vocabulary size as a function of corpus size 68

Figure 4. A small subset of the suffix tree 70

Figure 5. A particularly sparse region of the suffix tree 71

Figure 6. Frequency rank vs. coverage 72

Figure 7. Four-dimensional frequency distribution of Turkish simplex nouns 77

Figure 8. Registration page for participants 80

Figure 9. Potential stimuli located at the core of the word-cloud 84

Figure 10. Correlation as a function of sampling cylinder diameter 86

Figure 11. Distribution of subjects by age (Experiment 1) 92

Figure 12. Distribution of subjects by age (Experiment 2) 113

LIST OF ABBREVIATIONS

B_{drv}: derivational bundle

B_{inf}: inflectional bundle

BNC: British National Corpus

BOUN: Boğaziçi University

EEG: electroencephalography

ERP: event-related potentials

fMRI: functional magnetic-resonance imaging

PCA: principal component analysis

RT: response time

$T_{\emptyset}$: empty template

T_{CM}: compound template

T_{drv}: derivational template

T_{inf}: inflectional template

TSCorpus: The Turkish Corpus (www.tscorpus.com)

CHAPTER 1
INTRODUCTION

1.1 Overview

According to the "computational theory of mind", the human brain is a computational system that runs algorithms on representations (Markman, 2006; Rescorla, 2015). It creates, stores, accesses and processes representations as it interacts with the external world through the sensory systems, but it also creates, stores, accesses and processes non-sensory representations as it interacts with other parts of itself, i.e. with existing mental representations elsewhere in the brain. Understanding how the brain represents knowledge and what kinds of algorithms operate on those representations has been one of the main objectives of the cognitive sciences (Friedenberg & Silverman, 2011).

Language is one of the most widely studied human faculties in the scientific effort to understand the nature of mental representations and processes. Cognitive science studies language strictly as a mental phenomenon rather than as an abstract, ideal system detached from its biological realization in the human brain. Studying language is hoped to "offer a window into cognitive function, providing insights into the nature, structure and organization of thoughts and ideas", because it is assumed that language "reflect[s] certain fundamental properties and design features of the human mind" (Evans & Green, 2006).

The present study is an inquiry into how the human brain processes language. More specifically, it tries to shed light on how words are represented in the mind and how these representations are accessed during the act of reading (visual word recognition). Even more specifically, it tries to understand how morphologically simple and complex words in Turkish are represented in the minds of native speakers of Turkish, and how these representations are accessed during visual word recognition.

Since mental processes are largely unconscious, introspection is of limited value when studying the mind (Nisbett & Wilson, 1977, p. 231). No amount of introspection can reveal, for instance, that the signals coming in from the left visual fields of both eyes are initially processed by the primary visual cortex at the back of the brain's right hemisphere, or that language is primarily processed in the left hemisphere. Thus, hypotheses about the mental representation and processing of language can be better examined by using experimental methods (Hasson & Giora, 2007, p. 302). This is why this study follows the tradition of experimental psychology, and especially the tradition of "mental chronometry".

"Mental chronometry" refers to the systematic measuring of subjects' reaction times in perceptual and motor tasks, for the purpose of reaching conclusions about the nature of mental operations (Meyer, Osman & Irwin, 1988, p. 3). It is one of the oldest and most widely-used behavioral methods in experimental and cognitive psychology for understanding the workings of the human mind. The earliest studies go back more than 130 years: In 1885, the American psychologist James McKeen Cattell "pasted letters on a revolving drum … and determined at what rate they could be read aloud". Cattell found, among other things, that "it takes about twice as long to read … words which have no connexion as words which make sentences, and letters which have no connexion as letters which make words" (Cattell, 1886, p. 64, but also see Donders, 1969). These were the beginnings of psychology as an experimental science.

The type of behavioral experiment used in this study is known as a "lexical decision task". Subjects are shown strings of letters, usually on a computer screen, and are asked to decide, as quickly and accurately as possible, whether or not the letters constitute a valid word, by pressing either the "yes" button or the "no" button. The time between the presentation of the stimulus and the subject's pressing one of the two buttons is known as "response time" or "reaction time" (RT), and is the most widely used dependent variable, usually measured at millisecond accuracy.

Stimuli consist of valid words (like *desen* 'pattern' or *cüzdan* 'wallet'), non-words designed in accordance with the purpose of the experiment (like **cüzdar* or **fodtsz*), and a considerable number of "fillers", usually words from different parts-of-speech (like *geliyoruz* 'we are coming' or *hızla* 'quickly'), whose sole purpose is to prevent the subjects from guessing the purpose of the experiment. Although interesting inferences can be made based on the response times of the non-words as well, the experimenter is primarily interested in the response times of the real words.

In the two-condition experiments used in this study, the independent variable has two levels: low and high. Half of the real words belong to the low-level condition, and the other half to the high-level condition. For example, if the independent variable is word frequency, half of the real words used in the experiment are low-frequency words, and the other half are high-frequency words. Except for the independent variable, all variables which the experimenters think might affect the dependent variable are kept constant. In this way, changes in the dependent variable are hoped to be exclusively caused by changes in the independent variable.

All subjects of the experiments in this study have been exposed to all stimuli in the two frequency groups. This is known as a "within-subjects" or "repeated-measures" design, and has the considerable advantage of eliminating the variability arising from individual differences between subjects, as will be explained in more detail below.

The independent variables used in this study are various word frequency measures. There exists an extensive and long-standing literature showing that frequency of occurrence is a strong predictor of reaction time in word recognition experiments (Howes & Solomon, 1951; Forster & Chambers, 1973; Taft, 1979; Grainger, 1990; Meunier & Segui, 1999; Taft, 2004, among many others). There exists a statistically significant negative correlation between word frequency and reaction time. In other words, more frequent words are recognized more quickly (and more accurately).

However, as will be seen below, frequency is a surprisingly complex concept. Take the English word *institutionalized*, for instance. We can simply count how many times this letter-string occurs in a large-scale text collection like the British National Corpus (BNC)[1]. It turns out that *institutionalized* occurs 157 times in BNC. But this can either be the simple past of the verb *institutionalize*, or an attributive adjective as in the phrase *institutionalized violence*. Already, we have three different frequency measures, one for the past-tense form, one for the adjective form, and one for both forms in the aggregate. Moreover, the word *institutionalized* is clearly related to the verb *institutionalize,* the adjective *institutional*, the noun *institution*, and finally the verb *institute*. One could claim that, when a person is exposed to the word-form *institutionalized*, his/her brain also activates these related base-forms,

[1] A "corpus" is a large-scale collection of text or speech samples in a given language (pl. "corpora"). A corpus serves as a source of naturalistic data that can be used to discover statistical regularities, test hypotheses, write grammars, compile dictionaries or other reference materials, train machine learning systems, etc. The British National Corpus, on the other hand, is "a 100-million-word collection of samples of written and spoken language from a wide range of sources, designed to represent a wide cross-section of British English, both spoken and written, from the late twentieth century". It is available at http://www.natcorp.ox.ac.uk.

from which the word-form has been derived. Are their frequencies updated, too? Maybe we should define a system of "diminishing activation" based on the number of steps involved in the derivation, i.e. assume that a single activation of *institutionalized* results in, say, 0.8 activations of *institutionalize*, 0.6 activations of *institutional*, 0.4 activations of *institution*, etc., whatever such partial activations may mean. Or maybe we should define a single aggregate metric for all of them (activation of the "*institute* family"). But shall we simply follow the chain of derivation until we reach the bare root (*institutionalized* > *institutionalize* > *institutional* > *institution* > *institute*) and stop there, or does the family also include other "children" of the ultimate bare root (e.g. *institutionalization*)? Moving on to inflection, when a person is exposed to the inflected form *institutionalizes*, does his/her brain activate only that surface-form, only the base-form *institutionalize*, or both? The situation is even more complicated in the case of Turkish, which is the language used in this study.

The present study attempts to contribute to existing literature in three ways: (a) to propose a general notation for describing various frequency measures in a principled way, (b) to test if the frequency effect, which has been demonstrated so many times in the literature, also exists in a typologically very different language, and (c) to experimentally test existing models of morphological processing in a morphologically complex language. By defining various lexical frequency measures and by manipulating them in controlled experiments, we hope to shed light on the architecture of the linguistic representations and processes in the human brain.

1.2 Research questions and hypotheses

This study aims to experimentally investigate the effects of frequency on the time it takes subjects to visually recognize morphologically simple and complex Turkish words, and the accuracy of such recognition. In this context, the following three research questions will be addressed:

Research Question 1: Does frequency affect response times in the visual recognition of bare Turkish nouns?[2]

Research Question 2: Does frequency affect response times in the visual recognition of morphologically complex Turkish nouns?

Research Question 3: Are morphologically complex Turkish words parsed from left to right, from right to left, or as a single chunk?

The following hypotheses are derived from these research questions:

Hypothesis 1: Everything else being equal, the time it takes a native speaker of Turkish to visually recognize a bare Turkish noun decreases as the frequency of that noun increases.

Hypothesis 2: Everything else being equal, the time it takes a native speaker of Turkish to visually recognize a morphologically complex Turkish noun decreases as the frequency of the suffix-template (defined in Section 3.2.1) increases.

Hypothesis 3: Everything else being equal, the time it takes a native speaker of Turkish to reject a non-word that starts with a meaningless letter-sequence but ends with a valid sequence of suffixes (e.g. *gansien+lerinizin*) is shorter than the time it takes a native speaker of Turkish to reject a non-word that starts with a valid root but ends with meaningless letters (e.g. *bıldırcın+ganfirmaş*).

[2] Although the existence of the frequency effect has been demonstrated countless times in the literature, it has not been thoroughly and exhaustively investigated for Turkish, as will be discussed in Section 2.4 of the literature review.

1.3 Experimental findings

Two online experiments have been conducted to address these research questions and hypotheses:

The first experiment addresses the first research question, and tests the corresponding first hypothesis. It finds clear evidence that, everything else being equal, more frequent noun roots are processed significantly faster than less frequent noun roots, thus replicating a basic finding that has been demonstrated for a large number of languages since the 1950s.

The second experiment addresses the second and third research questions, and tests the corresponding second and third hypotheses. It finds that, everything else being equal, complex-words that contain high-frequency suffix sequences are processed significantly faster than complex-words that contain low-frequency suffix sequences. This finding supports the hypothesis that there exist separate mental representations for frequently occurring suffix sequences (not individual suffixes but entire suffix bundles such as *lAş+DIr+Il+mIş*). This is in line with “usage-based” accounts of grammar, which claim that linguistic structure emerges from language use, i.e. from repeated exposure to certain constructions over time (Bybee, 2011, p. 69).

Another finding of the second experiment is that novel complex word-forms such as *orangutanlaştırmamalıyız* ‘we should not orangutanize’ are processed step-by-step, starting with the root and then proceeding with the suffixes, as suggested by Hankamer (1989), rather than being processed as a single chunk, or from right-to-left, starting with the rightmost suffix and then proceeding with the other suffixes and the root.

Several other, more general and/or informal findings have been derived from the preparatory quantitative work reported in Chapter 3, which will be discussed in Chapter 5.

CHAPTER 2
LITERATURE REVIEW AND LINGUISTIC BACKGROUND

The scientific literature on language and the human brain is extensive, multidisciplinary, relatively young, and thus full of theoretical and practical divides, terminological issues, and contradictory results, as will be discussed below.

Methodological diversity is also striking: There are neurolinguistic studies that try to understand the neurological bases of language, using brain imaging techniques like functional magnetic resonance imaging (fMRI) and electroencephalography / event-related potentials (EEG/ERP); more traditional psycholinguistic studies (like the present study) that use behavioral experiments such as lexical decision, masked priming, progressive demasking, and naming to measure response times; modelling efforts that computationally or otherwise try to simulate the processing of language in a biologically plausible way.

This literature review is limited to the basic phenomena and methods discussed in this study: (a) visual word recognition (and especially the recognition of morphologically complex words) as the object of study, (b) frequency of occurrence as the independent variable, and (c) lexical decision as the experimental task.

"Visual word recognition" refers to a cognitive task where a human subject is visually exposed to a string of written symbols (as opposed to a sequence of sounds, which is the field of auditory word recognition) and recognizes those symbols as a valid lexical item. Note that this includes a wide range of cognitive processes from the low-level visual perception of letter shapes, to the higher-level (e.g. phonological, morphological and semantic) processes required to retrieve the mental entry for the relevant lexical item, its internal structure and meaning. Also note that the above definition is limited to the relatively more artificial task of recognizing single words presented in isolation, and thus excludes the relatively more realistic, everyday task of processing words in context (i.e. sentence processing).

The first three sections below summarize the literature on morphological processing, frequency effects and lexical decision, respectively. The fourth section provides a brief introduction to the Turkish language, with a special emphasis on its morphology. Equipped with this background knowledge, the fifth and last section offers a review of existing theoretical and empirical literature on the mental representation and processing of Turkish words.

2.1 Morphological processing

"Morphological processing" refers to the mental representation and processing of morphologically complex words, i.e. words composed of more than one morpheme, such as *normalization*, *worldwide* or *international*. The main dividing issue is if there exists a separate mental entry for each and every word, whether simple or complex (the "full-listing hypothesis"), or if roots and affixes are stored separately and are only combined during comprehension or production based on abstract morphological rules (the "decomposition hypothesis"). Three positions have been defended in the literature, including these two extremes, and a family of several positions in-between (Schreuder et al., 1990):

According to the more extreme versions of the "full-listing / whole-word representation / direct route" hypothesis (e.g. Butterworth, 1983; Seidenberg, 1987), words are directly represented and accessed in their entirety, regardless of whether they are morphologically simple or complex. In other words, there is a separate mental representation for each word-form (e.g. one entry each for *norm*, *normal*, *normalize*, *normalizes*, *normalizing*, *normalization*, etc.), and word recognition is simply a search/look-up process that does not involve any further computation.

According to the more extreme versions of the “decomposition / full-parsing” hypothesis (e.g. Taft & Forster, 1975; MacKay, 1978), on the other hand, only roots, affixes and morphological rules governing their combination are mentally represented, and word-forms are computed on-the-fly, through the application of the rules to those roots and affixes. In other words, there exists a single base-form entry for, say, *norm*, and the full-form *normalize*, for example, is recognized after being processed by the “morphological parser”, which identifies the constituents *norm*, *-al* and *–ize*, and computes the meaning of the complex word-form using the meanings of the three parts that have been identified (Baayen, Dijkstra, & Schreuder, 1997).

Finally, “dual-route models” (e.g. Caramazza, Laudanna, & Romani, 1988) are not a separate family of models but a range of intermediate positions between the two extremes described above. In dual-route models, both full-forms and roots/affixes/rules can be mentally represented. Thus, the full-form *normalize*, for example, can be represented both as *normalize* and as *norm+al+ize*.

The question that arises at this point is whether there exist both full-form and decomposed representations for all words regardless of their morphological structure, or whether some words prefer the full-form/direct route, while others use the decomposition/parsing route. Several positions have been defended in the literature:

Pinker (1991) and Pinker & Prince (1994), for instance, argue that phonologically and semantically regular and transparent forms are not stored as full-forms but are always processed on the basis of the morphemes that constitute them. Opaque complex forms that have at least one idiosyncratic property, on the other hand, have to be stored in their entirety. They further argue that inflected words tend to be stored in decomposed form, while derived words are more prone to being stored as full-forms, since inflectional processes are more regular and transparent than derivational processes. Other researchers, however, have empirically demonstrated that even the most regular and transparent inflected forms can be stored as full-forms, especially if they are used frequently (Stemberger & MacWhinney, 1986). According to the Augmented Addressed Morphology model (Burani & Caramazza, 1987; Caramazza, Laudanna & Morani 1988; Laudanna & Burani, 1985), for example, frequently-occurring words are processed directly via their full-form entry, while the decomposition / parsing route serves as a "backup" that is only used for rare or novel complex words that are regular and transparent. Taft (1979), on the other hand, proposes that there exists an obligatory parsing step prior to lexical access, where the parser initially, and blindly, attempts to separate any incoming letter-string into constituent morphemes, regardless of whether or not the string actually contains more than one morpheme (Baayen, Dijkstra, & Schreuder, 1997).

Finally, there are "parallel dual-route" models that use the "race" metaphor to conceptualize the respective roles of full-form storage and decomposition in visual word recognition. According to these models, there exist both full-form and decomposed representations for all word-forms, and two parallel and independent processes engage in a "race" with each other using these representations: The full-form/direct route uses the full-form representation to access the phonological and semantic representation of the word directly, while the decomposition/parsing route uses morphemes and morphological rules to attempt a combinatorial interpretation (e.g. Frauenfelder & Schreuder, 1992; Schreuder & Baayen, 1995).

Which of these models of morphological processing does empirical evidence support? What are the uncontroversial empirical facts? The picture does not seem to be bright in this regard: According to Amenta & Crepaldi (2012), for example, although a considerable amount of experimental data resulting from more than sixty years of research is available, the debate has become rather inconclusive, especially during the last ten years. New models are being proposed without clearly falsifying existing models, and as a result, knowledge on morphological processing does not progress in a cumulative fashion ("which means that . . . it does not progress at all", the authors add). Amenta & Crepaldi (2012) think that the basic reason for this state of affairs is that we still do not have a comprehensive list of uncontroversial empirical facts (Amenta & Crepaldi, 2012, p. 1).

2.2 Frequency effects

In its simplest form, "frequency effect" refers to the fact that more frequent words are recognized faster. For example, a high-frequency word such as *dog* is, on the average, recognized faster than a low-frequency word such as *doe*, "everything else being equal".[3] This effect is probably the most well-known phenomenon in psycholinguistic literature, with the earliest empirical studies going back more than sixty years (e.g. Howes & Solomon, 1951).

[3] As in many other fields of science, it is extremely difficult to make sure that "everything else is equal". In a rather humorous article that warns researchers that "we may not be able to run any psycholinguistic experiments at all in 1990", Cutler (1981) makes the following remark:

> Consider a judicious psycholinguist constructing materials for an experiment comparing nouns, verbs and adjectives. Ideally he would like to create matched triples of an unambiguous noun with an unambiguous verb and an unambiguous adjective. They should be matched, as we have seen, on both surface and combined frequency. Naturally they should be matched on length. At this point it is already clear to the experimenter that the task is probably impossible; and he has not even begun to consider further variables on which they might be matched, such as association, age of acquisition, autobiographical memory, categorizability, concreteness, digram frequency, imagery, goodness, letter frequency, number of meanings, orthographic regularity, meaningfulness, emotionality and recognition threshold. (Cutler, 1981, p. 68)

Frequency effects have been used extensively as a diagnostic tool to understand the mental representation and processing of language: They have been demonstrated for a large number of languages (e.g. Arabic, Dutch, English, Finnish, French, German, Hebrew, Italian, Korean, Russian), and using a variety of experimental tasks (e.g. lexical decision, priming, masked priming, progressive demasking, naming), independent variables (e.g. the frequency of surface-forms, base-forms, family members, orthographic neighbors, syllables, ngrams), and dependent variables (e.g. response time, eye-movement duration, electrophysiological brain responses) (Segui, Mehler, Frauenfelder, & Morton, 1982; Baayen, Milin & Ramscar, 2016).

The facilitatory effect of frequency on perception is probably not limited to the domain of language, and may also exist in unrelated domains such as color perception or event recall. In Broadbent's words, "the fact that common words are, other things being equal, more easily perceived is perhaps only a special case of the general influence of probability on perception" (Broadbent, 1967, p. 1).

2.2.1 "What's in a name?"

Why does the literature on probability and perception almost exclusively focus on the recognition of words if frequency effects are not limited to the domain of language? There are several reasons:

First of all, "unlike other types of stimuli like visual images, it is easy to quantify the probability of occurrence of words. . . [Words] thus provide a convenient special tool for investigating the general question of probabilistic effects in perception" (Broadbent, 1967, p. 1).

Secondly, although it looks simple at first sight, a single written word is an astonishingly complex artifact that can be analyzed, and quantified, at a large number of intrinsic and extrinsic levels: size, shape, contour, letter features, uniqueness point, surface frequency, base frequency, ngram frequency, number of letters, number of phonemes, number of syllables, number of morphemes, polarity, emotional content, imageability, family size, family frequency, neighborhood size, neighborhood frequency, age of acquisition, number of synonyms, number of senses, collocates, visual recognition time, visual recognition accuracy (this list is partially based on Rubin, 1980, and probably does not exhaust all possibilities).

Finally, the mental recognition of a word activates several types of associated mental representations at several levels including semantics, phonology, orthography, prosody, pragmatics and syntax, and is thus a suitable object of inquiry for studying the overall architecture of the mind (Seidenberg & Mclelland, 1989).

2.2.2 Defining frequency

Existing literature on word frequency is largely based on English and other morphologically impoverished languages, and therefore contains only a few, basic definitions of frequency, which prove inadequate in the case of an agglutinating language like Turkish, as will be seen in Section 3.2.2. Furthermore, existing frequency terms involve overlaps and omissions, and are in many cases used by researchers without even being defined. The following sub-headings critically discuss existing frequency measures in the literature:

2.2.2.1 Surface frequency

This is the most basic and uncontroversial frequency measure, and simply counts how many times a word-form occurs, exactly as it appears in the text, regardless of its internal structure, if any. For example, to compute the surface frequency of the complex Turkish word-form *denizdekiler* ‘those at sea’, one simply counts how many times the letter-sequence *_d-e-n-i-z-d-e-k-i-l-e-r_* occurs in a certain corpus. Note that this does not include occurrences of the related base-form *deniz* ‘sea’ or more complex forms like *denizdekilerle* ‘with those at sea’. Only the surface-form *denizdekiler* is included.

The terms “whole-word frequency” (e.g. Alegre & Gordon, 1999; Niswander, Pollatsek, & Rayner, 2000) and “word-form frequency” (e.g. Ford, Davis, & Marslen-Wilson, 2010; Keuleers, Brysbaert, & New, 2010) appear to be synonyms.

2.2.2.2 Base frequency

This is the most important family of measures in terms of the processing of morphologically complex words, and variously refers to the number of times the base / root / stem / lemma of a word-form (and sometimes also all inflected forms or even all forms whether inflectional, derivational or compound) occurs in a given corpus.

Unfortunately, the literature is full of nearly synonymous and partially overlapping definitions. For example, if one adopts the definition in Taft (1979) (“total frequency of the stem plus all inflected forms”), the base frequency of *agrees* would be the surface frequency of the base-form *agree* plus the surface frequencies of *agrees, agreed, agreeing,* etc. If one adopts the definition in Ford, Davis, & Marslen-Wilson (2010) (“the whole-form frequency of the base”), on the other hand, the base frequency of *agrees* would only include the surface frequency of the base-form *agree*. Finally, if one adopts the definition in Vannest, Newport, Newman, & Bavelier (2011) ("total frequency of all the words containing a given base morpheme"), the base frequency of *agrees* would be the sum of the surface frequencies of *agree, agrees, agreed, agreeing, agreeable, agreement, disagree,* etc.

"Stem frequency" (e.g. Schreuder & Baayen, 1997), "root frequency" (e.g. Vannest & Boland, 1999) "cumulative stem frequency" (e.g. Baayen, Dijkstra, & Schreuder, 1997), "cumulative root frequency" (e.g. Giraudo, & Grainger, 2000), "cumulative morpheme frequency" (e.g. Ford, Marslen-Wilson & Davis, 2003), "base morpheme frequency" (e.g. Ford, Davis, & Marslen-Wilson, 2010), "lemma frequency" (e.g. Solomyak, & Marantz, 2010), and "cluster frequency" (e.g. Alegre, & Gordon, 1999) appear to be closely-related and partially-overlapping terms, but a detailed definition of each of them will not be attempted here.

2.2.2.3 Family size and family frequency

The (morphological) "family" of a stem is the set of all words obtained from that stem through derivation or compounding (see, for example, Schreuder & Baayen, 1997). For example, the family of the stem *book* would include, among many others, the derived forms *bookmark*, *textbook*, *bookish*, and *unbook* (the literature is not clear on whether the stem *book* itself, or inflected forms like *bookmarks*, *textbooks*, *unbooking* are to be included in the family).

Based on this definition, the "family size" of a stem is the number of words in that stem's family. For example, if we assume that *bookmark*, *textbook*, *bookish*, and *unbook* are the only derived forms of the stem *book,* the family size of book would be equal to four. Similarly, the "family frequency" of a stem is the total frequency of all words in that stem's family. In other words, the family frequency of the stem *book* would be the total surface frequencies of *bookmark*, *textbook*, *bookish*, and *unbook*. As can be seen, family size is a type-based measure, while family frequency is token-based.

Family frequency is sometimes referred to as “cumulative family frequency” (e.g. Schreuder & Baayen, 1997).

2.2.2.4 Neighborhood size and neighborhood frequency

The “(orthographic) neighborhood” of a word is the set of all other words that can be obtained from the first word through one or more orthographic manipulations such as adding, substituting or deleting a letter or swapping the positions of two letters (see, for example, Coltheart, Davelaar, Jonasson, & Besner, 1977; Andrews, 1992; Sears, Hino & Lupker, 1995).[4] For example, assuming that we allow only a single substitution or deletion and ignore other possible manipulations, or repetitions of the same manipulation, the neighborhood of the word *start* would be the set {*tart*, *smart*, *stare*, *stark*}. Parallel to the above definitions of family size and family frequency, the “neighborhood size” of *start* would thus be equal to four, and its “neighborhood frequency” would be the total surface frequencies of *tart*, *smart*, *stare*, and *stark*. Once again, the literature is not clear on whether inflected forms such as *tarts*, *stared*, and *staring* are to be included.

2.2.3 Types of frequency effects

Having defined various frequency measures, we now turn to a brief discussion of the frequency effects demonstrated in the literature using these measures.

[4] This is closely related to the concept “Levehnstein distance”, which quantifies the similarity between two strings, and is equal to the number of deletions, insertions, or substitutions required to transform the first string into the second string.

2.2.3.1 Surface-frequency effect

This is the simplest form of frequency effect, since it ignores the internal structures of words and solely focuses on the actual word-forms that occur on the surface. Earlier theories of word recognition assumed that each word, whether simple or complex, was represented as a separate atomic entity (Ford, Davis, & Marslen-Wilson, 2010, p. 117).

Although there is not much to say on the surface frequency effect *per se*, except that it is the simplest possible form of a frequency effect and was the first such effect to be demonstrated in the literature, the picture becomes interesting when it is combined with the other frequency effects discussed in subsequent sections.

2.2.3.2 Base-frequency effects

Although the literature is inconclusive in several regards, a large number of studies agree that certain morphologically-motivated frequency measures (as opposed to, or in addition to, the simple surface frequency of the actual word-form) have an effect on response times in visual word recognition tasks. For example, in one of the most influential studies in the field, Taft (1979) finds that, everything else being equal, the time it takes to recognize a word-form that contains a certain root decreases as the total frequency of all word-forms that contain the same root ("base frequency") increases. Taft describes his reasoning as follows:

> If an inflected word (e.g., likes) were recognized by stripping off the suffix "s" and by then locating the lexical entry for its stem (like), then there would be no lexical entry represented as the inflected word (likes). Rather, "likes," "liking," and "liked" would all be accessed through a single entry, namely, "like". The effect of this on frequency would be that the frequency of the lexical entry "like" would be the summed frequency of "like," "likes," "liking," and "liked," [referred to as "base frequency"] This is the single-entry model. . .

> The separate-entry model . . . would say that “like,” “likes,” “liking,” and “liked” are accessed through separate lexical entries. According to this model, the frequency influencing the recognition time to “likes” will be the frequency of “likes” alone, and the frequency influencing recognition time to “like” will be the frequency of “like” alone.
>
> . . . For example, the words “sized” and “raked” both have a frequency of 4 according to Kučera and Francis (1967) and constitute a matched pair. However, the base frequency of the two words of a pair differed markedly. The frequency of “sized” + “size” + “sizes” + “sizing” is 154, while the frequency of “raked” + “rake” + “rakes” + “raking” is only 15. Therefore, if the single-entry model is correct, words like “sized” should be recognized more quickly than words like “raked,” since base frequency should influence lexical decision times. If, however, the separate-entries model . . . is correct, there should be no difference in response times to “sized” and “raked,” since they are matched on surface frequency. (Taft, 1979, pp. 267-268)

The existence of this “base-frequency effect” is assumed to show that the human brain uses morphemes as effective processing units in the recognition of complex words. In other words, there must be some abstract internal representation that combines all word-forms derived from the same root into a single family and computes a single frequency value for the entire family rather than (or in addition to) separate frequency values for each individual word-form (also see Murrel & Morton, 1974). Given that it has been replicated many times in the literature, all models of morphological processing must explain the base-frequency effect.

2.2.3.3 Family size and family frequency effects

Schreuder & Baayen (1997) demonstrate that family frequency (which they define as “the summed frequencies of the formations in the morphological family”) does not affect response time in a series of lexical decision, progressive demasking and subjective frequency experiments. Family size (i.e. the number of items in the morphological family), on the other hand, is found to be inversely related to speed of recognition in a lexical decision task that uses monomorphemic nouns. In other words, everything else being equal, words with a large family (e.g. *book*, whose family members include *booklet*, *bookish*, *bookcase*, *bookshelf*, *phonebook*, *bookstore*, among many others) are recognized faster than words with a small family (e.g. *hook*, whose only other family member seems to be *unhook*) (Schreuder & Baayen, 1997, p. 118).

Evidence shows that the family size effect is semantic in nature: Schreuder & Baayen (1997) report that the family size effect they have identified in a lexical decision task disappears when the "progressive demasking" task is used instead. Progressive demasking is an experimental technique that deals with the very early, pre-lexical stages of visual recognition. The authors thus conclude that the family size effect "arises at more central, post-identification stages of lexical processing", rather than the relatively earlier, mainly visual stages of the process. They also observe that the negative correlation between family size and response time is strengthened when semantically opaque members are removed from the family size counts, thus supporting the hypothesis that semantic transparency lies at the root of the family size effect. This hypothesis is also supported by evidence from Hebrew (del Prado Martín et al., 2005) and Finnish (del Prado Martín, Bertram, Häikiö, Schreuder & Baayen, 2004).

2.2.3.4 Neighborhood size and neighborhood frequency effects

"Neighborhood effect" refers to an interaction between the neighborhood size or neighborhood frequency of a word on the one hand, and the word's recognition time on the other hand. These effects were initially identified by Coltheart, Davelaar, Jonasson & Besner (1977), who defined the "orthographic neighborhood" (a.k.a "Coltheart's N") of a word as the set of "all other words of the same length that can be generated by changing just one letter to another, preserving letter positions"

(Grainger, 1990, p. 229). The authors found that non-words with more neighbors took longer to reject in a lexical decision task, but did not find any effect on response times to actual words. Andrews (1992), in contrast, found that words with large neighborhoods were responded to faster than words with small neighborhoods.

The literature is inconclusive as to the existence, direction, size, and reasons of neighborhood-size and neighborhood-frequency effects. Assuming that the neighborhood-frequency effect is driven by semantics, this inconclusiveness is probably caused by the fact that semantic factors are not as easily quantifiable, and thus cannot be as easily controlled for, as other lexical parameters.

2.3 Lexical decision

"Lexical decision" is an experimental task that has been used innumerable times in psycholinguistics literature. In a lexical decision experiment, a subject is visually exposed to a sequence of letters on a computer screen, and is asked to decide, as quickly and accurately as possible, if the letter sequence presented on the screen is a valid word or not, and to press the 'yes' button if the letters make up a word, and the 'no' button if not.

The stimuli consist of actual words selected in accordance with the objectives of the experiment, non-words that may or may not resemble actual words, and a number of fillers aimed at concealing the purpose of the experiment. Response time, usually measured at millisecond accuracy, is assumed to reflect the "mental accessibility" of the relevant word (Hasson & Giora, 2007, p. 303), thus providing insights into the mental organization of language.

Although it has been used extensively since the 1950s, the lexical decision task has certain weaknesses:

There is some evidence that the lexical decision task “exaggerates” the effect of frequency on response time. Balota & Chumbley (1984), for example, report the results of three experiments where they compare the impact of, among other factors, word frequency on performance in three different word recognition tasks (category verification, lexical decision, and pronunciation), using the same set of stimuli. They reach the following conclusion:

> The relationship of the lexical variables to reaction time varied significantly with the task within which the words were embedded. In particular, the effect of word frequency was minimal in the category verification task, whereas it was significantly larger in the pronunciation task and significantly larger yet in the lexical decision task. It is argued that decision processes having little to do with lexical access accentuate the word-frequency effect in the lexical decision task and that results from this task have questionable value in testing the assumption that word frequency orders the lexicon, thereby affecting time to access the mental lexicon. (Balota & Chumbley, 1984, p. 340)

Baayen (2014) mentions the following major weaknesses: (1) The lexical decision task is “a metalinguistic task far removed from normal comprehension”. This is because, when reading an actual text, the reader naturally assumes that all strings are real words. Hence, normal reading does not involve any lexical decision task. Furthermore, response time data from lexical decision experiments is contaminated by several non-lexical processes such as motor processes required to press the buttons.

(2) In a lexical decision task, words are presented in isolation. However, words always occur in context in normal tasks.

(3) The choice of non-words and fillers significantly affects the results.

(4) Response time tells us nothing about the time-course of events during reading. It just gives the aggregate time it takes the subject to complete the task (Baayen, 2014, p. 1).

Despite these drawbacks, Baayen predicts that lexical decision “will become more instead of less popular in the coming years”, “as psychologists have discovered crowd sourcing and have developed apps for smartphones that can easily harvest millions of lexical decisions” (Baayen, 2014, p. 18).

2.4 Background on Turkish

Before starting to discuss the literature on the mental representation and processing of Turkish words, this section provides brief information on Turkish, and especially its morphology.

Turkish is an agglutinating language with a complex morphology. Word formation, in terms of both inflection and derivation, is almost exclusively accomplished through suffixation, resulting in long word-forms that can in some cases only be expressed by a large number of words in analytic languages like English (Göksel & Kerslake, 2005, p. 43):

gör-üş-tür-ül-me-yecek-miş-siniz

see-RECIP-CAUS-PASS-NEG-FUT-NARR-A2PL[5]

‘I’ve been told that you will not be allowed to see each other.’

Nominal morphology is complex as well:

yasa-laş-tır-ıl-ma-sı-n-dan[6]

law-BECOME-CAUS-PASS-INF-P3SG-ABL

‘from its being made into law’

[5] See Appendix 1 for a complete list of suffixes and their properties.

[6] When the possessive/compound marker *-(s)I* is followed by the ablative marker *-DAn*, an additional -n- appears between the two suffixes, which is an irregular change (see Göksel & Kerslake p. 46).

Although morphologically simpler forms occur more frequently as can be seen in Appendix H, examples like the above are not exotic peculiarities of the language: The suffix sequence *BECOME-CAUS-PASS-INF-P3SG-ABL*, for example, occurs 8,597 times (around 17 per million) in the BOUN Corpus defined below.

Two additional salient characteristics of Turkish are consonant alternation and vowel harmony. These phonological processes are relevant for the purposes of this study because the resulting orthographic changes affect the visual-word-recognition experiments designed here. Before describing these phonological processes, let us briefly describe the relevant properties of Turkish vowels and consonants:

Vowels in Turkish allow a three-way categorization, in terms of the height of the tongue, the roundedness of the lips and the frontness of the tongue (see Figure 1).

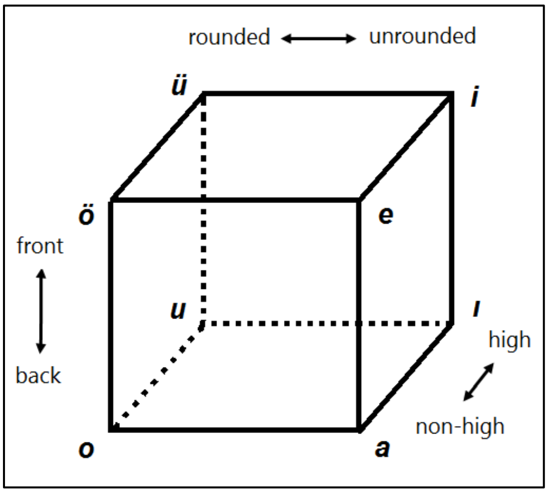

Figure 1. Vowel features

In Table 1, the eight consonants that are relevant for the purposes of the present study are categorized in terms of voicing:

Table 1. Four Voiceless Consonants and their Voiced Counterparts

Voiceless	Voiced
ç	c
k	g
p	b
t	d

There are five cases where the addition of a suffix to a stem results in an orthographic change in the stem[7]:

(1) Some stem-final voiceless consonants (see Table 1) become voiced: *dert* 'trouble' → *derdim* 'my trouble'.

(2) Some stem-final consonants get doubled: *hat* 'line' → *hattın* 'of the line'.

(3) Some high vowels disappear: *alın* 'forehead' → *alnım* 'my forehead'.

(4) Some stem-final *k'*s become *ğ*: *delilik* 'madness' → *deliliğe* 'to madness'

(5) Some stem-final *a, e, u* or *ü'*s become *ı* or *i*: *ağla* 'cry' → *ağlıyordum* 'I was crying'

Vowel harmony, on the other hand, can be described as follows:

(6) High vowels copy their frontness and roundedness features from the vowel in the preceding syllable, while back vowels copy their frontness features from the vowel in the preceding syllable: *boz* → *bozuldu* and *öp* → *öpüldü, ez* → *ezildi* and, *aç* → *açıldı*.

Apart from a few exceptions[8], suffixes in Turkish can be classified as "A-type" or "I-type". An I-type suffix contains high vowels which get their frontness and roundedness features from the preceding vowel in accordance with the fronting

[7] There are two additional types of consonant alternation, but these are not described here since they do not result in any orthographic change. See, for example, Göksel & Kerslake (2005), p. 14 for details.
[8] See Göksel & Kerslake, 2005, p. 24.

and rounding harmonies described above. The genitive suffix *-(n)In*, for instance, is an I-type suffix. An A-type suffix, on the other hand, contains unrounded and non-high vowels which get their frontness feature from the preceding vowel in accordance with fronting harmony. The plural marker *-lAr*, for instance, is an A-type suffix.

An A-type suffix prevents the vowels *u* and *ü* from surfacing in all subsequent suffixes. This phenomenon will be referred to as "blocking" below. For example, in the word *gör-se-ydi-niz* 'if you had seen', the A-type conditional suffix –*sA* blocks the allomorph *-ydü* of the subsequent suffix *-(y)dI* and the allomorph *-nüz* of the subsequent suffix *-nIz*.[9] Similarly, in the word *sor-sa-ydı-nız* 'if you had asked', the A-type conditional suffix –*sA* blocks the allomorph *-ydu* of the subsequent suffix *-(y)dI* and the allomorph *-nuz* of the subsequent suffix *-nIz*.[10]

2.5 Relevant work on Turkish

Equipped with the brief linguistic background in Section 2.4, this section offers an overview of existing work that theoretically or empirically discusses the mental representation and processing of Turkish words, with a special emphasis on the representation and processing of morphologically complex words. The review begins with two important theoretical contributions by Hankamer (1989) and Frauenfelder & Schreuder (1992), and then proceeds to empirical work by various authors.

Hankamer (1989) discusses the full-listing vs. decomposition issue described in Section 2.1, and claims that "all versions of the [full-listing hypothesis] are

[9] An "allomorph" is a phonological/orthographic variant of a morpheme. For example, the Turkish plural marker *-lAr* is realized on the surface either as *-ler* or *-lar*, depending on the last vowel of the stem (e.g. *at* 'horse' → *atlar* 'horses' vs. *et* 'meat' → *etler* 'meats').

[10] For additional details on the morphology, phonology and orthography of Turkish, the reader is referred to the following grammars: Göksel & Kerslake, 2005; Banguoğlu, 1974.

untenable as hypotheses about speakers of human languages in general". His counter-argument, based on the morphological complexity of Turkish, is very simple: The morphology of Turkish can generate 1,830,248 forms from a single verb root, and 9,192,472 forms from a single noun root. Assuming that an average educated speaker knows 20,000 noun roots and 10,000 verb roots, the full-listing hypothesis requires more than 200 billion entries to represent all possible word-forms. Based on a calculation of the total storage capacity of the human brain, Hankamer (1989) concludes that, storing the full-forms of so many words would only be possible if the brain was "dedicated to such storage and nothing else whatever"! Thus, in a language like Turkish, parsing must be involved in word recognition, "not just for rare or unfamiliar forms (unless one wants to call the majority of words occurring in ordinary text rare and unfamiliar)" (Hankamer, 1989, p. 403-404).

A serious problem with this argument is that it does not take into account the "sparsity" phenomenon discussed in Section 3.4.8. There may be millions of possible suffix combinations and billions of root-suffix combinations, but, as will be empirically demonstrated below, only a few thousand of them are used with any serious frequency. This means that those few thousand combinations that are used relatively frequently can very well have their full-form representations in the brain.

Hankamer (1989) also rejects any mechanism based on right-to-left parsing, which starts by stripping suffixes off the end of the word-form, to finally arrive at the root, and claims that the root must be recognized before suffixes are recognized. If parsing proceeds from right to left, "the set of stems determined by a suffix is always very large, and not necessarily even finite". If parsing proceeds from left to right, on the other hand, the number of possible suffixes that can be combined with the stem is finite, very small, and decreases at every step (Hankamer, 1989, p. 402). Based on the computational "wastefulness" of a right-to-left parsing algorithm, Hankamer (1989) concludes that morphological parsing must proceed from left to right, at least in an agglutinating language like Turkish.

In another important theoretical contribution, Frauenfelder & Schreuder (1992) discuss the issue of (morphological) productivity in word recognition, with frequent references to Hankamer (1989), and thus to the productivity of the morphology of Turkish. The paper summarizes several prominent models of word recognition, and discusses how they deal with the issue of productivity and the processing of word-forms that the speaker has never encountered before, and also how they explain the "morpheme frequency effects", which have been demonstrated many times in the literature.

According to Frauenfelder & Schreuder (1992), in an agglutinating language with a rich morphology and a transparent phonology, of which Turkish is a textbook example (Anderson, 1988, quoted in Frauenfelder & Schreuder, 1992), the speaker is constantly exposed to novel combinations that have never been encountered before. In such a language, the decomposition/parsing route (as opposed to the full-listing/direct route) would be expected to "win the race" in the processing of most morphologically complex word-forms, "because they are made up of morphemic combinations that occur rarely" (Frauenfelder & Schreuder, 1992, p. 180). The authors make an interesting suggestion at this point, which has been one of the inspirations behind the second experiment designed for this study:

> . . . it is possible that combinations of roots and affixes that a listener encounters frequently could get a separate access representation. Consequently, a single word form might be recognized through the cooperative efforts of the direct [i.e. full-listing] route and the parser [i.e. the decomposition route]. The *frequently co-occurring roots plus affixes* [emphasis added] would be recognized by the direct route, and the *rest of the*

> *word* [emphasis added] by the parser that combines the results of the direct route with the *remaining morphemes* [emphasis added] to be parsed. (Frauenfelder & Schreuder, 1992, p. 180)

The relevance of these two papers to the present study is clear: They specifically refer to the morphology of Turkish in their effort to “[constrain] psycholinguistic models of morphological processing and representation”. More importantly, they make testable predictions and invite researchers in the field to conduct “careful experimental research [to] resolve these questions” (Hankamer, 1989, p. 405).

The second experiment designed for the present study is an initial attempt to respond to this invitation. In fact, all word-forms used in Experiment 2 are “rare and unfamiliar” in the extreme: none of them occurs even once in a 283-million-word corpus. Thus, there cannot be any mental representation for these word-forms in the brains of most participants, at least before they finish the experiment. The experiment tests, among other things, how morphologically complex word-forms that have never been encountered before are processed by native speakers of Turkish. As a secondary issue, it also tries to understand if parsing proceeds from left to right as Hankamer (1989) suggests.

Having discussed two important theoretical papers, we now turn to empirical studies related to the mental processing of Turkish words:

Gürel (1999) tests the full-listing and decomposition hypotheses discussed in Hankamer (1989) and Frauenfelder & Schreuder (1992) using a lexical decision task involving morphologically simple and complex Turkish words. The two research questions Gürel addresses are (1) “to what extent does lexical access of multimorphemic words in Turkish involve morphological decomposition?”, and (2) "will all possible substrings of a word be parsed in word recognition?" (Gürel, 1999, p. 220).

The paper reaches the following conclusions:

> Not all multimorphemic words are accessed in a decomposed form in Turkish. Words with frequent suffixes seem to be accessed through a whole-word access procedure. Depending on the frequency of the suffix, a word can be accessed via the direct access route or the parsing route. (Gürel, 1999, p. 223)

However, the experiment presented in this paper involves certain methodological problems that might cast doubt on the validity of these conclusions:

(1) Although the experiment was conducted in 1999, word frequencies were obtained from a list of frequency counts published 39 years before the experiment (Pierce, 1960). Considering the speed with which language use can change over time[11], this puts into question the validity of the frequency values used in the study, and thus any conclusions that can be derived from it.

(2) According to Gürel (1999), the frequency counts in Pierce (1960) only contain the frequencies of bare roots, and the aggregate frequencies of suffixes regardless of where they occur. As a result, there is no way of knowing the surface frequencies of the complex word-forms used in the experiment. In other words, the independent variable of the experiment is in fact not the thing it purports to be (surface frequency of the complex form), but something different (total frequency of the relevant root).

[11] Since large-scale corpus studies showing changes in the frequencies of Turkish words over time are not available, at least to our knowledge, we will try to demonstrate how word usage can change over time using the Google nGram Viewer at books.google.com/ngrams and the two similarly-structured English words *accumulator* and *computer*: Between 1960 and 1999, the year Pierce collected the frequency counts and the year Gürel conducted the experiment, the frequency of the English word *accumulator* dropped from 1.8 pmw (per million words) to 0.7 pmw (a 61% decrease), while the frequency of the word *computer* increased from 19 pmw to 130 pmw (a 584% increase).

(3) As mentioned by Gürel, “nondecomposable words of three to four syllables in Turkish are generally of low frequency. Therefore matching these items with one-suffix words for frequency was not possible” (Gürel, 1999, p. 221). In other words, an important confounding variable was not controlled for.

(4) The experiment uses inflected nouns containing the locative, ablative and plural markers, and their combinations. Given that there are 98 suffixes in Turkish (see Appendix A), it is not clear why these three markers have been selected, and what unintended consequences this selection might have in terms of the phonology, orthography and semantics of the language.

(5) Word frequencies are not provided. Thus, it is impossible to know to what extent the “high-frequency” words and suffixes used in the experiment differ from the “low-frequency” words and suffixes in terms of frequency.

(6) The stimuli used in the experiment are also not provided. Hence, it is impossible to independently judge if stimulus selection was performed in a random and unbiased manner.

In a similar study conducted fourteen years after Gürel (1999), Gürel & Uygun (2013) discuss similar issues, this time with reference to “variability in the use of second-language morphology”. Since second-language learning is beyond the scope of this study, we will only briefly mention the results reported by the authors, and discuss methodological issues. Gürel & Uygun, (2013) reach the following conclusion, among others:

> It appears that in highly inflected languages like Turkish, for the sake of computational efficiency, complex forms are accessed via a direct [i.e. full-listing] route whenever possible. The extent of native-like performance of the advanced learners implies that full-listing can be accomplished after a certain degree of proficiency is attained in the [second language]. (Gürel & Uygun, 2013, p. 131)

The methodological problems mentioned for Gürel (1999) apply to this paper as well.

In a closely related article on “recognizing morphologically complex words in Turkish”, Durgunoğlu (2003) tests the left-to-right parsing hypothesis of Hankamer (1989), and more specifically its “counter-intuitive prediction that morphologically complex words should be recognized faster than matched simple words”. Applying word completion and word correction tasks to a group of children, the author observes that “morphologically [complex] forms . . . were completed or corrected just as accurately as morphologically simple forms. . .”, and concludes that “although some prefix stripping and searching the full list of morphemes may be a useful strategy in languages with more manageable morphological structures such as English, in the highly agglutinating language of Turkish, a left-to-right computational strategy seems to be the mode of operation” (Durgunoğlu, 2003, p. 89).

The paper involves some methodological problems: (a) all subjects in the experiment are children, meaning that the results cannot be generalized to the population of native speakers as a whole; and (b) the dependent variable of the experiment is the accuracy of a rather unorthodox paper-and-pencil word correction and completion task, meaning that the results cannot be compared to existing studies in the literature, and also cannot be used to decide whether parsing in Turkish proceeds from left to right or from right to left.

In an interesting article that inspired the second experiment designed for this study, along with Frauenfelder & Schreuder (1992), Durrant (2013) discusses whether the complex inflectional patterns of Turkish can be described as “formulaic”, where “formulaicity” refers to “the insight that some linguistic sequences which could potentially be analyzed into smaller units are, for one reason or another, better treated as wholes" (Durrant, 2013, p. 1). According to this view, any sequence of linguistic elements can become a formula "if it occurs so frequently that some form of independent storage in long-term memory is cognitively more efficient than creating the sequence from scratch each time it is needed" (Goldberg, 2006, quoted in Durrant, 2013). Durrant (2013) claims that, given the complexity of its morphology, inflected Turkish words are a good candidate for being processed in a formulaic manner.

The ideas presented in Durrant (2013) are closely related to "usage-based grammar", which treats grammar as a set of cognitive representations that emerge over time, from one's day-to-day experience with language. Bybee (2011) summarizes this view as follows:

> "The use of the same sounds, words, and patterns over thousands of usage-events has an impact on the cognitive storage and processing of linguistic experience that gives language its structure. As a result, then, linguistic structure is emergent from language use". (Bybee, 2011, p. 69)

To test the hypothesis that Turkish words are processed in a formulaic manner, Durrant (2013) compiles a "personal corpus" consisting of 765 texts totaling 374,590 words, taken from seven newspapers the author was personally exposed to between November 2009 and May 2010, and calculates the base frequencies of all words (i.e. the total frequency of all words that contain a given stem). He also calculates the frequencies of sequences made up of three to four morphemes, which he calls "morphemic bundles" (see the closely related concept of "template bundles" defined in Section 3.2.2.1).

In full agreement with Zipf (1949) and the results of the present study, Durrant (2013) discovers that there is a small number of very high-frequency complex words and morphemic bundles, and a large number of very low-frequency complex words and morphemic bundles. He suggests that this Zipfian distribution "fits comfortably within the dual-route processing model put forward for Finnish by Niemi et al. (1994)" (Durrant, 2013, p. 30). Based on these findings, Durrant (2013) concludes that "particular repeated patterns . . . might in some cases emerge as independently-represented entities, but would most likely exist within networks of associations with other morphological patterns . . . and with particular lexical roots. . ." (Durrant, 2013, p. 31).

Although Durrant (2013) is an "exploratory" survey that does not empirically test if morphemic bundles have psychological reality, it is in our view an important first step towards understanding the processing of complex words in Turkish and similar languages.

Finally, for the sake of completeness, let us mention, in passing, a few studies that are only indirectly related to the issues discussed in the present study:

Kırkıcı & Clahsen (2013) report a series of experiments investigating the morphological processing of inflectional and derivational words in native and non-native speakers of English, German and Turkish, and reach the following conclusion:

> "Adult native speakers . . . demonstrated efficient morphological priming effects for regularly inflected word forms [but that] this was not the case for L2 learners of these languages. For derivational processes, on the other hand, . . . [native] speakers showed the same significant morphological priming effects for productive derivational processes. . . Unlike for regular inflection, however, derived word forms also yielded significant masked priming effects for L2 learners. . ." (Kırkıcı & Clahsen, 2013, p. 787)

In a master's thesis presented to the cognitive science department of Middle East Technical University, Özer (2010) uses a picture naming task to investigate morphological priming effects in three types of Turkish nominal compounds. She finds "clear evidence for morphological priming effects, which are distinct from phonological effects and comparatively stronger”. According to Özer, this result supports processing models based on decomposition.

In a master’s thesis presented to the same department, Eren (2014) examines the effect of stem frequency and stem length on eye movement-parameters while reading (a) uninflected stems (e.g. *istasyon* ‘station’), (b) words with the single inflectional suffix *-lAr* (e.g. *istasyonlar* ‘stations’), and (c) words with a longer sequence of inflectional suffixes (e.g. *istasyonlardakilerden* ‘from those at the stations’). The experiments show that both the frequency and length of the uninflected stem influences gaze durations. More specifically, first-pass gaze duration and total gaze duration were longer for the long and low-frequency words than the short and high-frequency words. However, first-fixation duration in words with inflectional suffixes was shorter in longer words than in shorter words” (Eren, 2014, p. iv).

Finally, in a master’s thesis presented to the department of English Language Teaching of Middle East Technical University in the same year, Gacan (2014) examines how native speakers of Turkish process derivationally complex word-forms in their native language and in English as a second language. The masked priming experiments conducted as part of this study use the transparent, frequent and productive attributive suffix *-lI* and the privative suffix *-sIz,* as well as the corresponding English suffixes *-ful* and *-less*. Gacan (2014) reports “similar priming effects for L1 Turkish and the high proficiency L2 English group” and concludes that these findings support decompositional models of word recognition in the native and second-language morphological processing of derived words.

This concludes the review of relevant literature. Several important aspects of word recognition that have been widely discussed in the literature have not been included here, since they are only indirectly related to the objectives of the present study. These include the purely visual processing of the letters that make up a word (orthographic processing), the timing of the sub-processes involved in visual word recognition and the interactions between them (temporal sequence), the question whether morphological parsing is obligatory or optional, and the question if morphology is a phenomenon with a real neurological basis or simply an epiphenomenon that emerges from the interaction between the orthographic and semantic levels.

CHAPTER 3
PREPARATORY WORK

An unusual amount of preparatory work had to be done in order to design and implement the two lexical decision experiments reported in this study. This is mainly because the quantitative resources required for properly selecting the experimental stimuli and defining and quantifying independent variables and control variables were largely absent. Moreover, the methodological problems observed in existing literature have indicated that strict design decisions must be adopted in advance. Finally, the terminological problems described in Section 2.2.2 and the inadequacy of existing frequency definitions for an agglutinating language like Turkish have urged the development of a general notation for defining various measures of frequency. This chapter describes the steps of this large-scale preparatory work.

3.1 Design decisions

Four experimental design decisions have been made in advance with a view to maximizing the validity and reliability of results.

3.1.1 Principled stimulus selection

As pointed out by Forster (2000) and Baayen (2014), many experimental studies fail to select their experimental stimuli in a random and unbiased manner. Researchers often select, or refrain from selecting, experimental stimuli based on their own knowledge of the language, and intuitions about which stimuli would work, and which would not. In Baayen's words, "the consequences of non-random stimulus selection is, from a statistical perspective, disastrous" (Baayen, 2014, p. 4). In a similar vein, Clark (1973) advises researchers to "sample language by systematic, repeatable procedures" (Clark, 1973, p. 350).

This is why the present study makes a considerable effort to select stimuli in a principled way, based purely on mathematical descriptions, i.e. without human intervention. In fact, most of the work described in the subsequent sections of this chapter was performed with this purpose in mind.

3.1.2 Avoiding the "language-as-a-fixed-effect fallacy"

A scientific experiment would have very little point if the conclusions it reaches are limited only to the specific set of stimuli and/or to the specific individuals who participate in that experiment (Coleman, 1964). In other words, experimental results obtained from the necessarily small sample of stimuli and subjects used in the study should be generalizable to the population of all possible subjects on the one hand, and to the set of all possible stimuli on the other. Coleman (1964), however, notes that the actual situation is far from this ideal: "There is little statistical evidence that such studies could be successfully replicated if a different sample of language materials were used" (Coleman, 1964, p. 219). In a widely cited article, Clark (1973) defines the "language-as-a-fixed-effect fallacy" as the unfounded assumption that the findings obtained from the small sample of linguistic materials used in a given experiment are automatically true for language in general (also see Raaijmakers, Schrijnemakers, & Gremmen, 1999).

To reduce the potential impact of the "language-as-a-fixed-effect fallacy", multiple stimulus sets have been used in the experiments discussed in Chapter 4, using a fully-automated stimulus-selection algorithm specially designed for this study (see Section 4.2.4 for details).

3.1.3 Avoiding laboratory experiments

Behavioral / psycholinguistic experiments are generally conducted in controlled laboratory environments. This offers several advantages: (a) By using the same equipment for all subjects, the experimenter can be sure that the stimuli are presented, and responses recorded, in a uniform manner; (b) personal and demographic data collected from the subjects is reasonably reliable; (c) instructions can be given quite efficiently because subjects can ask questions if they are unsure about certain details of the task; (d) participant behavior and environmental parameters can be monitored closely; (e) preventing the same person from participating in the same experiment more than once is relatively easy.

However, the laboratory method also involves certain issues: (a) Subjects are usually unmotivated because the typical test subject is a student who is offered a nominal sum of money, extra course credit, and in many cases nothing, for participating in the experiment; (b) subjects are not naive: since the typical subject is a current student of one or more professors involved in the study, he/she has some idea about what the experiment might be trying to test; (c) subjects are not representative of native speakers as a whole: most subjects are undergrad students in their early twenties and thus represent a very specific sub-group with respect to age, sex, income, education, vocabulary size, linguistic preferences, cognitive skills, etc.; (d) sample size is small: the typical number of subjects in a laboratory experiment is between 20-70.

A second and increasingly popular option is to conduct the experiment online. An online experiment reduces the severity of the above-listed issues to varying degrees: (a) subjects participate in the experiment voluntarily and are thus more motivated than undergrad students forced to participate in the experiment; (b) since online advertisements of the experiment can reach anyone with an internet connection, subjects tend to come from all sections of society, and are thus both more representative of the typical native-speaker and are also more naïve as to the objectives of the experiment; (c) sample size is larger by several orders of magnitude. Online experiments also have the following added benefits: (a) Total cost is lower (no logistics, no equipment); (b) data collection is almost instantaneous.

On the other hand, online experiments have their own drawbacks: (a) The experimenter has no control over the equipment used by the subject during the experiment, and also no control over the environment / setting where the subject takes the experiment; (b) it is impossible to make sure that the subject has properly understood the instructions; (c) it is impossible to verify that the subject has provided correct demographic and personal data; (d) it is much more difficult than in a laboratory experiment to prevent the same person from participating in the experiment more than once.

Even if we assume that the benefits outweigh the drawbacks, online experiments are not a cure-all solution. The specific nature of the experimental task must be considered: If the experiment is designed to measure response time differences between subjects, the equipment issue would have disastrous consequences. The present study, however, is interested in within-subjects differences rather than between-subjects differences in response time. In other words, the two experimental conditions (high-frequency words vs. low-frequency words) are not administered to two separate groups; every subject is exposed to all (two) conditions, and thus every subject serves as its own control. This is why an online experiment can be conducted for the purposes of this study.

3.1.4 Quantifying anything that can be quantified

As mentioned earlier, there exists a large number of variables that can have an effect on response times in psycholinguistic experiments. Moreover, as mentioned in Section 2.2.1, the properties of a single word can be quantified at a surprisingly large number of levels. With these observations in mind, the last design decision is to try to quantify anything that can be quantified about Turkish words, and to use the resulting variables in the experiments.

The tree model discussed in Section 3.2.1, the generalized frequency notation proposed in Section 3.2.2, and the various metrics defined in Section 3.2.4 are the result of this effort to quantify and control. To the best of our knowledge, the values of these parameters have been estimated for the first time for Turkish. It is hoped that these general-purpose linguistic resources will both improve the validity and reliability of the results reported in the present study, and will also be useful in future studies in the field. The sections that follow offer a detailed description of the steps followed for the development of these resources.

3.2 Notation, definitions and models

This section summarizes the formal notation, definitions and mathematical models developed in the course of this study. The ideas proposed in this section were mainly borne of the effort to define independent variables and control for confounding variables during the experiments, but also go beyond that purpose. Although only a small part of the ideas proposed here have been used in the actual experiments, we think that this quantification and modelling effort will be useful for future work in the field.

3.2.1 The suffix tree and suffix sequences

In an agglutinating language, affixes accumulate on the root one after the other, as briefly demonstrated in Section 2.4. This "one-affix-at-a-time" mechanism offers the possibility to represent all possible affix sequences of the language in the form of a single tree, which will be referred to as the "suffix tree".[12] The suffix tree is the central modelling tool that will be used to describe, represent, and measure morphological phenomena in this study.

Figure 2 reproduces a small subset of the suffix tree for Turkish nouns, based on actual corpus data. The large black node at the top-left corner represents the root (in this case a bare noun), and each of the other nodes represents one of the suffixes listed in Appendix A.

Node color indicates suffix type (red: derivational suffix, green: plural marker, orange: compound marker, blue: inflectional suffix, yellow: non-terminal). The yellow nodes do not represent actual word-forms. They represent incomplete/intermediate verb forms that cannot stand alone because they do not contain the obligatory person, tense, aspect or modality markers, and have been added to the tree to make sure that the root can be reached from any given node.[13] All nodes except non-terminal nodes represent actual word-forms attested in the corpus (for a discussion of grammatically possible word-forms that are not attested in the BOUN Corpus , see Section 3.4.8).

---

[12] Since prefixation has a very limited range of application in Turkish (Göksel & Kerslake, 2005, p. 49) and infixation is non-existent, the affix tree will be called the "suffix tree" for simplicity.

[13] In fact, these incomplete/intermediate forms are orthographically identical to the second-person singular of the imperative, but we regard this as a superficial coincidence: Treating these forms as second-person imperative forms would not make sense since the child nodes of these forms do not have to be second-person and/or imperative. For example, if we treat the form *yaptır* as a second-person singular imperative (*Verb+Caus+Imp+A2sg*), the "child" *yaptırdık* would be expected to inherit this second-person singular imperative property, but it does not.

Moving from a given node to one of its children is equivalent to adding a suffix. For example, moving from the root node *R* to Node *15*, and then to Node *52* creates the sequence *Noun+Agt+Ness*, which is a highly productive pattern that creates profession names like *arabacılık* 'profession of a car-maker', *demircilik* 'profession of an ironsmith', *koruculuk* 'profession of a ranger', and *sütçülük* 'profession of a milkman'. The gray arrows in Figure 2 point to the nodes involved in the step-by-step creation of the template *R+15+52+04+60+09*, using the example *temsil+ci+lik+ler+imiz+i*.

The suffix sequence that remains when the root is removed from the word-form is a central concept in this study, and will be referred to as the "suffix template". Roots combine with suffix templates to create word-forms. In the above examples, for instance, the nominal roots *araba* 'car', *demir* 'iron', *koru* 'small forest' and *süt* 'milk' combine with the suffix template *Agt+Ness* to form stems like *arabacılık*, *demircilik*, *koruculuk* and *sütçülük*. There are more than 4,000 such stem types in the BOUN Corpus, and all of these are represented by the single red node denoted '*52*' in the center of Figure 2. Finally, node size is proportional to the total (logarithmic) frequency of all word-forms that use the suffix template in question.

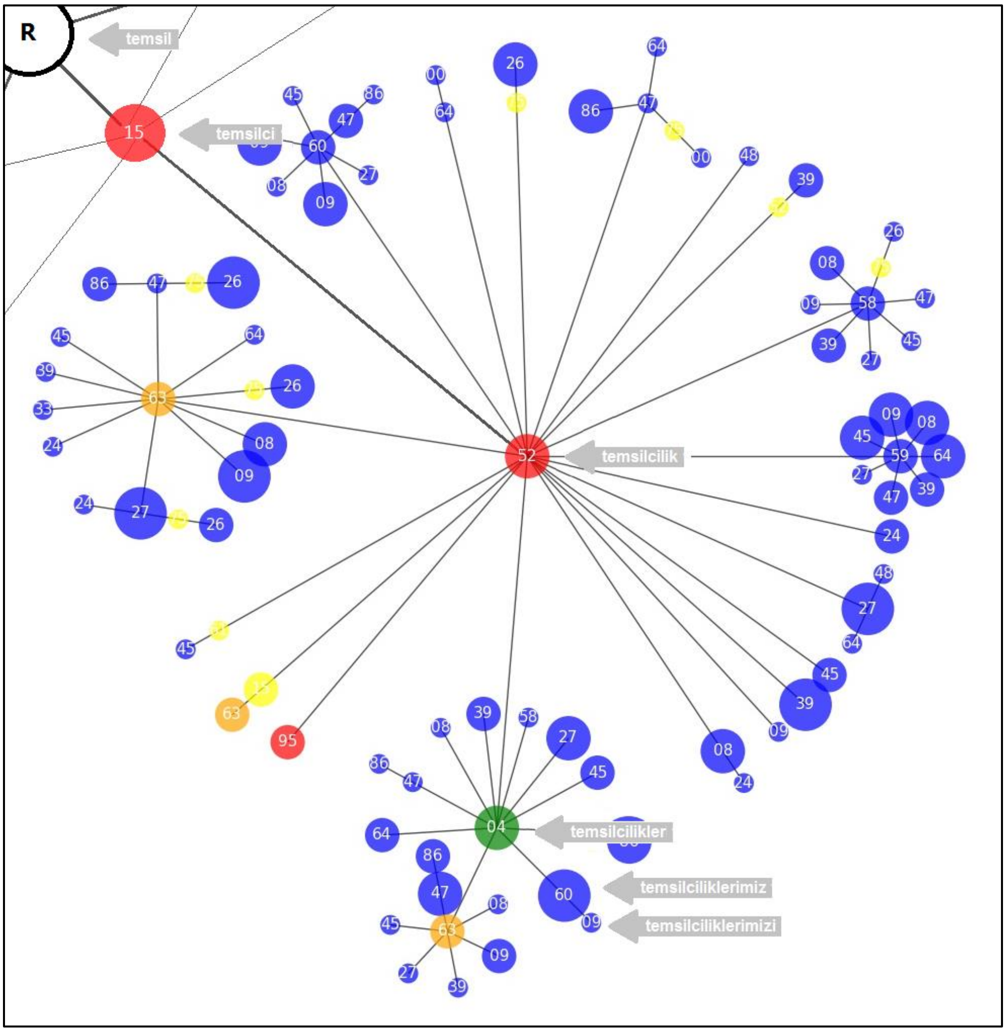

Figure 2. A suffix tree representing the accumulation of suffixes on a root

The tree model allows us to define suffix templates purely in mathematical terms. Before moving on to the categorization of suffix templates, note that a suffix template is not a single node but a *path* on the tree. More specifically, a suffix template is a path that (a) does not include the root node, (b) starts at an immediate child of the root node, and (c) ends at a terminal node. This being a tree structure, the relevant path automatically becomes uniquely identified when the name of its end-node is specified. Below are some suffix template examples based on Figure 2:

* *R-15-52* (*Noun+Agt+Ness*) is not a suffix template because it starts at *R*.

* *15-52* (*Agt+Ness*) is a suffix template.

* *15-52-75* (*Agt+Ness+Pres*) is not a suffix template because it ends at non-terminal node *75* (*Pres*).

* *15-52-04-60-09* (*Agt+Ness+A3pl+P1pl+Acc*) is a suffix template.

* *86* (*Rel*) is not a suffix template because it does not start at an immediate child of *R*.

Now, let us try to categorize suffix templates based on the types of suffixes (i.e. tree nodes) they contain:

T_{inf} : Inflectional template. Suffix sequence that produces a purely inflectional form. Cannot contain any derivational suffixes or the compound marker *-(s)I* (visually: any path without red or orange nodes). Examples: *+dA+ki+ler+in, +lAr+ImIz+lA, +ler+in+ki+ler+den*

T_{drv} : Derivational template. Suffix sequence that produces a derivational form. The first suffix after the root must be a derivational suffix, which may or may not be followed by one or more inflectional suffixes and/or the plural marker *–lAr* and/or the compound marker *-(s)I* (visually: any path that starts at a red node). Examples: *+CI, +lIklArdAki, +CI+(s)I+(n)dA+ki+lAr, +lAş+tIr+mA+mA+lI*

T_{CM} : Compound template. Suffix sequence that produces a compound form. The compound marker *-(s)I* may occur anywhere within the template. Cannot contain any derivational suffix (visually: any path that contains an orange node but no red nodes). Examples: *+(s)I+(n)dA+ki+lAr, +CI+lIk+(s)I+mIz, +lAş+Abil+mA+(s)I*

For the sake of completeness, let us also define the empty template $T_{Ø}$, which is assumed to be attached to bare roots, although it does not have any surface representation.

3.2.2 Defining frequency

In the context of exposure to linguistic events, frequency is a surprisingly complex concept. What types of frequency, if any, does a human brain track when exposed to complex forms like *sinema oyunculuğundaki* ‘the one in film acting’? Several things can be measured here: the frequency of the first word-form (*sinema*), the second word-form (*oyunculuğundaki*), the second word-form's root or stem(s) (*oyun*, *oyuncu*, *oyunculuk*), the phrase as a whole (*sinema oyunculuğundaki*), individual inflectional suffixes (*-(n)DA*, *-ki*), individual derivational suffixes (*-CI*, *-lIk*), the suffix template as a whole (*+CI+lIk+(s)I+(n)+DA+ki*), etc.

Considering the terminological problems discussed in Section 2.2.2 of the literature review, a generalized formal notation will be proposed here, and equivalent terms used in the literature, if any, will be indicated in footnotes. To be able to systematically define frequency as exhaustively as possible, the following section uses the theory-neutral concept “bundle”.

3.2.2.1 Bundles

For the purposes of this study, a “word-form bundle” is defined as the set of all word-forms that are related to each other in some systematic way. For example, the word-forms *gözü, göze, gözde, gözden, gözün, gözler, gözle* constitute a meaningful bundle in that they are all formed by adding a single inflectional suffix to the root *göz* ‘eye’[14]. Another meaningful bundle of the same root would be *gözlem*, *gözlemci*, *gözlemcilik*, *gözlük*, *gözlükçü*, *gözlükçülük*, etc., the set of all word-forms that can be obtained by adding any number of derivational suffixes to the root *göz*.

[14] Although this is closely related to the concept “linguistic paradigm”, the term “paradigm” will be avoided here, since the definitions in this section go well beyond the traditional usage of that term in linguistics literature. To avoid overgeneralizing an existing term, the rather neutral term “bundle” will be used.

Similarly, a "template bundle" is defined as the set of all suffix templates that are related to each other in some systematic way. For example the suffix templates *+Ness+Acc* (*+lIk+I*), *+Ness+Dat* (*+lIk+A*), *+Ness+Loc* (*+lIk+DA*), *+Ness+Abl* (*+lIk+DAn*), *+Ness+Gen* (*+lIk+(n)In*), *+Ness+A3pl* (*+lIk+lAr*) constitute a meaningful bundle in that they are all formed by adding a single inflectional suffix to the derivational template *+Ness*. When combined with the root *tüccar* 'merchant', for instance, these bundle members would form the words *tüccarlığı*, *tüccarlığa*, *tüccarlıkta*, *tüccarlıktan*, *tüccarlığın*, and *tüccarlıklar*. Another meaningful bundle of the derivational template *+Ness* (*+lIk*) would be *+Ness+Agt* (*+lIk+CI*), *+Ness+With* (*+lIk+lI*), *+Ness+Without* (*+lIk+sIz*), *+Ness+Agt+Ness* (*+lIk+CI+lIk*), the set of all suffix templates that can be obtained by adding at most two derivational suffixes to the suffix template *+Ness*. Once again, when combined with the root *tüccar* 'merchant', these bundle members would form the (hardly interpretable but possible) words *tüccarlıkçı*, *tüccarlıklı*, *tüccarlıksız*, and *tüccarlıkçılık*.

(a) Inflectional bundles

For the purposes of this study, the "inflectional bundle" $B_{inf:n}(S)$ of a stem S is defined as all word-forms obtained by adding to S any inflectional template $T_{inf:n}$, where n refers to the maximum number of inflectional suffixes allowed to be added to the stem, i.e. the maximum allowed length of the inflectional template. Additionally, $B_{inf:max}(S)$ denotes the maximal inflectional bundle where n is allowed to be as large as the morphological rules of the language permit.[15]

A second distinction will be made based on whether membership in the bundle requires corpus frequency to be non-zero. If word-forms that are not attested

[15] Note that the traditional definition of an inflectional paradigm in linguistics literature is equivalent to $B_{inf:1}(S)$.

in a certain corpus are excluded, a smaller bundle will be obtained, which will be referred to as the “attested inflectional bundle” and denoted by $B'_{inf:n}(S)$.

Finally, a third distinction will be made based on whether or not the stem itself is included in the bundle. If the stem is included, the resulting bundle will be referred to as the “extended inflectional bundle” and denoted by $B^*_{inf:n}(S)$.

The following made-up example using the stem *göz* ‘eye’ illustrates some of the inflectional bundles resulting from these definitions.[16] On each row, the members of the relevant bundle are marked ‘x’:

Table 2. Inflectional Bundles of the Stem göz ‘eye’

Form	*göz*	*~ü*	*~ün*	*~de*	*~dedir*	*~ünki*	*~ünkiler*	*~ünkileri*	…
Attested?	yes	yes	yes	no	yes	yes	no	yes	…
$B_{inf:1}(S)$		x	x	x					…
$B'_{inf:1}(S)$		x	x						…
$B^*_{inf:1}(S)$	x	x	x	x					…
$B^{*\prime}_{inf:1}(S)$	x	x	x						…
$B_{inf:2}(S)$		x	x	x	x	x			…
$B'_{inf:2}(S)$		x	x		x	x			…
$B^*_{inf:2}(S)$	x	x	x	x	x	x			…
$B^{*\prime}_{inf:2}(S)$	x	x	x		x	x			…
...	…	…	…	…	…	…	…	…	…
$B_{inf:max}(S)$		x	x	x	x	x	x	x	…
$B'_{inf:max}(S)$		x	x		x	x		x	…
$B^*_{inf:max}(S)$	x	x	x	x	x	x	x	x	…
$B^{*\prime}_{inf:max}(S)$	x	x	x		x	x		x	…

[16] The total frequency of all members of the inflectional bundle is sometimes referred to as “base frequency” in the literature. See Section 2.2.2.2 for details.

This notation is unnecessarily complicated for analytic languages like English. In fact, applying the above definitions to the English stem ‘eye’, we obtain Table 3.

Table 3. Inflectional Bundles of the Stem ‘eye’

Form	*eye*	*~s*	*~'s*	*~s'*
Attested?	yes	yes	yes	yes
$B_{inf:1}(S)$		x	x	
$B'_{inf:1}(S)$		x	x	
$B^*_{inf:1}(S)$	x	x	x	
$B^{*\prime}_{inf:1}(S)$	x	x	x	
$B_{inf:2}(S)$		x	x	x
$B'_{inf:2}(S)$		x	x	x
$B^*_{inf:2}(S)$	x	x	x	x
$B^{*\prime}_{inf:2}(S)$	x	x	x	x
$B_{inf:max}(S)$		x	x	x
$B'_{inf:max}(S)$		x	x	x
$B^*_{inf:max}(S)$	x	x	x	x
$B^{*\prime}_{inf:max}(S)$	x	x	x	x

As can be seen in Table 3, there are only three different sets of word-forms: The largest, four-member bundle contains (1) the nominative, (2) the plural, (3) the possessive, and (4) the plural-possessive; the three-member bundle contains (2) and (3) and (4), while the two-member bundle only contains (2) and (3). The template-length parameter *n* has very little effect, because a noun can take at most two inflectional suffixes in English.

Also note that all four forms have been marked as “attested”. This is because there are only four inflectional forms in English nominal morphotactics, and a moderately large corpus would probably contain all of them at least once. In other words, the sparsity phenomenon described in Chapter 3.4.8 below does not exist, or is much weaker, in languages like English, at least in the case of inflection. But in Turkish:

(a) a single stem can take as many as seven consecutive inflectional suffixes, as in the made-up example *ev-ler-in-de-ki-ler-imiz-den*, which gives rise to seven possible bundle lengths;

(b) a single inflectional bundle can have tens of thousands of members, which gives rise to the sparsity phenomenon, where thousands of grammatically possible forms do not occur in a rather large corpus even once, thus necessitating the distinction between attested and unattested bundle members.[17]

(b) Derivational bundles

Applying the same idea to derivational suffixes, the "derivational bundle" $B_{drv:n}(S)$ of a stem S is defined as all word-forms obtained by adding to S any derivational template $T_{drv:n}$, where all of the n nodes on the path between S and T_{drv} are derivational. As before, $B_{drv:max}(S)$ denotes the maximal derivational bundle where n is allowed to be as large as the morphotactics of the language permits.[18]

The second distinction is based on whether membership in the bundle requires corpus frequency to be non-zero. If unattested word-forms are excluded, a smaller bundle is obtained, which will be referred to as "attested derivational bundle" and denoted by $B'_{drv:n}(S)$.

[17] This is the distinction Hankamer (1989) ignores in his treatment of the productivity of Turkish morphology.
[18] The total frequency of all members of the derivational bundle is known as "family frequency" in the literature. See Section 2.2.2.3 for details.

As before, the third distinction is based on whether or not the stem itself is included in the bundle. If the stem is included, the resulting bundle will be referred to as "extended derivational bundle" and denoted by $B^*_{drv:n}(S)$.

The made-up example in Table 4 using the stem *göz* 'eye' illustrates some of the derivational bundles resulting from the above definitions.

Table 4. Derivational Bundles of the stem göz 'eye'

Form	*göz*	*~lük*	*~cü*	*~lükçü*	*~lükçülük*	*~lemcisizlik*	…
Attested?	yes	yes	no	yes	no	yes	…
$B_{drv:1}(S)$		x	x				…
$B'_{drv:1}(S)$		x					…
$B^*_{drv:1}(S)$	x	x	x				…
$B^{*\prime}_{drv:1}(S)$	x	x					…
$B_{drv:2}(S)$		x	x	x			…
$B'_{drv:2}(S)$		x		x			…
$B^*_{drv:2}(S)$	x	x	x	x			…
$B^{*\prime}_{drv:2}(S)$	x	x		x			…
...	…	…	…	…	…	…	…
$B_{drv:max}(S)$		x	x	x	x	x	…
$B'_{drv:max}(S)$		x		x		x	…
$B^*_{drv:max}(S)$	x	x	x	x	x	x	…
$B^{*\prime}_{drv:max}(S)$	x	x		x		x	…

Unlike the situation in inflectional bundles, derivational bundles defined in this complicated manner can be meaningful also for languages like English. This is because, unlike its inflectional morphology, the derivational morphology of English is quite complex, where a stem can take up to five affixes, as in the example *deinstitutionalization*.[19]

(c) Template bundles

The inflectional and derivational bundles defined above are sets of word-forms. However, it is also possible and meaningful to define bundles that package suffix templates instead of word-forms. Simply by replacing the stem *S* in the above definitions with a suffix template *T*, we obtain two additional sets of bundles, shown in Table 5 and Table 6. Note that the column headings now use the allomorph notation with capital letters, because they now represent all possible allomorphs (surface forms) of the relevant suffix as attached to any stem, rather than a particular allomorph of the relevant suffix as attached to a particular stem:

Table 5. Inflectional Template Bundles

Form	*Ø*	*-(y)I*	*-DA*	*-DAdIr*	*-(n)In*	*-(n)Inki*	*-(n)InkilAr*	*-(n)InkilArI*	…
Attested?	yes	yes	no	yes	yes	yes	no	yes	…
$B_{inf:1}(T)$		x	x		x				…
$B'_{inf:1}(T)$		x			x				…
$B^*_{inf:1}(T)$	x	x	x		x				…
$B^{*\prime}_{inf:1}(T)$	x	x			x				…
$B_{inf:2}(T)$		x	x	x	x	x			…
$B'_{inf:2}(T)$		x		x	x	x			…
$B^*_{inf:2}(T)$	x	x	x	x	x	x			…
$B^{*\prime}_{inf:2}(T)$	x	x		x	x	x			…
…	…	…	…	…	…	…	…	…	…
$B_{inf:max}(T)$		x	x	x	x	x	x	x	…
$B'_{inf:max}(T)$		x		x	x	x		x	…
$B^*_{inf:max}(T)$	x	x	x	x	x	x	x	x	…
$B^{*\prime}_{inf:max}(T)$	x	x		x	x	x		x	…

[19] However, as this example demonstrates, English uses both prefixes and suffixes, and this might justify the introduction of separate bundle types for prefixed and suffixed word-forms. Since this study uses Turkish data, this will not be attempted here.

Table 6. Derivational Template Bundles

Form	Ø	*-lIk*	*-CI*	*-lIkCI*	*-lIkçIlIk*	*-lAmcIsIzlIk*	…
Attested?	yes	yes	no	yes	no	yes	…
$B_{drv:1}(T)$		x	x				…
$B'_{drv:1}(T)$		x					…
$B^*_{drv:1}(T)$	x	x	x				…
$B^{*\prime}_{drv:1}(T)$	x	x					…
$B_{drv:2}(T)$		x	x	x			…
$B'_{drv:2}(T)$		x		x			…
$B^*_{drv:2}(T)$	x	x	x	x			…
$B^{*\prime}_{drv:2}(T)$	x	x		x			…
…	…	…	…	…	…	…	…
$B_{drv:max}(T)$		x	x	x	x	x	…
$B'_{drv:max}(T)$		x		x		x	…
$B^*_{drv:max}(T)$	x	x	x	x	x	x	…
$B^{*\prime}_{drv:max}(T)$	x	x		x		x	…

Assuming that at most seven inflectional suffixes and at most four derivational suffixes can be attached to Turkish bare nouns, ((4 x 7) + (4 x 4)) x 2 = 88 different frequency metrics can be defined using the above bundle notation. Since it is impractical to exemplify them one-by-one, six arbitrary examples are given below, using the stem *göz* and the derivational template *Agt*:

$f(P_{inf:1}$(‘göz’)): Total frequency of all words formed by adding a single inflectional suffix to the stem *göz*. This includes, but is not limited to, the frequencies of *gözü, göze, gözde, gözden, gözün, gözler, gözle*…

$f(P^*_{inf:max}$(‘göz’)): Total frequency of all words formed by adding any number of inflectional suffixes to the stem *göz*, plus the frequency of the stem *göz* itself. This includes, but is not limited to, the frequencies of *göz, gözü, gözlerimiz, gözlerimizin, gözlerimizdekiler…*

$f(B_{drv:1}$(‘göz’)): Total frequency of all words formed by adding a single derivational suffix to the stem *göz*. This includes, but is not limited to, the frequencies of *gözlük*, *gözcü*, and *gözlem*.

$f(B^*_{drv:max}$('göz')): Total frequency of all words formed by adding any number of derivational suffixes to the stem *göz,* plus the frequency of the stem *göz* itself. This includes, but is not limited to, the frequencies of *göz*, *gözlük*, *gözcü*, *gözlem,* *gözlükçü*, *gözlemci*, *gözlükçülük,* and *gözlemcisizlik.*

$f(B_{inf:max}$('Agt')): Total frequency of all suffix templates formed by adding any number of inflectional templates to the suffix template *Agt*. This includes, but is not limited to, the frequencies of *Agt+Gen*, *Agt+Gen+Rel*, *Agt+Gen+Rel+A3pl+Acc* …

$f(B^*_{drv:1}$('Agt')): Total frequency of all suffix templates formed by adding a single derivational suffix to the suffix template *Agt*, plus the frequency of the suffix template *Agt* itself. This includes, but is not limited to, the frequencies of *Agt*, *Agt+Ness*, *Agt+With*, *Agt+Without*…

3.2.3 Does this make sense?

When a human brain is exposed to a visual linguistic event (i.e. printed or handwritten word or phrase)[20], several of the above-defined frequency values would potentially increase by one. But does the brain really update tens of different frequency values when exposed to a single complex expression?

It would require extensive sets of experiments to see if all (or any) of these proposed frequency measures have "psychological reality". We have used only a few of them in the two experiments described in Chapter 4 below. Still, it is clear that frequency is far from being a straightforward concept, and expressions like "the frequency of the word X" are ambiguous at best.

Since it is largely based on English and other morphologically impoverished languages, existing literature does not treat frequency at this level of detail. But a

[20] Note that we only focus on "visual lexical events", and ignore phonology, semantics and syntax almost completely, for which separate sets of frequency measures can be defined.

systematic treatment with a special emphasis on morphology is indispensable in the case of an agglutinating language.

3.2.4 Quantifying suffix templates

Combined with some metrics defined with the help of the suffix tree described in Section 3.2.1. and some additional metrics that will be defined in this section, a suffix template will be represented by the seventeen structural parameters below. In addition to offering a systematic way of describing suffix templates, these parameters will also allow us to control for several potential confounding variables while designing the experiments:

(a) Template frequency: Number of times a suffix template occurs in the corpus. We refer to the corresponding node on the tree as the “focus node”.

(b) Parent count: Given a focus node, this is the number of parents, excluding any non-terminal nodes.

(c) Parent frequency: Total frequency of parent nodes.

(d) Child count: Given a focus node, this is the number of children, excluding any non-terminal nodes.

(e) Child frequency: Total frequency of child nodes.

(f) Sibling count: Given a focus node, this is the number of its siblings on the tree, excluding any non-terminal nodes.

(g) Sibling frequency: Total frequency of sibling nodes.

(h) Unigram count: Number of suffixes (suffix unigrams) in the template. For example, the template *+Ness+Agt+A3pl* contains the suffix unigrams *Ness*, *Agt*, and *A3pl*, and its unigram count is therefore equal to three.

(i) Mean unigram Frequency: Arithmetic mean of the frequencies of all suffixes contained in the template. For example, this is equal to the arithmetic mean of the three frequencies corresponding to the three suffixes listed in (i) above.

(j) Bigram count: Number of suffix bigrams, after adding two dummies to mark the beginning (‘#’) and end (‘|’) of the suffix template. For example the template +*Ness*+*Agt*+*A3pl* contains the following suffix bigrams: {(*#*, *Ness*), (*Ness*, *Agt*), (*Agt*, *A3pl*), (*A3pl*, |)}.

(k) Mean bigram frequency: Arithmetic mean of the frequencies of all suffix bigrams contained in the template. For example, the arithmetic mean of the four frequencies corresponding to the four bigrams listed in (j) above.

(l) Trigram count: Number of suffix trigrams, after adding two dummies to mark the beginning (‘#) and end (‘|) of the suffix template. For example the template +*Ness*+*Agt*+*A3pl* contains the following suffix trigrams: {(*#*, *Ness*, *Agt*), (*Ness*, *Agt*, *A3pl*), (*Agt*, *A3pl*, |)}.

(m) Mean trigram Frequency: Arithmetic mean of the frequencies of all suffix trigrams contained in the template. For example, the arithmetic mean of the three frequencies corresponding to the three trigrams listed in (l) above.

(n) Inflectional suffix count: Number of inflectional suffixes in the suffix template. Please refer to Appendix A for a list of inflectional suffixes.

(o) Derivational suffix count: Number of derivational suffixes in the suffix template. Please refer to Appendix A for a list of derivational suffixes.

(p) Existence of the compound marker -*(s)I*: This is a Boolean parameter that is set to 1 if the suffix template contains the compound marker -*(s)I*, and to 0 otherwise (see Section 3.4.6 for a discussion on why the compound marker -(*s)I* requires special treatment).

(q) Blocking: The blocking parameter specifies the position where the phonological/orthographic blocking described in Section 2.4 occurs: Zero means there is no blocking, two means the second suffix is an A-type suffix, etc.

(r) Orthographic length: Number of characters in the word-form. For example, the orthographic length of the word-form *gör* 'see' is three, while the orthographic length of *görüldü* '(it) has been seen' is seven.

3.3 Measuring frequency

Having defined various frequency metrics and structural parameters, this section estimates their actual values for individual words, phrases, morphemes and suffix templates.

3.3.1 Objective vs. subjective frequency

The frequency of a linguistic event can be measured objectively or subjectively (Howes, 1954; Gernsbacher, 1984; Balota, Pilotti, & Cortese, 2001). "Objective frequency" refers to the number of times a linguistic event occurs in a certain (usually large-scale and electronic) corpus.

Frequency is not an intrinsic property of a word. "It cannot be measured directly on the word . . . but can be determined by counting occurrences of the word in a finite specimen of text" (Popescu, 2009, p. 1). It is impossible to estimate the true population value of a word's frequency "because there are no true populations in language" (Popescu, 2009, p. 1). Hence, expressions like "the frequency of the word *W* is *n*" are meaningless; all that can be asserted is that "the frequency of the word *W* in corpus *C* is *n*". In other words, frequency is a strictly corpus-dependent, extrinsic measure.

Despite the above-mentioned limitation, frequency counts obtained from corpora are assumed to provide an objective estimate of an average person's exposure to a linguistic event. “Subjective frequency”, on the other hand, is measured by asking people how familiar a given lexical item is, and by averaging the self-reported values. Subjective measures of frequency aim to overcome the sampling bias usually involved in corpus counts, but are harder to obtain and may be contaminated by certain other, unintended variables (Balota, Pilotti, & Cortese, 2001, p. 640).

In this study, we have used the objective method to estimate frequencies. In particular, we have used the BOUN Corpus described in Sak, Güngör & Saraçlar (2008).

3.3.2 The BOUN Corpus

This is a “web corpus” in the sense that it is a collection of documents harvested from the World Wide Web. It consists of 491 million tokens including punctuation marks, and has two sub-components. The first sub-component called "NewsCor" is based on three dailies published in Turkish, and contains 212 million tokens. The other sub-component called "GenCor" is a general sampling of Turkish webpages and contains 279 million tokens. The exact publication dates of the texts included in the corpus have not been reported in the relevant papers, but a safe guess would be that they all belong to the period after the year 2005.

Although the BOUN Corpus is a medium- to large-scale corpus even by modern standards, it is by no means representative of the Turkish language as a whole. Spoken language is represented only to the extent the news items in the dailies included in the corpus contain transcribed speech (such as statements by politicians), while formal written language, and especially news-reporting language, is overrepresented. This might cause significant sampling errors, especially at the lower end of the frequency spectrum.

In any study that uses a language corpus as an input, the question of representativeness inevitably arises. Is the BOUN Corpus suitable for the experiments described in this study? In our opinion, a corpus used in a set of visual-word-recognition experiments should be representative of the visual-word-recognition task rather than being "representative of the language as a whole". Once again, "because there are no true populations in language" (Popescu, 2009, p. 1), no corpus, however large and balanced, can claim to be "representative of the language as a whole". In this sense, a corpus dominated by newspapers and webpages is probably more suitable than a more "balanced" corpus that contains carefully selected samples from genres like poetry, law, medicine, sociology, etc. This is simply because newspapers and webpages presumably have a high share in the overall reading experience of our subjects (see Baayen, Milin & Ramscar (2016) for a detailed discussion of this issue).

3.3.3 Morphological analysis and disambiguation

Raw strings in a corpus are ambiguous. To repeat a famous example[21], the letter-string *koyun* has six interpretations, indicated in Table 7 together with their morphological analyses and English translations[22]:

[21] In fact, this example is problematic because *koyun(ii)* 'bosom' never occurs in the nominative. It must occur in the possessive, in which case the /u/ in the root disappears (e.g. *koynunda*, *koynuna*, *koynumuzda*, etc.).

[22] The number of possible parses reaches eleven if we include the five parses that treat the nominals above as predicates: is a sheep / is a bosom / is of the bay / is your bay / is your dark one.

Table 7. Six Possible Interpretations of the String koyun

Root	Morphological analysis	English translation
koyun(i)_Noun 'sheep'	*Noun*	sheep
koyun(ii)_Noun 'bosom'	*Noun*	bosom
koy_Verb 'to put'	*Verb+Imp+A2pl*	put (it)
koy_Noun 'bay'	*Noun+Gen*	of the bay
koy_Noun 'bay'	*Noun+P2sg*	your bay
koyu_Adj 'dark'	*Adj+P2sg*	your dark one

Thus, the more than 400 million tokens in the corpus must be morphologically analyzed, disambiguated[23] and counted before accurate frequencies can be calculated. For this purpose we used the "annotated corpus search engine" at tscorpus.com, which is described in Sezer & Sezer (2013) (referred to as "TSCorpus" below).

TSCorpus uses Sak, Güngör & Saraçlar (2007) for morphological analysis and disambiguation, which in turn uses the tagset described in Oflazer, Say, Hakkani-Tür & Tür (2003). The complete tagset is provided in Appendix A.

Sak, Güngör & Saraçlar (2007) report a 96.8% accuracy for their morphological disambiguator, but our personal experience with the data suggests that its performance on the BOUN Corpus is significantly worse, mainly because the disambiguation algorithm makes systematic errors in certain cases where the morphology of Turkish involves an inherent ambiguity that cannot be resolved without syntactic parsing, or even semantic analysis and/or world knowledge in some cases. The most important of these systematic errors concerns the accusative-

[23] "Disambiguation" in the context of morphological analysis refers the task of choosing the correct morphological analysis from among all possible morphological analyses.

possessive ambiguity that effects nouns ending in a consonant. Consider the sequence *hüzün insanı derinleştirir*, where the part *hüzün insanı* has the two parses indicated in Table 8, in the absence of additional information.

Table 8. Two Possible Interpretations of the Sequence *hüzün insanı derinleştirir*

Morphological analysis of *hüzün insanı*	English translation of entire sequence
Noun Noun+Acc	Sadness deepens a person.
Noun Noun+P3sg	The sadness person deepens [it].

Since the disambiguator described by Sak, Güngör & Saraçlar (2008) almost exclusively prefers the latter possessive interpretation in such cases (probably because this form is much more frequent in the aggregate), TSCorpus and other sources based on this disambiguator frequently use the possessive suffix where they should have used the accusative suffix. Since this systematic error is rather difficult to correct during post-processing, we had to leave this problem unresolved.

Another problem with the morphological analysis data is related to the list of roots used by the program. To give an example, the analyzer treats *telsizcilik* 'the profession of a radio operator' as a nominal root, probably simply because the root-list used by the program has been obtained from a printed dictionary that happens to have a separate entry for this complex word-form. The identically formed *oyunculuk* 'the profession of an actor', on the other hand, is analyzed as *oyun+Agt+Ness*. This inconsistency contaminates the word-form and template frequencies obtained from the corpus. To minimize the impact of this problem, we have manually prepared a list of all complex roots that were erroneously marked as simplex by TSCorpus, and have updated the relevant frequencies (see Section 3.4.12 for a more detailed description of this marking task).

3.4 Processing corpus data

The raw data obtained from TSCorpus has been processed by a series of computer programs developed for this study, mainly using the Python programming language. The steps are described below:

3.4.1 Download raw data

All word-forms that start with a letter other than *ğ*[24] have been downloaded by running the following CQP queries[25] on tscorpus.com: *[word="[aA].*"], [word="[bB].*"], …, [word="[zZ].*"]*

The raw data consists of the surface-form, lemma (stem) and morphological analysis of a single word, and the surface-form, lemma (stem) and morphological analysis of the words to its left and right. The present study only uses the word in focus, but the words to its left and right would be useful, for instance, for resolving the accusative-possessive ambiguity described above, for including clitics in morphological analyses, or for identifying repetitions.

The resulting raw data consists of 283,695,791 such lines. In other words, all quantitative measures reported in this study are based on an approximately 283-million-word corpus.

3.4.2 Normalize the data

In the context of data processing, "normalization" refers to any pre-processing performed to eliminate non-standard features and idiosyncrasies from raw data and

[24] There are no words in Turkish that start with the letter *ğ*.
[25] CQP is a query language that allows users to perform complex queries on annotated corpus data. For details, see http://cwb.sourceforge.net.

bring raw data to a consistent, standard format, thus preparing it for further analysis. We applied the following normalizations to the raw data:

(1) Remove circumflexes: The use of the circumflexed characters *â*, *î*, and *û* is not consistent in modern Turkish orthography. The Turkish word for 'paper', for instance, can be written as *kâğıt*, with a circumflex above the second letter, or much more frequently as *kağıt*, without the circumflex. In fact, the non-circumflexed version appears 31,813 times in the BOUN Corpus, while the circumflexed version appears only 9,619 times. To give a more striking example, the non-circumflexed and circumflexed versions of *milli* 'national' appear 30,469 and 1,863 times, respectively. To be able to count both forms as tokens of the same type, all circumflexes have been removed from the raw data.[26]

(2) Convert to lowercase: This allows us to treat *MASA*, *masa*, *Masa*, and even *mAsA*, as tokens of a single type, rather than as tokens of four separate types.

(3) Remove sense numbers: In some rare cases the raw data contains sense numbers of the form (i), (ii), (iii), etc., for certain homonymous roots such as *yaz* 'write' or 'summer', *çay* 'tea' or 'stream', *yüz* 'hundred' or 'swim', *koy* 'put' or 'bay'. These sense numbers have been removed.

3.4.3 Aggregate tokens

In this step, we simply count the number of times each word-form occurs in the dataset, and aggregate each form in a single line:

[26] The undesired consequence of this removal is that *hala* 'aunt' and *hâlâ* 'still' are now treated as a single type. However, such cases are extremely rare since the general tendency is to omit the circumflex.

Table 9. Aggregating the Tokens in BOUN Corpus

Before	After
jenerasyonu|jenerasyon|Noun+A3sg+P3sg+Nom	jenerasyonu|jenerasyon|Noun+A3sg+P3sg+Nom, 3
japonya|japonya|Noun+Prop+A3sg+Pnon+Nom	japonya|Japonya|Noun+Prop+A3sg+Pnon+Nom, 2
jenerasyonu|jenerasyon|Noun+A3sg+P3sg+Nom	
japonya|japonya|Noun+Prop+A3sg+Pnon+Nom	
jenerasyonu|jenerasyon|Noun+A3sg+P3sg+Nom	

The corpus contains 1,236,526 unique word-forms. When compared to the 200 billion forms proposed by Hankamer (1989), this shows the surprising extent of the sparsity phenomenon described in Section 3.4.8.

3.4.4 Calculate letter-ngrams

The above data allows the computation of letter-ngram[27] frequencies for the entire corpus, which will be needed to control for related variables during stimulus selection. The most frequent letters, letter-bigrams and letter-trigrams of Turkish are shown in Appendix B, Appendix C and Appendix D, respectively.

Although a detailed discussion of the letter-ngram statistics is beyond the scope of this study, let us mention a few salient characteristics: The corpus contains 865 distinct letter-bigrams and 10,120 distinct letter-trigrams. As in many other types of linguistic data (see Zipf, 1949; Baayen, 2001), the distributions are extremely uneven, with the top 10% letter-bigrams accounting for 66.4% of all letter-bigrams, and the top 10% letter-trigrams accounting for 82.3% of all letter-trigrams.

In the next step, the two procedures described above will be applied to suffix templates instead of word-forms:

[27] A sequence of two things is known as a “bigram”, a sequence of three things as a “trigram”, and a sequence of n things as an “ngram”. For example, the Turkish word *kedi* ‘cat’ contains the letter-bigrams *ke*, *ed*, and *di*, and the letter-trigrams *ked* and *edi*.

3.4.5 Aggregate templates

In this step, we count the number of times each suffix template occurs in the dataset, and aggregate each template in a single line:

Table 10. Aggregating the Templates in the BOUN Corpus

Before	After
Noun+A3sg+P3sg+Gen	*Noun+A3sg+P3sg+Gen*, 3
Noun+A3sg+Pnon+Nom	*Noun+A3sg+Pnon+Nom*, 2
Noun+A3sg+P3sg+Gen	
Noun+A3sg+Pnon+Nom	
Noun+A3sg+P3sg+Gen	

The corpus contains 28,313 unique suffix templates. Once again, when compared to the more than 10 million templates proposed by Hankamer (1989), the extent of the sparsity phenomenon described in Section 3.4.8 is striking. The implications of this surprising finding in terms of language processing will be discussed in Section 5.1.4.

3.4.6 Calculate template-ngrams

The above data also allows us to compute template ngram frequencies for the entire corpus. This data will be needed to control for ngram-related variables during stimulus selection. The most frequent suffix-bigrams and suffix-trigrams or Turkish are shown in Appendix F and Appendix G.

Once again, let us mention a few salient characteristics: The corpus contains 940 distinct suffix-bigrams and 4,459 distinct suffix-trigrams. As in the letter-ngram data, the distributions are extremely uneven, with the top 10% suffix-bigrams accounting for 87.3% of all suffix-bigrams, and the top 10% suffix-trigrams accounting for 95.3% of all suffix-trigrams.

Another striking result is the extreme frequency of the morpheme *-(s)I* (tag: *P3sg*), both in terms of types and tokens. First of all, *-(s)I* is by far the most frequent suffix of Turkish: It occurs 37,531,526 times in the BOUN Corpus, followed by the plural marker, which occurs 22,780,373 times, and the passive marker, which occurs 10,402,113 times (Appendix E). In the list of the most frequent 200 suffix templates (Appendix H), *P3sg* occurs 48 times, including the 2nd, 5th, 8th, 10th, 11th, 12th and 16th top positions. All of the top seven middle-position suffix bigrams (Appendix F), and all of the top fifteen middle-position suffix trigrams (Appendix G) contain *P3sg*. In short, this morpheme dominates the statistics of Turkish morphology at every level.

P3sg is the possessive suffix of the third person singular and can be expressed as *-(s)I* using the allomorph notation (i.e. it has the following eight variants: *-ı*, *-i*, *-u*, *-ü*, *-sı*, *-si*, *-su*, *-sü*). It appears in two constructions: (a) genitive constructions like *Evrem'in arabası* 'Evrem's car' or *pencerenin camı* 'the glass of the window', and (b) compound nouns like *pencere camı* 'window-glass'.

According to Hayasi (1996), the genitive constructions in (a) have a phrase-like nature, while the compound constructions in (b) have inherent word status. Traditional grammars do not distinguish between (a) and (b), and treat both under genitive constructions, while others (e.g. Swift, 1963) treat the *-(s)I* in the possessive compounds as in (b) above as a "compound marker". Despite acknowledging that Swift's treatment is superior to the traditional treatment, Hayasi (1996) discusses several problems involved in describing *-(s)I* as a compound marker, and concludes that "possessive compounds in Turkish cannot be exclusively dealt with either in morphology or in syntax" (Hayasi, 1996, p. 125).

In view of the above discussion and the extreme frequencies identified in this study, we will treat *-(s)I* as a separate category of its own.

3.4.7 Verify corpus size

It is known since Zipf (1949) that lexical data is characterized by extremely uneven distributions: There are a relatively few items that are extremely frequent, and a very large number of items that are extremely rare. This tendency applies at every level of linguistic analysis including words, syllables, morphemes, phonemes and graphemes, as well as sequences made up of them. An important consequence that is relevant for our purposes is that lexical parameters are highly dependent on sample (i.e. corpus) size. In Baayen's words: "This property sets lexical statistics apart from most other areas in statistics, where an increase in sample size leads to enhanced accuracy and not to systematic changes in basic measures and parameters." (Baayen, 2001, p. 1).

To see how the BOUN Corpus behaves in this regard, we have counted the number of lemmas encountered for the first time, as corpus size increases by chunks of 10 million words. The resulting vocabulary-size histogram is shown in Figure 3:

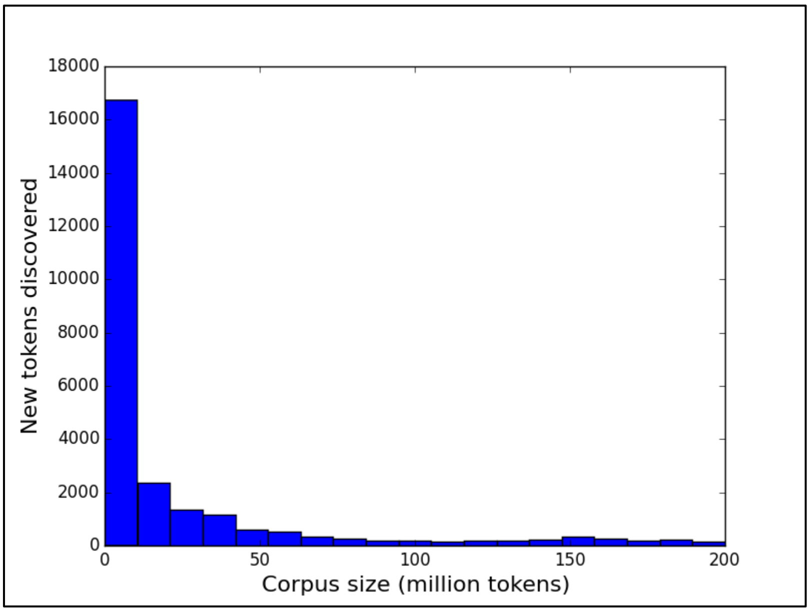

Figure 3. Vocabulary size as a function of corpus size

Figure 3 shows that, although new lemmas continue to be discovered even as corpus size increases from 190 to 200 million, the rate of discovery decreases dramatically after the first 10 million tokens. This histogram verifies the basic findings of Zipf (1949) and Baayen (2001), and suggest that the BOUN Corpus does not exclude a significant number of word-forms and suffix templates, and is thus not an inappropriate source for the purposes of this study.

3.4.8 Create suffix tree

In this step we generate the suffix tree defined in Section 3.2.1 above. Figure 4 shows a very small subset of the suffix tree derived from the BOUN Corpus (162 nodes out of 30,666). The node in the center represents the (non-terminal) template *Acquire+Caus+Pass+Able*, which attaches to nominal roots and generates words of the form *-lendirilebil-*.

The full version of the tree consists of 30,666 nodes. 7,320 of the nodes are non-terminal nodes added by us in order to link any given node to the topmost node. The remaining 23,346 nodes represent actual suffix templates attested in the BOUN Corpus. Of these, 11,808 attach to verb roots, 7,733 to noun roots, 2,919 to adjective roots, 829 to pronoun roots, 29 to postposition roots and 24 to the question clitic *-mI*. Since they do not have any derivational or inflectional forms, adverbs, conjunctions, determiners and interjections are each represented by a single node on the tree.

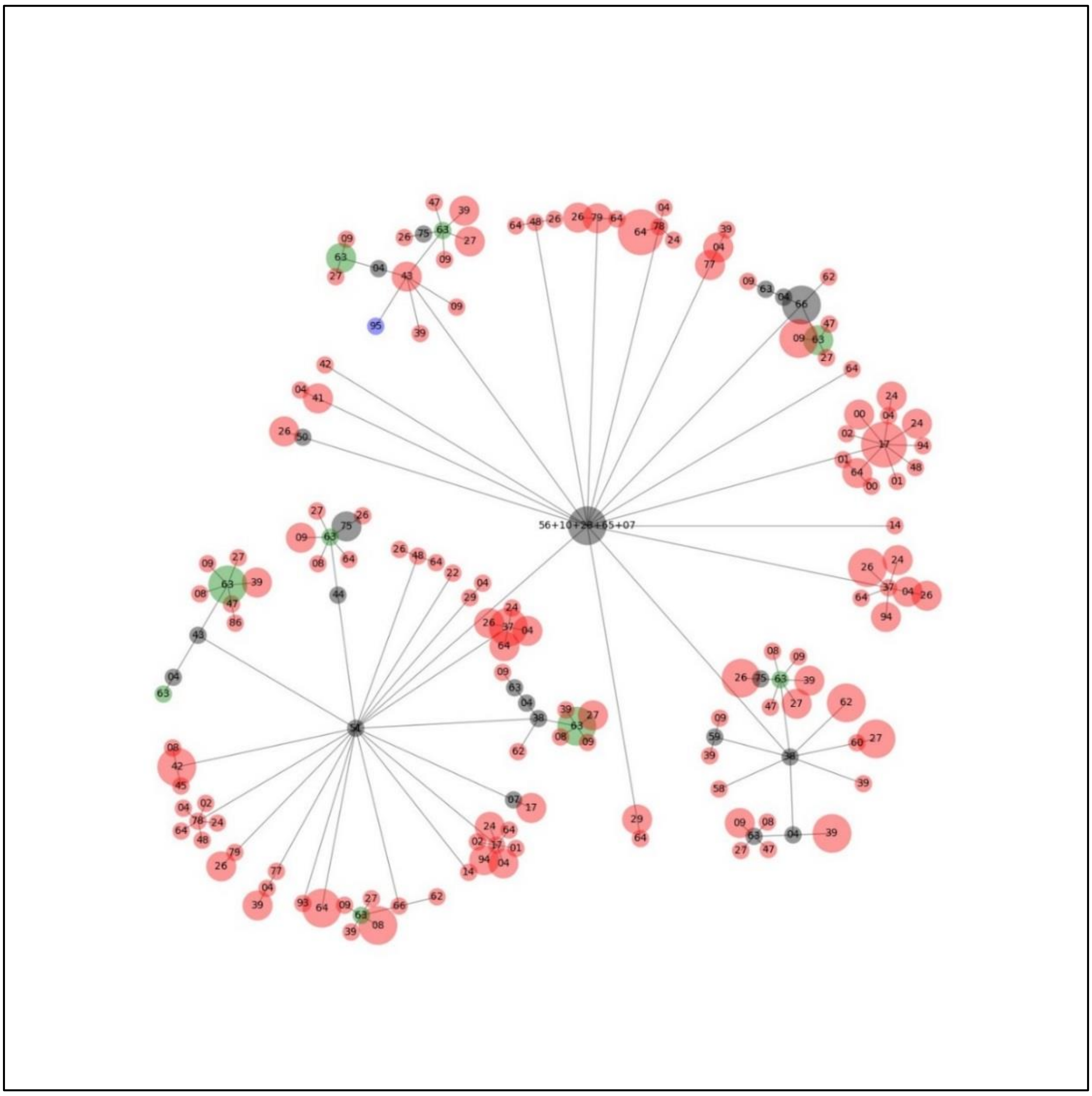

Figure 4. A small subset of the suffix tree

Although it looks complicated for a tree that visualizes only the children of the rather complicated suffix template *Noun+Acquire+Caus+Pass+Able*, the above suffix tree does not contain a node for every grammatically possible child of that template. The 30,666 nodes in the full version of the suffix tree represent only those suffix sequences that are attested in the BOUN Corpus. Literally tens of thousands of grammatically possible suffix sequences are absent from the corpus, even though it is a rather large corpus even by modern standards.

The extended suffix tree in Figure 5 demonstrates this sparsity: The (non-terminal) node in the center represents the sequence *Noun+Agt+Become+Neg*, whose children can form words like *politikacılaşmamalısın*, *politikacılaşmaması* or *politikacılaşmadığına*. As before, the yellow nodes are non-terminal nodes that have been artificially added to make sure that there exists a path between every node of the tree and the root. They can thus be disregarded. The single blue node represents the suffix template *Noun+Agt+Become+Neg+PastPart+P3sg+Dat*, which occurs only once in the corpus, in the form of *politikacılaşmadığına* 'to the fact that he/she has not turned into a politician'.

The interesting part is the gray nodes, which represent grammatically valid suffix templates that do not occur in the corpus even once. As will be seen in the section on Experiment 2 below, this sparsity phenomenon creates a unique opportunity for experiment designers to create words that are perfectly grammatical and still have zero surface-frequency.

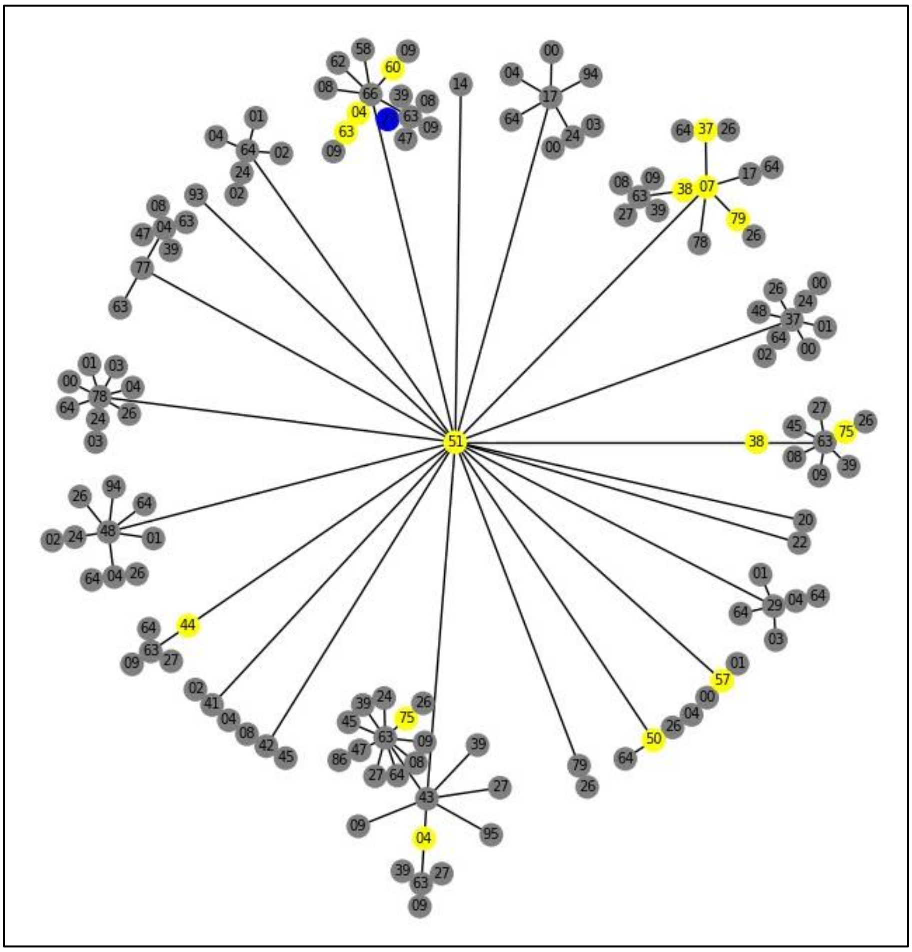

Figure 5. A particularly sparse region of the suffix tree

3.4.9 Calculate template parameters

In this step, we calculate the parameters defined in Section 3.2.4 for the 23,346 suffix templates attested in the BOUN Corpus. This dataset is hoped to contribute to the growing literature on quantitative surveys on Turkish (see Güngör, 2003; Aksan et al., 2012; Durrant, 2013; Sezer & Sezer, 2013; Erten, Bozşahin, & Zeyrek, 2014 and Tolgay, 2015 for related work).

3.4.10 Form the frequency matrix

Using the resources described above, we now form a single matrix to represent all nominal roots[28], and the frequencies of their various morphological forms.

As mentioned in Section 3.4.8 above, 7,733 nominal suffix templates are attested in the BOUN Corpus. Once again, these exhibit an extremely uneven distribution, with the most frequent 200 nominal templates accounting for 99.3% of all tokens with a noun root. These are listed in Appendix H and the corresponding rank-coverage graph can be seen in Figure 6.

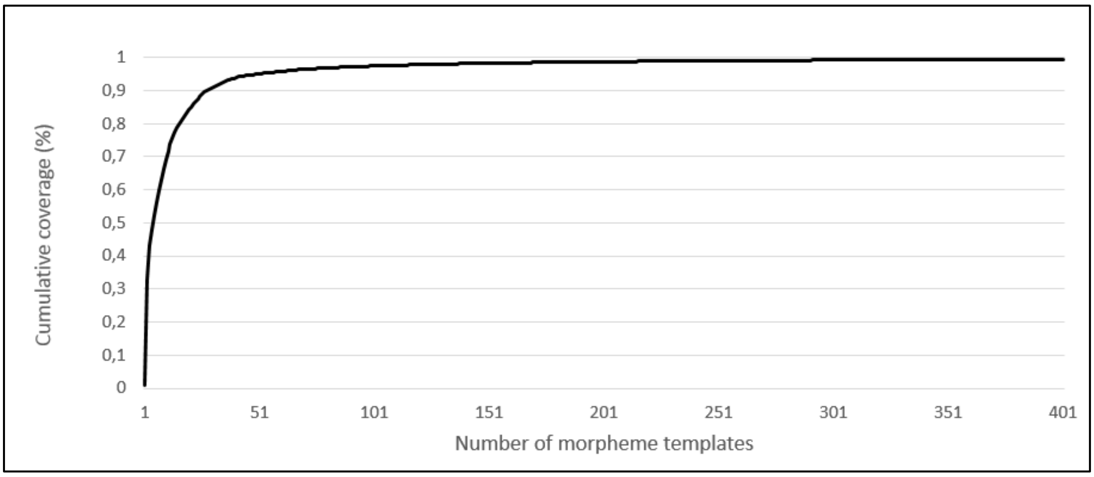

Figure 6. Frequency rank vs. coverage

[28] As before, we focus only on nouns because the experiments use only nouns as stimuli.

Considering that the remaining 7,533 suffix templates account for only 0.7% of the tokens, we have chosen to prepare a 200-dimensional frequency matrix for noun roots, using only the top 200 suffix templates instead of using all 7,733. The resulting 20,661 x 200 matrix has the following structure:

$$f_{200}\begin{bmatrix} R_1 \\ R_2 \\ R_3 \\ \dots \\ R_n \end{bmatrix} = \begin{bmatrix} f(R_1+T_1), f(R_1+T_2), f(R_1+T_3), \dots, f(R_1+T_{200}) \\ f(R_2+T_1), f(R_2+T_2), f(R_2+T_3), \dots, f(R_2+T_{200}) \\ f(R_3+T_1), f_7(R_3+T_2), f(R_3+T_3), \dots, f(R_3+T_{200}) \\ \dots\ \dots\ \dots \\ f(R_n+T_1), f(R_n+T_2), f(R_n+T_3), \dots, f(R_n+T_{200}) \end{bmatrix}$$

$$where\ f(T_1) > f(T_2) > \cdots > f(T_{200})$$

3.4.11 Reduce to four dimensions

In this step, we classify the above-mentioned 200 nominal suffix templates into four categories, thus gaining the ability to reduce the frequency matrix from 200 to 4 dimensions. The four categories in question are “bare root”, “inflectional forms”, “derivational forms” and “*–(s)I* forms” (as discussed in Section 3.4.6, the compound marker *–(s)I* has been represented as a separate dimension of its own).

The resulting 20,661 x 4 matrix has the following structure:

$$f_4\begin{bmatrix} R_1 \\ R_2 \\ R_3 \\ \dots \\ R_n \end{bmatrix} = \begin{bmatrix} f(R_1), f\left(R_1+T_{inf}\right), f(R_1+T_{drv}), f(R_1+T_{CM}) \\ f(R_2), f\left(R_2+T_{inf}\right), f(R_2+T_{drv}), f(R_2+T_{CM}) \\ f(R_3), f\left(R_3+T_{inf}\right), f(R_3+T_{drv}), f(R_3+T_{CM}) \\ \dots\ \dots\ \dots \\ f(R_n), f\left(R_n+T_{inf}\right), f(R_n+T_{drv}), f(R_n+T_{CM}) \end{bmatrix}$$

3.4.12 Improve the root dictionary

As mentioned in Section 3.3.3, the morphological disambiguator treats certain complex stems as simplex (bare) roots. To continue the *telsizcilik* example, this incorrect treatment has the following undesired consequences:

i. The derivational-form frequencies of all tokens that start with *telsiz* are underestimated.

ii. All template ngram frequencies that involve the suffixes –*CI* and–*lIk* are underestimated.

iii. The data contains superfluous frequency measures for *telsizcilik.*

To partially solve these problems, 20,661 lemmas in the "root" dictionary have been manually scanned and marked as either (a) simplex (10,247 entries), (b) complex with a transparent structure (5,984 entries), or (c) complex with an opaque structure (4,430 entries). Below are some comments about the method used in this marking task:

(a) Rule of thumb: A word is complex if some subset of its characters significantly overlaps with another, semantically related word, even if the remainder is not a valid and productive affix and even if certain phonological and/or orthographic transformations are involved. Here are some words marked as complex in accordance with this rule: *Türkolog-Türk, likidasyon-likid, sinematik-sinema, bankamatik-banka, antidot-anti, biberiye-biber, yanılsama-yanıl, nadirat-nadir, kanserojen-kanser, gramaj-gram, daktilograf-daktilo, dekoratör-dekor, otlak-ot, garantör- garant[i]*.

(b) Certain apparently complex words proposed as part of the "Turkish language reform" starting in the 1930s have been treated as simplex, considering that their internal structure is unclear: *izlenim-izle?*, *birey-bir?, kalıtım-kal?, sınav-sına?*, *etmen-et?, dışkı-dış?, koşul-koş?, tanım-tanı?*, *kapsam-kapsa?*, *tümce-tüm?...*[29]

(c) Compound words have been treated as complex: *telgraf+hane, anlam+bilim, baş+müdür…*

(d) Certain onomatopoeia have been treated as complex: *çağıltı, şırıltı, tıkırtı…*

(e) Members of certain extended word families formed by widely used Persian and Latin suffixes have been treated as complex: *-name*, *-hane*, *-metre*, *-loji*, *-syon*, *-izm…*

(f) Some complex lexicalizations have been treated as simplex: *kahvaltı, tersane…*

(g) Many words that are complex according to the morphotactics of Arabic, Persian and Latin but would appear simplex to the average native speaker of Turkish have been treated as simplex: *malikane*, *rençber*, *tezyinat*, *züccaciye*, *permanganat*, *anemi*, *ameliye*, *kameriye*, *sürmenaj…*

The result of this manual processing task is that the root dictionary of Sak, Güngör & Saraçlar (2007) erroneously lists 5,984 complex-transparent forms like *telsizcilik* and 4,430 complex-opaque forms like *ultrason* as simplex, meaning that only 49.6% of the lemmas in the "root" dictionary are actual roots.

3.4.13 Recalculate 4D frequencies

In this step, we recalculate the four-dimensional frequencies described in Section 3.4.11, using the root-dictionary improvement described in Section 3.4.12. This involves two operations:

[29] For a detailed discussion of the Turkish language reform, see Lewis (1999).

i. Remove all complex stems erroneously marked as simplex (e.g. *telsizcilik*, *ultrason*, *kapsam*).

ii. In complex-transparent/simplex pairs like *telsizcilik*/*telsiz*, augment the third frequency value (derivational-forms frequency) of the simplex entry (*telsiz* in this case) by the sum of the four frequency values of the complex-transparent form (*telsizcilik* in this case).

The result is a 10,247 x 4 matrix that contains the bare-form, inflectional-form, derivational-form and compound-form frequencies of 10,247 Turkish simplex noun roots.

3.4.14 Visualize 4D frequencies

Based on the data prepared in Section 3.4.13, Figure 7 visualizes the four frequencies of all simplex nouns of Turkish in three-dimensional space. Each of the 10,247 data points represents a simplex noun. The x, y and z axes represent the frequencies of inflectional forms, derivational forms and *–(s)I* compound forms, respectively, while color represents the frequency of the bare root ("hot" colors: high frequency, "cold" colors: low frequency).

Probably the most salient characteristic of the distribution is the positive correlation between the four frequency measures. The correlation matrix is shown in Table 11.

Table 11. Correlation Between the Four Frequency Measures

	Root	Inflectional	Derivational	Compound
Root		+0.68	+0.35	+0.47
Inflectional			+0.31	+0.65
Derivational				+0.25
Compound				

A second observation is that the “word cloud” gets thinner as frequency increases. This is a visual confirmation of Zipf (1949): There is a large number of low-frequency events and a small number of high-frequency events.

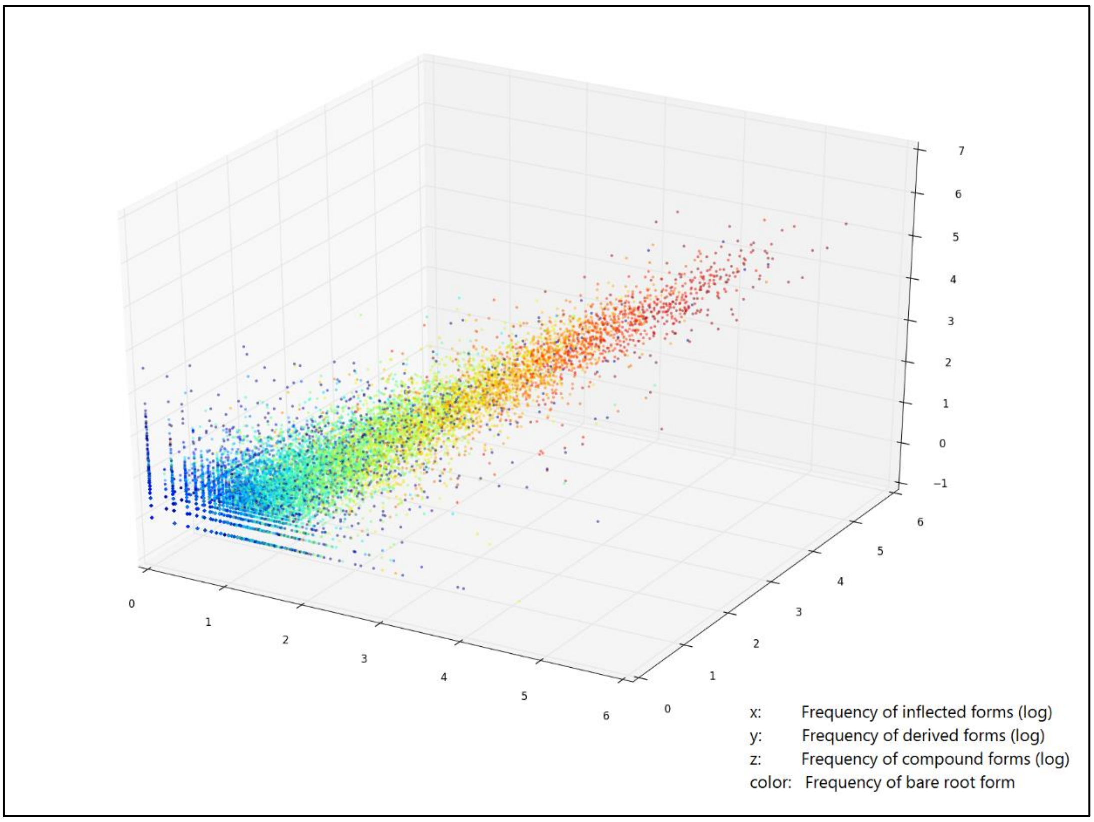

Figure 7. Four-dimensional frequency distribution of Turkish simplex nouns

This concludes the discussion on the preparatory quantitative work that had to be done in order to design and implement the experiments described in Chapter 4. Most of the work described up to this point has been independent of the experiments, and have general applicability. The resulting datasets can be used as general language resources.

CHAPTER 4

METHODOLOGY AND THE EXPERIMENTS

This chapter describes the design of an online platform for conducting lexical decision experiments outside the laboratory environment, and the design, implementation, and results of two lexical decision experiments using the same online platform.

4.1 Designing a platform for online experiments

For the reasons mentioned in Section 3.1.3, the experiments have been conducted online. The following sections describe the software platform developed for this purpose, and the relevant experiment design decisions.

4.1.1 Using the 'jsPsych' JavaScript library

jsPsych is a publicly available JavaScript library for "creating and running behavioral experiments in a web browser".[30] The highly customizable open-source library offers built-in functionality for displaying instructions, presenting stimuli and measuring response times (de Leeuw, 2015).

4.1.2 Personal and demographic data

To participate in the experiments, subjects are required to fill in the registration form in Figure 8. In addition to providing basic information regarding the experiment and its purpose, the registration form collects gender, age, handedness, education and

[30] The library and related documentation is available at www.jspsych.org.

native-language data from participants. The relationships between these variables and response times is an interesting issue, but will not be discussed in this study.

Figure 8. Registration page for participants

4.1.3 Choice of keys

Subjects in a lexical decision experiment are asked to press either the ‘yes’ button or the ‘no’ button. In an online experiment, these will have to be two keys on the subject’s own keyboard. Considering that we require subjects to be native speakers of Turkish, a straightforward choice would be to use ‘E’ for ‘yes’ (*evet* in Turkish) and ‘H’ for ‘no’ (*hayır* in Turkish). But in an online experiment, the experimenter has no control over the subjects’ keyboard layouts: Some might be using a standard QWERTY keyboard, others might be using some version of the “Turkish-F keyboard”, while still others use some other unknown layout.

A second complication is related to handedness: A pilot study conducted before the first experiment revealed that subjects tended to prefer pressing the ‘yes’ button with their dominant hand, and the ‘no’ button with their non-dominant hand, regardless of button position. Several pilot study subjects have declared that pressing the ‘yes’ button with their non-dominant hand “felt awkward/difficult”. To be able to systematically investigate this anecdotal handedness effect, the position of the ‘yes’ and ‘no’ keys have been randomized in all experiments: Subjects were randomly assigned to one of the two setups. A similar approach has been adopted by Snodgrass & Jarvella (1972) and Rice & Robinson (1975).

To measure the effect of handedness on response times, we also need to make sure that the subjects use separate hands for the two buttons. Using the keys for letters, whose position on the keyboard is unknown, is therefore not a viable option. A reasonable choice would be to use ‘1’ for ‘yes’ and ‘0’ for ‘no’. But ‘1’ and ‘0’ are not neutral numbers: Subjects would probably tend to assume that ‘1’ represents ‘yes’ and ‘0’ represents ‘no’, and would get confused when asked to press ‘1’ for ‘no’ and ‘0’ for ‘yes’.

To summarize, we need two non-letter buttons with relatively stable positions across keyboard models, minimum symbolic content and maximum physical distance from each other. The optimum choice seems to be ‘2’ and ‘9’.

4.1.4 Lowercase vs. uppercase

The earlier stages of the lexical decision task probably involve purely visual pattern recognition (Dehaene, 2009). This means that the physical features of letters might introduce unintended visual biases, thus confounding response time measurements. Considering that there exist significant visual differences between several lowercase and uppercase letters of the Latin alphabet (compare, for instance, the pairs a-A, b-B, d-D, e-E, g-G, l-L, r-R), we used two sets of stimuli, one containing lowercase letters and the other their uppercase versions.

4.1.5 Randomization of the order of stimuli

To prevent earlier stimuli from affecting the processing of subsequent stimuli in unexpected ways, the order of presentation has been randomized for each subject, using the randomization module of the jsPsych library.

4.1.6 Within-subjects design

Both experiments reported in this study use a "within-subjects / repeated-measures" design, which is the typical experimental design in the field of memory and language.

What are the benefits of a "within-subjects / repeated-measures" design? Imagine that the high-frequency stimuli were given to a group of 50 people, and the low-frequency stimuli to a an entirely different group of 50 people. In this case, random differences between the individuals in the two groups would be a confounding factor: Subjects in the first group can, by pure chance, be younger, more intelligent, less-educated, etc. than subjects in the second group, or vice versa, which would have an unintended impact on the mean response times of the two groups. In a within-subjects / repeated-measures design, by contrast, the same person is exposed to both the high-frequency stimuli and the low-frequency stimuli. In other words, every subject serves as his/her own control, thus eliminating the possibility of chance differences between subjects (Raaijmakers, Schrijnemakers & Gremmen, 1999, p. 416).

This concludes the description of the experimental setup that has been used in both lexical decision experiments. The sections that follow describe the specific details of these experiments, and their results.

4.2 Experiment 1

The first experiment asks the most basic question possible: Is there any frequency effect at all in the visual processing of Turkish words? Frequency effect is probably the most well-established outcome of psycholinguistic research, but its existence in Turkish has not been adequately demonstrated before, to the best of our knowledge.[31] The following sections describe the issues that have been taken into consideration while designing the experiment.

4.2.1 The collinearity issue

As seen in Table 11 and Figure 7, all four frequency measures proposed in Section 3.4.11 are highly correlated. This collinearity issue can be addressed in two ways:

The first method would be to vary one of the frequency measures while keeping the other three constant. The main problem with this approach is that there wouldn't be enough items to choose from, because fixing the three remaining frequency measures defines a small region within the word cloud in Figure 7, from which all stimuli must be selected. Furthermore, even if this sparsity problem is somehow overcome, most of the selected items would be atypical in terms of their frequency distribution, and an experimental result based on atypical stimuli cannot be generalized to the language as a whole. Another problem is that fixing the values of the three frequency measures at certain levels will inevitably involve arbitrary choices.

A second, alternative method would be to allow the four frequency measures to covary. Visually, this corresponds to taking a sample from the core of the "word cloud" in Figure 7. The first advantage of this approach is that it solves the sparsity

[31] See Section 2.4 of the literature review for a critique of existing studies.

problem described above. Secondly, this approach does not require the fixing of frequencies at arbitrary levels. Another major advantage is that the "core of the cloud" consists of the most typical words of the language, at least in terms of the four frequency measures visualized here. Finally, the "core of the cloud" is a well-defined concept in mathematical terms. In view of these advantages, this "core sampling" method has been used in Experiment 1. Figure 9 reproduces the word cloud in Figure 7, where potential stimuli at the core of the cloud are represented by larger dots.

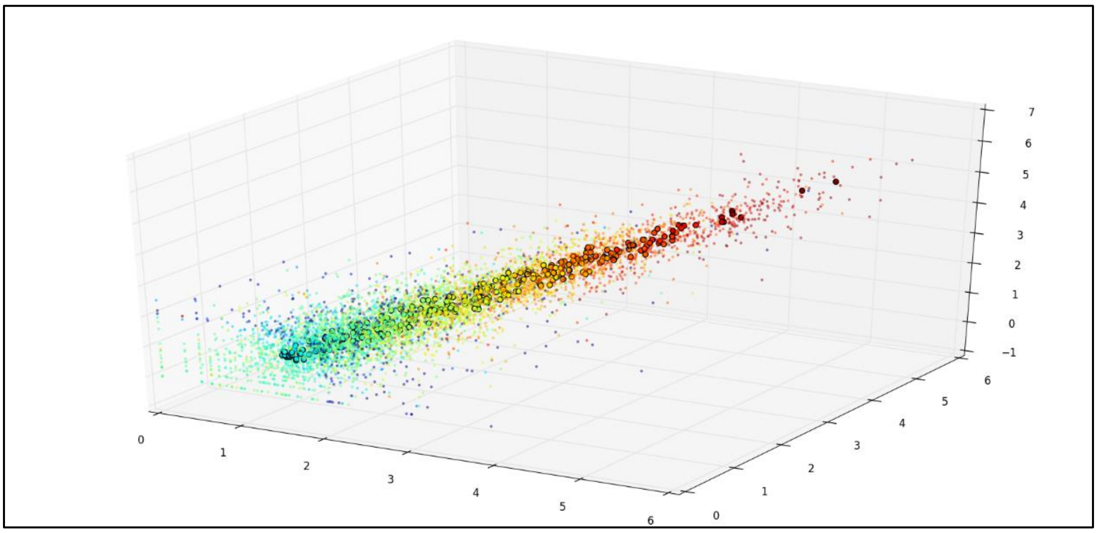

Figure 9. Potential stimuli located at the core of the word-cloud

In mathematical terms, the above-described stimulus selection area can be described through principal component analysis (PCA), which finds the vectors (components) that have maximum variance in terms of the data points. The so-called "core of the cloud" lies along the first component discovered through PCA.

4.2.2 Choosing the right diameter for the "sampling cylinder"

Since Experiment 1 aims to test the existence of a frequency effect in the broadest possible sense of that term, the four frequency measures of the stimuli should covary to the maximum extent possible. Expressed statistically, the pairwise correlation coefficients between the four frequency measures should be as close to 1.0 as possible. Geometrically, this means that the stimuli should be as close to the core of the word cloud as possible, i.e. the diameter of the "sampling cylinder" should be as small as possible.

The six correlation coefficients calculated for the six possible pairs of the four frequency measures have already been shown in Table 11. These are the lowest possible values of the correlation coefficients because all "outlier" data points remain within the cylinder when diameter exceeds a certain level. As diameter is reduced, many of the outliers are excluded, and the correlation coefficients begin to increase.

On the other hand, as the diameter of the sampling cylinder is reduced, the number of suitable candidates falls. In the extreme case where the diameter is reduced to zero, there are no items at all to include in the experiment.

This trade-off requires the experimental selection of an optimum diameter, where there is an acceptably high correlation between the four frequency metrics and at the same time a sufficient number of stimulus candidates. Figure 10 shows the decrease in the six pairwise correlation coefficients and the increase in the number of candidates as diameter increases from 0.05 to 2.50.

As can be seen in Figure 10, all six correlation coefficients are above 0.90 for diameter values smaller than 0.25. Moreover, setting the diameter of the sampling cylinder at 0.25 gives us 183 stimulus candidates to choose from. This is clearly sufficient considering that we want to choose 50 stimuli out of 183 candidates, and 183C50 = 2.67×10^{45} is a huge number. Hence, setting the diameter at 0.25 seems to be a good choice.

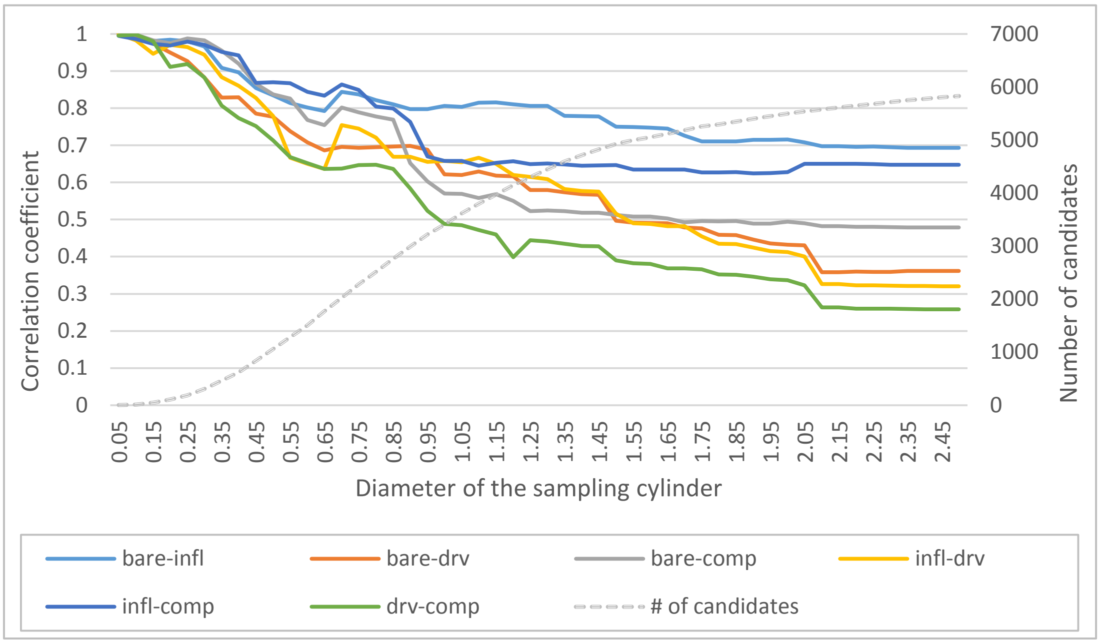

Figure 10. Correlation as a function of sampling cylinder diameter

4.2.3 Control variables

Following the literature on lexical decision experiments, the stimuli have been matched for (1) number of characters, (2) number of syllables, (3) mean bigram frequency, and (4) number of orthographic neighbors.

Although the first two variables are straightforward, the third and the fourth one might require some discussion:

The "mean bigram frequency" of the Turkish word *kedi* 'cat', for example is calculated as follows: The string *kedi* contains the three bigrams: *k-e*, *e-d*, and *d-i*, which occur 7,204,289, 7,127,686, and 15,186,164 times, respectively, in the BOUN Corpus. Thus the mean bigram frequency of *kedi* is equal to the sum of these three values, divided by three (29,518,139/ 3 = 9,839,379).

Ferrand *et al.* (2010) defines "orthographic neighborhood" as "words that differ by changing, adding, or deleting a letter or by swapping two adjacent letters" (Ferrand et al., 2010, p. 3). Davis, Perea, & Acha (2009), on the other hand, distinguish between five types of neighbors: (1) Substitution neighbors (e.g. *gone* –

done); (2) transposition neighbors (e.g. *trail* – *trial*); (3) neighbors once-removed, i.e. substitution+transposition (e.g. trial – trawl); (4) addition neighbors (e.g. *dive* – *drive*); (5) deletion neighbors (e.g. *drive* – *dive*).

This study uses the "KelimetriK" tool described by Erten, Bozşahin, & Zeyrek (2014) to calculate orthographic neighborhood sizes, because the tool covers four of the five types of neighbors described by Davis, Perea & Acha (2009) (only "neighbors once-removed" is not covered).

4.2.4 Selection of two matched stimulus sets

Stimuli used in the high-frequency and low-frequency conditions should be matched for the control variables mentioned in Section 4.2.3. This is arguably the most important step in the design of the experiment, because it eliminates the potential impact of confounding variables to the maximum extent possible, while varying the independent variable at the same time. However, it is extremely difficult, if not impossible, to perform this task manually. A custom-made algorithm has been designed to automate this critical step.[32]

The brute-force algorithm first randomly selects 25 + 25 = 50 items from among the 183 candidates described in Section 4.2.2, where the frequencies of the first 25 items are all lower than the frequencies of the second 25 items. It then computes a single score that quantifies the difference between the two groups in terms of the means and standard deviations of the control variables.

Before calculating this score, however, "feature scaling" is used to normalize the four frequency metrics, so that each frequency metric has a comparable effect on

[32] There exist some proprietary and open-source algorithms and software tools that have been designed for similar purposes. See, for example, van Casteren & Davis (2007) for details of the "Match" algorithm.

the overall score. This normalization brings all frequency values to the range [0, 1], using the following formula:

$$f' = \frac{f - \min(f)}{\max(f) - \min(f)}$$

The score mentioned above is equal to the absolute difference between the means plus the absolute difference between the standard deviations of the frequency measures in the two groups. The smaller the score, the closer the two stimulus groups will be to each other in terms of their means and standard deviations. The algorithm performs millions of random trials and reports the stimulus groups that have the lowest score, or groups whose score is below a certain minimum value.

The performance of the algorithm in selecting the stimuli of Experiment 1 can be seen by examining the last four rows of Table 12.

4.2.5 Creation of non-words

Experiment 1 contains 45 non-words in three categories: (1) phonotactically plausible non-words generated by the Turkish localization of the software tool "Wuggy"[33] (e.g. *netaet*), (2) phonotactically plausible non-words generated by Wuggy plus an actual suffix (e.g. *pasagınçlı*), (3) random letter strings (e.g. *möptd*). The full list of these non-word stimuli is provided in Appendix I.

[33] "Wuggy" is a software tool used for generating phonotatically valid pseudo-words for psycholinguistic experiments. It was originally developed by Keuleers & Brysbaert (2010) and uses an algorithm that decomposes words into their sub-syllabic components, and constructs bigram frequency chains using the full words' onset, nucleus and coda patterns. The Turkish localization of Wuggy is described by Erten, Bozşahin, & Zeyrek (2014).

4.2.6 Training effect / fatigue effect

Training effect is the phenomenon that, up to a certain point, subjects get better at performing the task presented to them during an experiment. This results in a continuous reduction in response times and error rates, not because of the intrinsic properties of the stimuli, but simply as a result of practice. The 20th stimulus of a lexical decision experiment, for instance, is on average processed faster and more accurately than the 10th stimulus, which in turn is processed faster and more accurately than the first stimulus, etc., regardless of the nature of the stimuli in question (see, for example, Howes & Solomon, 1951, p. 407).

Fatigue effect refers to a similar phenomenon: After a certain point, the subjects get tired and begin to lose interest in the experimental task. The result is a reversal of the training effect, evidenced by a gradual increase in response times and error rates.

The training effect can be minimized (a) by adding a training section at the beginning of the experiment and explicitly telling the subjects that this section is for training and will not affect the results; and/or (b) by examining experiment results and accordingly removing the first *n* items from the experiment, until the point the training effect disappears. In the present study, we used a 10-item training section at the beginning of the experiments.

4.2.7 Two sets of stimuli

To minimize the generalizability problem described in Section 3.1.2, the automatic stimulus selection algorithm described in Section 4.2.4 has been run twice, to obtain two disjoint sets of stimuli that satisfy the relevant criteria of the experiment, and the two sets have been randomly assigned to participants. The two have similar characteristics in terms of the independent variable (frequency) and the control variables (described below), but none of the words in one set appears in the other. The hope is that this will further reduce the impact of uncontrolled chance factors. In a sense, in each experiment, there are two experiments that replicate each other. This is also a demonstration of the validity of the stimulus selection mechanism described here.

Experiment 1 has the following 2 x 2 x 2 = 8 versions, and the computer randomly chooses one of them as the subject presses the start button:

Version 1: Stimulus Set 1, lowercase, ‘yes’ button on left

Version 2: Stimulus Set 1, lowercase, ‘yes’ button on right

Version 3: Stimulus Set 1, uppercase, ‘yes’ button on left

Version 4: Stimulus Set 1, uppercase, ‘yes’ button on right

Version 5: Stimulus Set 2, lowercase, ‘yes’ button on left

Version 6: Stimulus Set 2, lowercase, ‘yes’ button on right

Version 7: Stimulus Set 2, uppercase, ‘yes’ button on left

Version 8: Stimulus Set 2, uppercase, ‘yes’ button on right

4.2.8 No fillers

No filler items have been used in Experiment 1. The motivation is to limit the average duration of the experiment, thus reducing dropouts. However, it should be admitted that the absence of filler items will make it easier for the subjects to guess the purpose of the experiment, thus gaining a better ability to develop unpredictable response strategies, and as a result confounding the results.

4.2.9 Stimuli of Experiment 1

50 real words have been used in Experiment 1. Although using a larger number of words would have been better in terms of statistical power, the number has been limited considering that this is an online experiment where all participants join voluntarily, and would not be motivated to spend more than a few minutes in front of the computer.

The stimulus matching algorithm described in Section 4.2.4 has selected the (25 + 25) x 2 = 100 nouns in Appendix J to serve as the two disjoint stimulus sets of Experiment 1.

The means and standard deviations calculated for the four elements of the independent variable (various measures of word frequency) and the four controlled variables are shown in Table 12.

Table 12. Means and Standard Deviations of Independent and Control Variables

	STIMULUS SET 1				STIMULUS SET 2			
	Low frequency		High frequency		Low frequency		High frequency	
Variable	MEAN	STD	MEAN	STD	MEAN	STD	MEAN	STD
Root freq.	723	435	11228	13297	522	422	14003	12240
Inflected forms freq.	388	268	6960	8307	322	286	8045	8868
Derivational forms freq.	35	27	631	637	28	24	994	1056
Compound forms freq.	227	139	5937	6950	194	201	8617	8871
Word length	5.60	1.44	5.60	1.44	5.76	1.36	5.76	1.45
Number of syllables	2.32	0.73	2.32	0.79	2.24	0.71	2.24	0.81
Mean bigram frequency	135	92	139	92	145	92	142	97
Orthographic neighbors	4.00	5.61	4.16	5.62	3.72	5.29	3.68	5.30

4.3 Results of Experiment 1

The first experiment has been completed 664 times[34]. As will be described in Section 4.3.1, below, 182 of the subjects have been removed from the results since their accuracy was below 90%. Thus, demographic data will be reported only for the remaining 482 subjects:

(1) In 431 of the trials (89.4%), the subject reported to be right-handed, and in 51 trials (10.6%) left-handed.

(2) In 230 trials (47.7%), the subject reported to be male, and in 243 (50.4%) female, while in 9 cases (1.9%) the subject preferred not to specify any gender.

(3) In 391 trials (81.1%), the subject reported to speak only one native-language (Turkish), and in 57 cases (18.9%) more than one native-language including Turkish.

Figure 11 shows the distribution of the subjects by age, Table 13 the distribution of the subjects by education level, and Table 14 the distribution of the subjects between the experiment's eight versions.

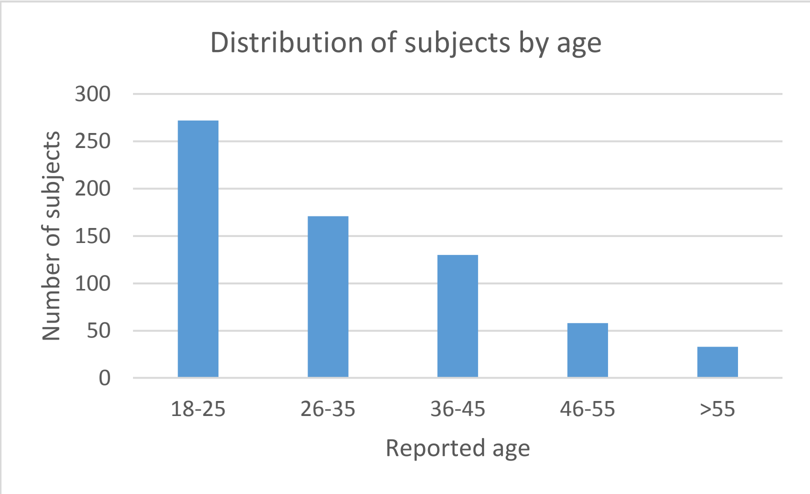

Figure 11. Distribution of subjects by age (Experiment 1)

[34] We cannot claim that the experiment has been completed by 664 different individuals since the experiment was conducted online and no measures have been taken to prevent the same person from completing the experiment more than once. The same uncertainty also applies to the demographic information provided by the subjects.

Table 13. Distribution of Subjects by Education Level (Experiment 1)

Primary-school	13	Secondary-school	4
High-school	38	Pre-graduate	29
Undergraduate	203	Master's degree	109
Ph. D.	86		

Table 14. Distribution of Subjects by Experiment Version (Experiment 1)

Version	#	*Version*	#
1	49	*5*	66
2	81	*6*	46
3	60	*7*	68
4	55	*8*	57

4.3.1 Outlier removal and response time adjustment

Before proceeding to the statistical analysis of results, we first discuss how outliers have been removed and how response times have been adjusted to accommodate for the delay caused by the jsPsych library used in the experiment:

All subjects with an accuracy below 90% (i.e. those who correctly responded to less than 85 of the 95 stimuli) have been removed from the results. Since 182 subjects performed below 90%, only 482 subjects out of the 664 who completed the experiment have been included in the results.

The overall accuracy of the 482 subjects to all stimuli of the experiment was 93.5% The twelve stimuli in Table 15 have been removed from the results because at least one of their lowercase or uppercase versions had an accuracy rate below 80%.

Table 15. Stimuli Removed from Experiment 1 Results due to Low Accuracy

Stimulus	Versions	Accuracy
antrepo	1 & 2	67%
ANTREPO	5 & 6	61%
danak	3 & 4	70%
DANAK	7 & 8	77%
gaya	3 & 4	74%
GAYA	7 & 8	71%
itrah	1 & 2	75%
İTRAH	5 & 6	76%
mene	1 & 2	71%
MENE	5 & 6	72%
şimlik	1 & 2	68%
ŞİMLİK	5 & 6	72%

Interestingly, the lowest-performing twelve stimuli (out of 380) only contain the uppercase and lowercase versions of the same six stimuli (*antrepo*, *danak*, *gaya*, *itrah*, *mene*, and *şimlik*). This finding will be discussed in Chapter 5.

In the next step, 27 ms has been deducted from all RT measurements based on the finding that the JavaScript technology used in the experiment results in a 26.8 ms delay on the average (de Leeuw & Motz, 2016).

For each subject, all RT measurements that are more than two standard deviations above or below the mean RT of that particular subject have been removed from the results.

Finally, the mean RTs of all subjects in a given version of the experiment have been recalculated following the above-mentioned removal of outlier RTs, and subjects whose mean RT is more than two standard deviations above or below the mean RT of all subjects who completed the same version have been excluded from the results. A total of 22 subjects have been removed in this way. Their distribution between the versions is shown in Table 16.

Table 16. Number of Subjects Removed from Experiment 1, by Version

Version	#	*Version*	#
1	2	*5*	4
2	4	*6*	2
3	2	*7*	3
4	1	*8*	4

4.3.2 Descriptive statistics

Tables showing descriptive statistics for the eight versions of Experiment 1 are available in Appendix N.

4.3.3 Pairwise comparison of the conditions

The mean RT values reported in the eight tables in Appendix N have been reproduced in Table 17 for convenience:

Table 17. Mean RT Values in the Five Conditions of Experiment 1

Version	Mean RT (LOW)	Mean RT (HIGH)	Mean RT (WUGGY)	Mean RT (SUFFIX)	Mean RT (RANDOM)
1	668	615	809	823	635
2	662	607	758	801	626
3	684	618	838	910	643
4	693	632	849	917	644
5	682	631	801	848	646
6	649	604	799	849	642
7	691	634	836	894	640
8	673	611	808	859	627

As can be seen in Table 17, the inequality in (1) holds for each of the eight versions of Experiment 1, without a single exception:

(1) *RT(SUFFIX) > RT(WUGGY) > RT (LOW) > RT (RANDOM) > RT(HIGH)*

4.3.4 Tests for normality

The hypothesis that the mean values of two experimental conditions are significantly different from each other can be tested using the "paired-samples" version of the t-test. In accordance with the normality assumption of this test, it should be checked in advance if the RT data in each of the conditions being compared is distributed "approximately" normally. The Shapiro-Wilk test has been used to test normality. The resulting p-values for the eight versions and five conditions of the experiment are shown in Table 18.

Table 18. p-values for the Shapiro-Wilk Normality Tests

Version	n	p-value (LOW)	p-value (HIGH)	p-value (WUGGY)	p-value (SUFFIX)	p-value (RANDOM)
1	48	0.064	0.062	0.011	0.013	0.074
2	79	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001
3	54	0.436	0.122	0.360	0.028	< 0.001
4	53	0.248	0.079	0.020	0.002	0.008
5	63	0.008	< 0.001	0.003	< 0.001	0.002
6	47	0.009	0.060	0.121	0.132	0.091
7	61	0.067	0.180	0.008	0.034	0.327
8	50	0.004	0.195	0.003	0.002	0.017

Since the null-hypothesis of the Shapiro-Wilks test is that the data is distributed normally, we reject the normality hypothesis for those experiment conditions where the p-value is below $\alpha = 0.05$ (confidence level 95%). As can be seen in Table 18, all eight versions contain at least one condition whose p-value is below 0.05. In other words, the distribution of RT data cannot be assumed to be normal in any of the test versions. Thus, using the t-test to see if there is a statistically significant difference between two conditions is not a feasible option.

4.3.5 Wilcoxon signed-rank tests

Considering the results of the Shapiro-Wilk Normality Test in Table 18, the Wilcoxon signed-rank test has been used to test the hypothesis that there is a statistically significant difference between mean RT differences for low-frequency and high-frequency words. The results are summarized in Table 19.

Table 19. Results of the Wilcoxon Test for Exp. 1 (HIGH vs. LOW)

Version	n	Test statistic	p-value
1	48	31.5	< 0.0000001
2	79	69.5	< 0.0000001
3	54	12.5	< 0.0000001
4	53	4.5	< 0.0000001
5	63	70.0	< 0.0000001
6	47	36.0	< 0.0000001
7	61	24.0	< 0.0000001
8	50	23.5	< 0.0000001

4.3.6 Significance, difference, and effect size

"Statistical significance" only tests the existence of an effect, but does not tell us anything about the size or importance of that effect. One might obtain an arbitrarily low p-value in a significance test, but this only tells us that there is an extremely small probability that the observed difference is purely due to chance.

A more meaningful analysis would be to look at the actual differences between the mean RTs obtained in the relevant two conditions. Although this is a better measure, it still does not tell us whether the observed effect is small or large, and also does not allow us to compare the effects observed in another pair of conditions, or in another experiment.

To determine the relative size of the observed effect, one can calculate the variable known as “effect size”, which is a standardized measure independent of the absolute values measured in the experiment. Effect size is traditionally reported by a measure known as “Cohen’s D” in the literature. A Cohen’s D value around 0.20 is traditionally interpreted as indicating a “small” effect, a value around 0.50 is assumed to point to a “medium” effect, while a value around 0.80 indicates a “large” effect.

P-values calculated for Experiment 1 have already been reported in Table 19, and show that there most probably exists an effect in all eight versions. Table 20 below supports this data with the mean differences between the HIGH and LOW conditions, and the Cohen’s D measure.

Table 20. Mean RT Differences and Effect Sizes for Exp. 1 (HIGH vs. LOW)

Version	Mean RT (LOW)	Mean RT (HIGH)	Mean Diff. (LOW - HIGH)	Effect size (Cohen's D)
1	668	615	53	0.50
2	662	607	55	0.51
3	684	618	66	0.74
4	693	632	61	0.59
5	682	631	51	0.45
6	649	604	45	0.47
7	691	634	57	0.49
8	673	611	62	0.63

4.3.7 Overview of the results of Experiment 1

Experiment 1 has produced extremely clear results: As mentioned in Section 4.3.3, the following inequality holds for all eight versions of the experiment, without a single exception:

RT(HIGH) < RT(RANDOM) < RT(LOW) < RT(WUGGY) < RT(SUFFIX)

In other words, high-frequency bare nouns like *duman* 'smoke' are the most rapidly processed among the stimuli in the five conditions, followed by random letter-sequences like *hdggıüok*, low-frequency bare nouns like *yosun* 'moss', phonotactically plausible non-words like *fekotan*, and finally, pseudo-suffixed phonotactically plausible non-words like *kürazarlık*, which are the slowest.

If we focus only on the HIGH and LOW conditions, which are the two conditions involved in Hypothesis 1, there is an extremely-statistically-significant difference[35] between the visual recognition times of the high-frequency words and the low-frequency words used in Experiment 1. Moreover, Table 20 shows that the observed effect is in the same direction for all versions (HIGH < LOW), and is between 45 and 66 milliseconds. More importantly, Table 20 also shows that this effect can be classified as a “medium-to-large” effect according to the traditional interpretation of the Cohen’s D measure.

These results clearly support Hypothesis 1, which claims that high-frequency words in Turkish are recognized faster than low-frequency words, everything else being equal, and thus serves as evidence that there is a medium- to large-sized frequency effect in the visual recognition of bare Turkish nouns.

This result is not very interesting in itself, though. The existence of strong frequency effects has been demonstrated countless times in the literature, for a wide range of typologically unrelated languages. The importance of the result lies in the facts that (a) this is, to our knowledge, the first experiment to demonstrate the existence of a frequency effect in Turkish; (b) all eight versions of the experiment have produced exactly the same result, thus demonstrating the validity and reliability of the general-purpose resources developed and the methodology used; (c) the experiment has been conducted online, achieving an unusually large sample-size of 664 subjects, thus demonstrating the usefulness of the software platform developed for this study.

[35] “Extremely-statistically-significant” means that the relevant p-value is less than 0.0001. Some researchers use four stars (****) to report such results (see, for example, http://www.graphpad.com/guides/prism/7/statistics/index.htm?extremely_significant_results.htm. Retrieved on 6 September 2016).

4.4 Experiment 2

The first experiment has identified a clear frequency effect in the visual processing of bare nouns in Turkish. The second experiment investigates the visual processing of morphologically complex Turkish nouns.

As in the first experiment, frequency is used as the independent variable, and response time as the dependent variable. But this time, the independent variable is the frequency of the suffix template rather than various measures of root frequency: Half of the real words consist of complex nouns made up of high-frequency suffix templates (e.g. *+Loc+Pres+A1sg*, which forms word-forms like *evdeyim* 'I am at home'*)*, and the other half consists of complex nouns made up of low-frequency suffix templates (e.g. *+P2pl+Pres+A1sg,* which forms word-forms like *hocanızım* 'I am your teacher'), keeping several other variables under control. The first hypothesis of Experiment 2 can be formulated as follows:

Hypothesis 1: The high-frequency group will be processed significantly faster than the low-frequency group. If the results support this hypothesis, and all other factors that may affect response times have been adequately controlled for, the only source of this difference can be the frequency of the suffix template. If response times are sensitive to the frequencies of the suffix templates, this means that suffix templates (rather than individual suffixes) are being used as effective processing units at some point of the word recognition process. This would constitute evidence that suffixes attached to nouns are not processed one-by-one but as chunks made up of several suffixes, as suggested by Frauenfelder & Schreuder (1992) and Durrant (2013). In other words, the brain of a native speaker of Turkish contains some representation of the suffix templates defined in Section 3.2.1.

Unlike the non-words in Experiment 1, the non-words in Experiment 2 have been designed for the purpose of testing the following additional hypothesis:

Hypothesis 2: Non-words made up of a pseudo-root and a real suffix template (e.g. *gansien+diler*) will be rejected faster than non-words made up of a real root and a pseudo-suffix-template (e.g. *bıldırcın+ganıluf*). This would constitute evidence that the parsing of complex Turkish words proceeds from left to right, as suggested by Hankamer (1989).

4.4.1 Design decisions

The following design decisions have been adopted to minimize confounds, maximize validity and reliability, and ensure maximum participation in the online experiment:

4.4.1.1 Zero surface-frequency

The complex word-forms used in the experiment do not occur in the BOUN Corpus. Thus, the subjects will most probably be exposed to the stimuli of Experiment 2 for the first time in their lives. In other words, surface-frequency has been controlled for by being eliminated from the picture. This design allows us to avoid the full-listing / whole-form route altogether, and study the parsing / decomposition route in isolation.

4.4.1.2 Nouns only

Experiment 2 has been limited to nouns, considering that verbs are subject to a separate set of rules in terms of inflection and derivation, and are morphologically, syntactically and semantically more complex than nouns. Another consideration is that the frequency effect identified in Experiment 1 was obtained for nouns.

4.4.1.3 Only animals, tools, and feelings as roots

For the purpose of eliminating root-related (mainly semantic) confounds to the maximum extent possible, the stimuli of Experiment 2 have been generated by combining the suffix templates with names of animals in Set 1, with names of edible plants in Set 2, and with names of tools and household items in Set 3.

This design ensures that stimuli within the same set have similar semantic features, thus reducing the impact of any semantic confounds within the set, while stimuli across sets are as semantically different as possible, thus making it intentionally more difficult to obtain similar results from the three sets, and thus increasing validity.

4.4.1.4 Zero neighbors

All roots used in the experiment have zero orthographic neighbors. In other words, none of the roots can be transformed into another valid Turkish word by adding, deleting or changing a single letter. This design decision is aimed at controlling orthographic confounds to a certain extent. The roots shown in Appendix K have been selected as a result.

4.4.1.5 Use roots twice

In an attempt to further eliminate root-related confounds, each of the animal, plant and tool names mentioned above have been used twice in the respective real-word conditions, once attached to a low-frequency suffix template and once attached to a high-frequency suffix template. In other words, the high-frequency condition and the low-frequency condition use the same set of roots.

4.4.1.6 Avoid the accusative-possessive ambiguity

To reiterate, the accusative-possessive ambiguity described in Section 3.3.3 affects nouns ending in a consonant: The inflected word *insanı* (= *insan+ı*), for example, is ambiguous in that the second morpheme can be interpreted either as an accusative marker or a possessive marker, giving rise to the two parses *insan+Acc* and *insan+P3sg*. To avoid this ambiguity, roots ending in a vowel have been preferred whenever the relevant suffix template begins with the compound marker *-(s)I* (*P3sg*). For example, the vowel-ending root *kanguru* 'kangaroo', rather than the consonant-ending root *akrep* 'scorpion', has been preferred for the suffix template *+P3sg+Loc+While,* resulting in the unambiguous word-form *kangurusundayken*, instead of the (at least temporarily) ambiguous word-form *akrebindeyken.*

4.4.1.7 Two or three suffixes

The stimuli selected for Experiment 2 initially contained between one and seven suffixes. However, an informal pilot study using these stimuli has shown that most subjects took a very long time to finish the experiment (up to 15 minutes, compared to approximately 4 minutes in Experiment 1), and reported that the words were 'too long and difficult'. With this anecdotal evidence in mind, and also in view of the increased difficulty of matching for the control variables when the stimuli contain anything between zero and seven suffixes, the maximum number of suffixes has been forced to be three. Indeed, a second pilot study conducted after the introduction of this limitation has shown that the subjects no more complained about the difficulty and length of the stimuli, and finished the experiment within a time period comparable to Experiment 1.

4.4.1.8 Avoid *Agt*, *Ness*, *With*, *Without*

As mentioned in Section 3.4.12, the root dictionary of the morphological disambiguator is inconsistent in its treatment of derived forms: It arbitrarily treats some derived forms as complex stems and some others as simplex roots, although the two have identical structures (e.g. *telsizcilik* 'the profession of a radio operator' and *oyunculuk* 'the profession of an actor'). This problem especially affects the highly productive derivational suffixes *Agt*, *Ness*, *With* and *Without*.

These four derivational suffixes have been avoided while selecting suffix templates for Experiment 2, since they are involved in a large number of lexicalizations with varying degrees of opacity, and are thus more likely to appear more frequently than other suffixes in the disambiguator's imperfect "root dictionary".

In fact, of the 10,414 roots removed from the dictionary, 3,215 are combinations of these 4 derivational suffixes. The breakdown is shown in Table 21.

Table 21. Stems That Contain Agt, Ness, With and Without

Form	# of types
R+Ness	947
R+Agt+Ness	888
R+Agt	805
R+Without+Ness	409
R+With	85
R+With+Ness	81
TOTAL	3,215

4.4.1.9 Same suffix max. five times

The algorithm that selects matched suffix-templates for the two conditions of Experiment 2 initially did not involve any limitation as to the maximum number of times a given experimental condition contains a given suffix. In other words, nothing prevented the algorithm from selecting, say, fifteen of the twenty suffix-templates in a condition from among templates that contain the suffix *Acquire*. This would mean that an experimental condition could be dominated by a single suffix, along with its orthographic, phonological, syntactic and semantic idiosyncrasies.

To prevent domination by any one suffix, the following additional requirement has been added to the template matching algorithm: the same suffix can appear at most five times among the twenty templates that constitute an experimental condition. This upper limit has been determined experimentally, and reflects the trade-off between the desire to minimize the number of occurrences of a single suffix in the same condition, and the need to have a sufficient number of candidates that can be used as stimuli.

4.4.1.10 Maximum difference three

The requirement described in Section 4.4.1.9 solves the domination problem to a certain extent, but there is another problem: What if a suffix occurs five times in the high-frequency group, but only once in the low-frequency group, for example? The orthographic, phonological, syntactic and semantic idiosyncrasies of that suffix would have a five-fold impact on the high-frequency group compared to the low-frequency group.

To minimize this confounding effect, the following additional requirement has been added to the template–matching algorithm: the difference between the number of times a given suffix occurs in an experimental condition and the number of times the same suffix occurs in the other experimental condition has been limited to three. Once again, this upper limit has been determined experimentally, and reflects the trade-off between the desire to minimize said difference, and the need to have a sufficient number of candidates that can be used as stimuli.

4.4.1.11 Only 40 stimuli

20 + 20 = 40 real-word stimuli have been used in Experiment 2, compared to the 25 + 25 = 50 in Experiment 1. The reason for this 20% reduction is that the stimuli used in Experiment 2 are much more complex than those used in Experiment 1. This poses the risk of considerably increasing the time it would take subjects to complete the experiment, thus increasing the rate of dropouts, i.e. unfinished experiments. Obviously, using fewer stimuli reduces the statistical power of the experiment.

4.4.1.12 No fillers

As in Experiment 1, no filler items have been used in Experiment 2. Once again, this decision will probably increase participation in the experiment by making it considerably shorter, but will also make it easier for subjects to guess the purpose of the experiment, thus possibly reducing the validity of the results.

4.4.1.13 Limit child frequency

The correlation coefficients between the independent variable (suffix-template frequency) and the sixteen remaining template variables described in Section 3.2.4 are shown in Table 22.

Table 22. Correlation Btw. Independent Var. and Control Variables in Exp. 2

Parameter	Corr.	Parameter	Corr.
parent_count	-0.03	*bi_count*	-0.08
parent_freq	-0.01	*bi_mean*	+0.04
child_count	+0.88	*tri_count*	-0.07
child_freq	+0.86	*tri_mean*	+0.10
sib_count	+0.05	*inf_count*	+0.05
sib_freq	+0.14	*drv_count*	-0.03
uni_count	-0.07	*blocking*	-0.03
uni_mean	+0.01	*length*	-0.06

As can be seen in Table 22, the independent variable (suffix-template frequency) is highly correlated with the control variables *child_count* (number of child nodes) and *child_freq* (total frequency of child nodes). This makes it difficult to match the two stimulus sets for the two control variables in question. To control for *child_count* and *child_frequency*, the matching algorithm uses the additional requirement that the ratio of *child_freq* to template frequency does not exceed 0.20. This ratio has been determined experimentally: The lower the ratio, the better child frequency will be controlled for. On the other hand, the lower the ratio, the fewer stimulus candidates will remain. The ratio 0.20 establishes an experimentally determined balance between the desire to control for a variable that is highly correlated with the independent variable and the need to find a sufficient number of stimuli.[36]

[36] To clarify, there is a very high positive correlation between the number of times the corpus contains, for example, the suffix template *+P3sg+Loc+Rel+A3pl* (which generates word-forms such as *arabasındakiler*), and the number of "child-templates" that begin with the same template (e.g. in word-forms like *arabasındakilerle*, *arabasındakilermişcesine*), as well as the total number of times such child-templates occur in the corpus. This correlation is probably just another expression of the correlation demonstrated between roots, inflected-forms, derived-forms and compound-forms in Section 3.4.13.

4.4.1.14 Control variables

Stimuli in the two conditions have been matched for the 16 template parameters described in Section 3.2.4, as well as the mean letter-bigram and letter-trigram frequencies of the root-template combination. Table 23 shows the relevant means and standard deviations across the two conditions of the experiment.

Table 23. Means and Std. Deviations in Experiment 2 (HIGH & LOW)

		ARITHMETIC MEAN		STD. DEVIATION	
Variable		High Freq.	Low Freq.	High Freq.	Low Freq.
Template Frequency*		12,784	1,426	10,783	701
Parent Count		1.95	1.95	0.67	0.67
Total Parent Frequency		71,529,909	72,015,072	6,356,508	6,367,672
Child Count		2.90	1.90	3.33	2.53
Total Child Frequency		188,500	56,400	398,984	92,125
Sibling Count		11.55	10.80	6.48	6.69
Total Sibling Frequency		1,597,382	1,541,580	4,369,072	4,396,925
Template Unigram Count		2.65	2.65	0.48	0.48
Mean Template Unigram Frequency		10,529,172	9,910,446	8,449,962	8,651,270
Template Bigram Count		3.65	3.65	0.48	0.48
Mean Template Bigram Frequency		5,033,371	4,069,153	3,938,471	3,769,465
Template Trigram Count		2.65	2.65	0.48	0.48
Mean Template Trigram Frequency		272,502	257,494	551,867	561,769
Inflectional Suffixes		2.00	2.00	0.71	0.77
Derivational Suffixes		0.50	0.50	0.74	0.74
Blocking Position		0.90	0.90	0.54	0.62
ANIMAL	Mean Letter Bigram Freq.	11,832,965	10,914,101	3,536,054	3,431,335
	Mean Letter Trigram Freq.	1,914,670	1,695,827	890,090	730,542
	Root+Template Length	13.50	13.35	2.46	1.90
PLANTS	Mean Letter Bigram Freq.	11,547,812	11,112,942	2,778,426	2,790,900
	Mean Letter Trigram Freq.	1,990,518	1,895,101	808,445	687,512
	Root+Template Length	13.50	13.45	2.04	1.60
TOOLS	Mean Letter Bigram Freq.	11,999,300	11,112,428	2,751,607	2,834,529
	Mean Letter Trigram Freq.	2,139,989	1,920,523	835,091	606,650
	Root+Template Length	13.10	13.00	2.66	2.17

* = independent variable

4.4.1.15 Minimize interpretability

The roots listed in Appendix K have been combined with the suffix templates listed in Appendix L manually. This is the only manual step in the generation of stimuli for Experiment 1 and Experiment 2. Although we must admit that this manual matching involves a certain risk of biased stimulus selection, best efforts have been made to avoid any bias. The root-suffix template combinations have been chosen such that the resulting word form is as uninterpretable as possible. For example, the suffix-template *+Acquire+Pass+Past* would form a much more interpretable word-form when combined with the root *poşet* 'bag' (i.e. *poşetlenildi* 'it has been bagged'), compared to the root *kiremit* 'roof tile' (i.e. *kiremitlenildi* '?it has been roof-tiled'). This is why *kiremit* 'roof tile' has been preferred and *poşet* 'bag' avoided for this particular suffix-template. The resulting stimulus sets are shown in Appendix L

4.4.1.16 Generation of non-words

Experiment 2 contains 45 non-words in the following three categories, which have already been briefly described in Section 4.4:

(1) The 'PP' condition: 15 non-words where a pseudo-root is combined with a pseudo-suffix-template (e.g. *kındün+matgadan* based on *kunduz+lanmadan*);

(2) the 'PR' condition: 15 non-words where a pseudo-root is combined with a real suffix template (e.g. *gansien+diler* based on *penguen+diler*);

(3) the 'RP' condition: 15 non-words where a real root is combined with a pseudo-suffix-template (e.g. *bıldırcın+ganıluf* based on *bıldırcın+lanılır*).

All pseudo-roots and pseudo-suffix-templates have been generated by Wuggy, based on real roots and real suffix templates used in the experiment.

4.4.1.17 Modified instructions

As revealed by an informal pilot study, standard lexical decision instructions prove inadequate in Experiment 2. Since all "real words" used in this experiment are novel combinations with zero surface-frequency, subjects are confused when they are simply told to press 'yes' if what they see on the screen is a "valid word in Turkish", because they tend to think that a word is only valid when they have been exposed to it before, or at least when they can make sense of it. To avoid this confusion, the (inevitably complicated and long) instructions in Appendix M have been used.

4.4.2 The twelve version of Experiment 2

In accordance with the seventeen design decisions described in Section 4.4.1, Experiment 2 has the following 3 x 2 x 2 = 12 versions, and the computer randomly chooses one of them as the subject presses the start button:

Version 1: Animals as roots, lowercase letters, 'yes' button on left

Version 2: Animals as roots, lowercase letters, 'yes' button on right

Version 3: Plants as roots, lowercase letters, 'yes' button on left

Version 4: Plants as roots, lowercase letters, 'yes' button on right

Version 5: Tools as roots, lowercase letters, 'yes' button on left

Version 6: Tools as roots, lowercase letters, 'yes' button on right

Version 7: Animals as roots, uppercase letters, 'yes' button on left

Version 8: Animals as roots, uppercase letters, 'yes' button on right

Version 9: Plants as roots, uppercase letters, 'yes' button on left

Version 10: Plants as roots, uppercase letters, 'yes' button on right

Version 11: Tools as roots, uppercase letters, 'yes' button on left

Version 12: Tools as roots, uppercase letters, 'yes' button on right

4.4.3 Predictions

The two hypotheses described in Section 4.3 have been repeated here for convenience:

Hypothesis 2: Everything else being equal, the time it takes a native speaker of Turkish to visually recognize a morphologically complex Turkish noun decreases as the frequency of the suffix-template increases.

Hypothesis 3: Everything else being equal, the time it takes a native speaker of Turkish to reject a non-word that starts with a meaningless letter-sequence but ends with a valid sequence of suffixes (e.g. *gansien+lerinizin*) is shorter than the time it takes a native speaker of Turkish to reject a non-word that starts with a valid root but ends with meaningless letters (e.g. *bıldırcın+ganfirmaş*).

In other words, Hypothesis 2 predicts that mean RTs in the 'HIGH' condition will be significantly lower than mean RTs in the 'LOW' condition in all 12 versions of Experiment 2. Similarly, Hypothesis 3 predicts that mean RTs in the 'PR' condition will be significantly lower than mean RTs in the 'RP' condition in all 12 versions of Experiment 2.

4.4.4 Results of Experiment 2

A total of 1,996 subjects have completed Experiment 2. In other words, 1,996 x 70 = 139,720 response time measurements have been collected.[37] To the best of our knowledge, Experiment 2 is by far the largest-scale psycholinguistic experiment on Turkish so far.

[37] As in Experiment 1, we cannot claim that the experiment has been completed by 1,996 different individuals since the experiment was conducted online and no measures have been taken to prevent the same person from completing the experiment more than once. The same uncertainty also applies to the demographic information provided by the subjects.

As will be discussed in Section 4.4.4.1 below, 578 of the subjects have been removed from the results since their accuracy was below 90%. Thus, demographic data will be reported only for the remaining 1,418 subjects:

(1) In 1,201 of the trials (84.7%), the subject reported to be right-handed, and in 217 trials (15.3%) left-handed.

(2) In 729 trials (51.4%), the subject reported to be male and in 665 (46.9%) female, while in 24 cases (1.7%) the subject preferred not to specify any gender.

(3) In 1,223 trials (86.3%), the subject reported to speak only one native-language (Turkish), and in 195 trials (13.7%) more than one native-language including Turkish.

Figure 12 shows the distribution of subjects by age, Table 24 the distribution of subjects by education level, and Table 25 the distribution of subjects between the experiment's twelve versions.

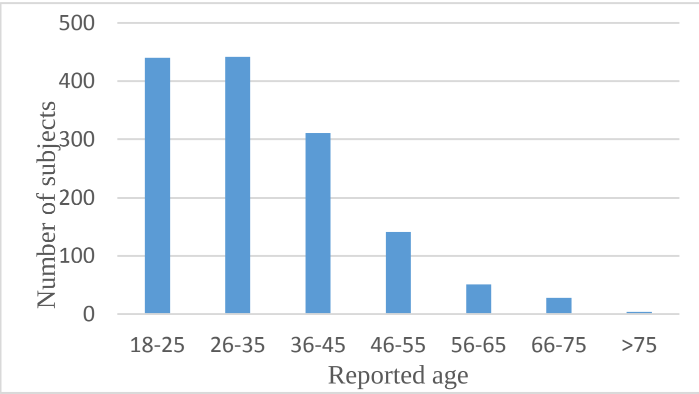

Figure 12. Distribution of subjects by age (Experiment 2)

Table 24. Distribution of Subjects by Education Level (Experiment 2)

Primary-school	59	Secondary-school	18
High-school	141	Pre-graduate	64
Undergraduate	680	Master's degree	318
Ph. D.	138		

Table 25. Distribution of Subjects by Experiment Version (Experiment 2)

Version	#	*Version*	#	*Version*	#	*Version*	#
1	123	*4*	119	*7*	107	*10*	123
2	114	*5*	103	*8*	139	*11*	105
3	154	*6*	115	*9*	122	*12*	94

4.4.4.1 Outlier removal and response time adjustment

Before moving on to the statistical analysis of results, we first discuss how outliers have been removed and how response times have been adjusted to accommodate for the delay caused by the jsPsych library used in the experiment:

All subjects with an accuracy below 90% (i.e. those who correctly responded to less than 63 of the 70 stimuli) have been removed from the results. Since 578 subjects performed below 90%, only 1,418 subjects out of the 1,996 who completed the experiment have been included in results. Overall accuracy of the 1,481 subjects to all stimuli of the experiment is 96.6%.

The eight stimuli in Table 26 have been removed from the results because at least one of their lowercase or uppercase versions had an accuracy rate below 80%.

Table 26. Stimuli Removed from Experiment 2 Results due to Low Accuracy

Stimulus	Versions	Accuracy
pırasayadır	3 and 4	63%
PIRASAYADIR	9 and 10	73%
zımbasınca	5 and 6	64%
ZIMBASINCA	11 and 12	65%
kurbağasınca	1 and 2	79%
KURBAĞASINCA	7 and 8	82%
kelesaldırlar	1 and 2	78%
KELESALDIRLAR	7 and 8	82%

Interestingly, the lowest-performing eight stimuli (out of 420) only contain the uppercase and lowercase versions of the same four stimuli (*pırasayadır*, *zımbasınca*, *kurbağasınca*, and *kelesaldırlar*). This finding will be discussed in Chapter 5.

In the next step, 27 ms has been deducted from all RT measurements based on the finding that the JavaScript technology used in the experiment results in a 26.8 ms delay on the average (de Leeuw & Motz, 2016).

For each subject, all RT measurements that are more than two standard deviations above or below the mean RT of that particular subject have been removed from the results.

Finally, the mean RTs of all subjects in a given version of the experiment have been recalculated following the above-mentioned removal of outlier RTs, and subjects whose mean RT is more than two standard deviations above or below the mean RT of all subjects who completed the same version have been excluded from the results. A total of 58 subjects have been removed in this way. Their distribution between the version is shown in Table 27.

Table 27. Number of Subjects Removed by Version (Experiment 2)

Version	#	*Version*	#	*Version*	#	*Version*	#
1	3	*4*	4	*7*	3	*10*	10
2	5	*5*	3	*8*	4	*11*	4
3	10	*6*	5	*9*	4	*12*	3

4.4.4.2 Descriptive statistics

Tables showing descriptive statistics for the twelve versions of Experiment 2 are available in Appendix O.

4.4.4.3 Pairwise comparison of the conditions

The mean RT values reported in the twelve tables in Appendix O have been summarized in Table 28 for convenience:

Table 28. Mean RT Values in the Twelve Conditions of Experiment 2

Version	Mean RT (HIGH)	Mean RT (LOW)	Mean RT (PP)	Mean RT (RP)	Mean RT (PR)
1	1238	1305	1396	1485	1294
2	1196	1272	1341	1447	1266
3	1145	1218	1318	1463	1386
4	1251	1300	1392	1546	1433
5	1327	1375	1382	1579	1547
6	1139	1215	1228	1436	1395
7	1327	1370	1520	1556	1405
8	1301	1365	1512	1546	1353
9	1204	1254	1414	1565	1478
10	1280	1342	1461	1681	1562
11	1330	1401	1384	1556	1530
12	1375	1446	1446	1576	1546

As an analysis of Table 28 reveals, the following two inequalities hold for each of the twelve versions of Experiment 2, without a single exception:

(2) *RT(HIGH) < RT(LOW)*

(3) *RT(PR) < RT(RP)*

Moreover, if we assume that a pairwise inequality between two conditions is valid if it is supported by at least eight of the twelve versions, we can assert the more ambitious inequality in (3)[38]:

[38] PP < PR is supported by eight versions, LOW < PR by nine versions, and LOW< PP by ten versions. The remaining seven pairwise inequalities are supported by all twelve versions of the experiment.

(4) $RT(H) < RT(L) < RT(PP) < RT(PR) < RT(RP)$

4.4.4.4 Tests for normality

As in Experiment 1, the Shapiro-Wilk test has been used to test normality. The resulting p-values for the twelve versions and five conditions of Experiment 2 are shown in Table 29.

Table 29. p-values for the Shapiro-Wilk Normality Tests (Experiment 2)

Version	n	p-value (HIGH)	p-value (LOW)	p-value (PP)	p-value (RP)	p-value (PR)
1	119	0.004	0.078	0.001	0.110	< 0.001
2	109	0.075	0.133	0.034	0.311	0.010
3	144	0.030	0.528	0.003	0.256	0.003
4	114	0.003	0.016	0.005	0.012	0.006
5	98	0.053	0.102	0.014	0.502	0.026
6	109	0.004	0.008	< 0.001	0.009	< 0.001
7	101	0.088	0.040	0.003	0.038	0.008
8	134	0.040	0.017	0.019	0.020	0.005
9	116	0.118	0.133	0.015	0.054	0.006
10	112	0.342	0.439	0.507	0.075	0.038
11	101	0.012	0.043	0.001	0.056	0.006
12	90	0.037	0.445	0.050	0.046	0.038

Since the null-hypothesis of the Shapiro-Wilks test is that the data is distributed normally, we should reject the normality hypothesis for those experiment conditions where the p-value is below $\alpha = 0.05$ (confidence level 95%). As can be seen in Table 29, all twelve versions contain at least one condition whose p-value is below 0.05. In other words, the distribution of RT data cannot be assumed to be normal in any of the test versions. Thus, using the paired-samples t-test to see if there is a statistically significant difference between mean RTs in two conditions is not a feasible option.

4.4.4.5 Wilcoxon signed-rank tests

Considering the results of the Shapiro-Wilk normality test in Table 29, the Wilcoxon signed-rank test has been used to test the hypothesis that there is a statistically significant difference between mean RTs in two conditions of the experiment. The first analysis is for the “low-frequency suffix template” and “high-frequency suffix template” conditions (‘LOW’ vs. ‘HIGH’), whose results are summarized in Table 30. The second analysis is for the “pseudo root + real suffixes” and “real root + pseudo suffixes” conditions (‘PR’ vs. ‘RP’), whose results are summarized in Table 31.

Table 30. Results of the Wilcoxon Test for Experiment 2 (HIGH vs. LOW)

Version	n	Test statistic	p-value
1	119	1604.5	< 0.001
2	109	1518.0	< 0.001
3	144	1924.0	< 0.001
4	114	1765.5	< 0.001
5	98	1505.5	< 0.001
6	109	1123.0	< 0.001
7	101	1605.0	< 0.001
8	134	2444.0	< 0.001
9	116	1993.5	< 0.001
10	112	1612.0	< 0.001
11	101	1155.0	< 0.001
12	90	952.0	< 0.001

Table 31. Results of the Wilcoxon Test for Experiment 2 (PR vs. RP)

Version	n	Test statistic	p-value
1	119	650.5	< 0.0001
2	109	689.0	< 0.0001
3	144	3162.5	< 0.0001
4	114	1629.5	< 0.0001
5	98	2139.0	0.310
6	109	2299.0	0.048
7	101	1014.0	< 0.0001
8	134	1014.0	< 0.0001
9	116	2006.5	< 0.0001
10	112	1719.5	< 0.0001
11	101	2150.0	0.197
12	90	1723.0	0.192

4.4.4.6 Effect sizes

As in Experiment 1, Table 32 summarizes the mean difference between HIGH and LOW conditions, and the Cohen's D measure calculated for each version of the experiment. Table 33 does the same thing for the PR and RP conditions.

Table 32. Mean RT Differences and Effect Sizes for Exp. 2 (HIGH vs. LOW)

Version	Mean RT (HIGH)	Mean RT (LOW)	Difference (ms)	Effect size (Cohen's D)
1	1238	1305	67	0.17
2	1196	1272	76	0.24
3	1145	1218	73	0.22
4	1251	1300	49	0.14
5	1327	1375	48	0.12
6	1139	1215	76	0.22
7	1327	1370	43	0.12
8	1301	1365	64	0.19
9	1204	1254	50	0.16
10	1280	1342	62	0.24
11	1330	1401	71	0.24
12	1375	1446	71	0.18

Table 33. Mean RT Differences and Effect Sizes for Exp. 2 (RP vs. PR)

Version	Mean RT (RP)	Mean RT (PR)	Difference (ms)	Effect size (Cohen's D)
1	1485	1294	191	0,43
2	1447	1266	181	0,50
3	1463	1386	77	0,19
4	1546	1433	113	0,26
5	1579	1547	32	0,07
6	1436	1395	41	0,09
7	1556	1405	151	0,33
8	1546	1353	193	0,47
9	1565	1478	87	0,20
210	1681	1562	119	0,34
11	1556	1530	26	0,07
12	1576	1546	30	0,07

4.5 Results of Experiment 2

Experiment 2 has produced extremely clear results that fully support Hypothesis 2, and relatively clear results that support Hypothesis 3 to a certain extent. As mentioned in Section 4.4.4.3, the following two inequalities hold for all twelve versions of the experiment, without a single exception:

(2) *RT(HIGH) < RT(LOW)*

(3) *RT(PR) < RT(RP)*

In terms of Hypothesis 2, there is an extremely-statistically-significant difference between the visual recognition times of stimuli that contain high-frequency suffix templates and stimuli that contain low-frequency suffix templates. Moreover, Table 32 shows that the observed effect is in the same direction for all versions (HIGH < LOW), and is between 43 and 76 milliseconds. Table 32 also shows that this effect can be classified as a "small" effect according to the traditional interpretation of the Cohen's D measure.

These results clearly support Hypothesis 2, which claims that words containing high-frequency suffix templates are processed faster than words containing low-frequency suffix templates, everything else being equal. This supports the idea that there exist separate mental representations for frequently occurring suffix sequences (not individual suffixes but entire suffix bundles such as *lAş+DIr+Il+mIş*). This is in line with "usage-based" accounts of grammar, which claim that linguistic structure emerges from language use, i.e. from repeated exposure to certain constructions over time (Bybee, 2011, p. 69).

In terms of Hypothesis 3, on the other hand, in eight of the twelve versions of the experiment, there is an extremely-statistically-significant difference between the visual recognition times of non-words where a pseudo-root is combined with a real suffix template (e.g. *gansien+diler*) and non-words where a real root is combined with a pseudo-suffix-template (e.g. *bıldırcın+ganıluf*). Moreover, Table 33 shows that the observed effect is in the same direction for all versions (PR < RP), and is between 26 and 193 milliseconds. Table 33 also shows that this effect can be classified as a "small- to medium-sized" effect according to the traditional interpretation of the Cohen's D measure.

The only problem is that the Wilcoxon test results for experiment versions 5, 6, 11 and 12 do not support Hypothesis 3, as the relatively high p-values in Table 33 show. Interestingly, these four are the only versions that use the "names of tools" set in Appendix K as roots. Several analyses on stimulus accuracies, subject accuracies, descriptive statistics, and sample size have been performed to understand the root cause of this problem, but no conclusion could be reached. The only thing that can be said is that there seems to be something wrong with the “names of tools” set used exclusively in these four versions, and this could very well have a semantic reason.

Apart from this problem, the results clearly support Hypothesis 3, which claims that the time it takes a native speaker of Turkish to reject a non-word that starts with a meaningless letter-sequence but ends with a valid sequence of suffixes (e.g. *gansien+lerinizin*) is shorter than the time it takes a native speaker of Turkish to reject a non-word that starts with a valid root but ends with meaningless letters (e.g. *bıldırcın+ganfirmaş*). This is in line with Hankamer’s (1989) claim that Turkish words are processed from left to right.

CHAPTER 5

GENERAL DISCUSSION

The experiments have provided statistical support for all three hypotheses defined at the outset. We now have some evidence that (a) more frequent simple words are processed faster than less frequent simple words, (b) complex words are probably parsed from left to right, and most importantly, (c) there probably exist mental representations for frequently occurring morpheme sequences.

Before discussing the details of these three main findings, however, let us first report "secondary findings".

5.1 Secondary findings

Several secondary findings have been obtained more or less incidentally during the course of this study, especially during the preparatory work described in Chapter 3. The following sections describe these findings.

5.1.1 Frequency is complex

"Word frequency" is a deceptively simple concept, especially when dealing with isolating languages like English. The true extent of the complexity of frequency becomes apparent in an agglutinating language like Turkish.

This study started with the few frequency metrics available in existing literature, e.g. surface frequency, base frequency, and family frequency. However, it soon became apparent that these metrics were not adequate for describing Turkish data. The end result has been a general notation that can be used to define around 90 different measures of word frequency (see Section 3.2.2 for details). Although the proposed notation might need to be modified in several ways, this is, to our knowledge, the first attempt to define the frequencies of Turkish words at this level of detail.

5.1.2 Frequency measures are highly correlated

As has been visually demonstrated in Figure 7, there is a high correlation between the number of times a noun occurs in its bare form, in inflectional forms, derivational forms, and *-(s)I* compound forms. Another correlation that has been discovered is the very high correlation between the newly-defined measure of template frequency, on the one hand, and the number and total frequency of that template's children, on the other hand. This second correlation is probably just another appearance of the first correlation at the stem level.

Any study that uses word frequencies as a variable must duly take into account these collinearity issues, because they have important implications in terms of experiment design, statistical analysis, and interpretation of results.

5.1.3 Distributions are extremely uneven

This study has empirically demonstrated that frequency distributions are extremely uneven at several levels, whether for letters, roots, suffixes, or their ngrams (see Appendices 2-8 for the relevant data). For example, although 7,733 nominal suffix templates are attested in the BOUN Corpus and many more are grammatically possible, the most frequent 200 nominal suffix templates listed in Appendix H account for 99.3% of all word-forms in the same corpus.[39]

[39] Although not included in this study, the picture is probably similar at sub-lexical levels like syllables and syllable sequences, as well as supra-lexical levels like verb subcategorization preferences.

One of the implications for future research is that, standard distributions in statistics literature (normal, binomial, exponential, beta, chi-square, etc.), and any tests and tools that assume such distributions, cannot be used when dealing with language data.[40]

Another implication is that experimental stimuli must be selected with the utmost care, considering that the frequency differences between seemingly similar items can reach several orders of magnitude. For example, it sounds quite natural to assume that the obligative marker *-mAlı* and the aorist marker *-(A/I)r/-z* have similar frequency distributions, since both occupy the same slot of the Turkish verb (see Göksel & Kerslake, 2005, p. 73). However, as can be seen in Appendix E, the aorist occurs 5,184,027 times in the BOUN Corpus, while the obligative occurs only 237,551 times, which means there is almost a 22-fold difference between their frequencies. In other words, stimuli in psycholinguistic experiments should not be selected simply based on grammatical properties; distributional properties that reflect language use must be taken into account.

5.1.4 Most grammatical forms are never used

This is probably the most interesting secondary finding of this study. If the calculation in Hankamer (1989) is correct, 9,192,472 forms can be generated from one noun root. In other words, the morphology of Turkish allows the generation of more than 9 million nominal suffix templates. When one counts the suffix templates in the BOUN Corpus, however, only 7,733 nominal suffix templates are attested. Considering that this corresponds to 0.00084% of all possible suffix templates, the

[40] Lexical frequency distributions are better described by a type of distribution know as LNRE (Large Number of Rare Events) (see, for example, Khmaladze, 1988).

conclusion is that 99.99916% of possible suffix templates are never used in a corpus of almost 300 million words.[41]

This empirical finding has important implications for the mental representation and processing of agglutinating languages like Turkish: The brain of a native speaker of Turkish who has been exposed to each and every word that occurs at least once in the BOUN Corpus does not need to store more than 9 million suffix templates, but only 23,346.

When combined with the "uneven distributions" finding in Section 5.1.3, things get even more interesting: Since these 23,346 attested suffix templates are themselves distributed in an extremely uneven fashion, a native speaker can process the overwhelming majority of the word-forms he/she encounters simply by storing the most frequent few hundred suffix templates as undecomposed full-form entries.

These empirical findings falsify Hankamer's (1989) conclusion that, since the human brain does not have enough capacity to store all possible forms, full-listing is impossible in an agglutinating language like Turkish. Storing a few hundred forms is clearly within the capacity of the human brain, and these few hundred forms can efficiently deal with the majority of the word-forms people encounter in their daily lives.

These findings are also in line with "usage-based" accounts of grammar, which claim that linguistic structure emerges from language use, i.e. from repeated exposure to certain constructions over time (Bybee, 2011, p. 69).

[41] An analysis of word-forms instead of suffix templates provides an even more striking picture: The BOUN corpus contains 1,236,526 unique word-forms. When compared to the 200 billion forms proposed by Hankamer (1989), it can be concluded that 99.999996% of possible word-forms are never used in a corpus of almost 300 million words.

5.1.5 Morphology dominates even sub-lexical distributions

According to Appendix D, the most frequent 15 middle-position letter-trigrams of Turkish are the following: *lar*, *ler*, *eri*, *arı*, *bir*, *ara*, *nda*, *ile*, *ası*, *lan*, *ini*, *ını*, *rin*, *yor*, *ınd*. In other words, these are the most frequent letter combinations to occur in the middle of word-forms.

Except for the letter sequences *b-i-r* and *i-l-e*, which are identical to the function-words *bir* ‘one, a’ and *ile* ‘with’, all letter-triples on this list look like the inflectional and derivational suffixes of Turkish: *l-e-r* and *l-a-r* correspond to the two variants of the plural marker *+A3pl*; *e-r-i* and *a-r-ı* are identical to the last three letters of the frequent suffix sequence *+A3pl+Acc* or *+A3pl+P3sg*; *y-o-r* corresponds to the only variant of the present-tense marker *+Prog1*, etc. This means that the morphology of the language has a deep impact even at a sub-lexical level such as the distribution of letter-triples.

5.1.6 Online experiments are feasible

This study has also demonstrated the feasibility of online psycholinguistic experiments. To the best of our knowledge, the two experiments described in Chapter 4 are the most widely-participated psycholinguistic experiments on Turkish so far.

The main benefit of an online experiment is that it can collect large amounts of data in a short time. In fact, the two experiments reported here have collected hundreds of thousands of RT measurements from more than 2,500 people within a couple of weeks. The availability of such a large sample also allows the experimenter to design complex experiments that consist of several versions, each probing another aspect of the experimental task, and to use more than one stimulus set, each constituting a replication of the others, thus improving the validity and reliability of the results.

5.1.7 Letter shape plays no role in lexical decision

Both experiments in this study had a factorial design in that one of the versions used lowercase stimuli while the other used the uppercase versions of the same stimuli. The purpose of this design was to test if the visual shapes of letters played a role in lexical decision.

Although a formal statistical analysis of this parameter has not been included in this study, an informal examination shows that the uppercase and lowercase versions have similar results in terms of response times and recognition accuracies. The recognition accuracies of individual stimuli, in particular, support an interesting finding: As mentioned in Section 4.3.1 and Section 4.4.4.1, the lowest-performing items of both experiments consist of the uppercase and lowercase versions of the very same stimuli. In other words, subjects had a very low accuracy in recognizing these stimuli, regardless of their physical shape. This coordinated behavior suggests that this low performance was caused by the inherent, abstract properties of the stimuli, rather than the physical shapes of the letters that make them up.

Although we are usually unaware, there exists almost no visual similarities between several lowercase and uppercase letters of the Latin alphabet (e.g. a-A, b-B, d-D, e-E, g-G, l-L, r-R). According to Dehaene (2009), the very early stages of visual word recognition involve purely visual pattern recognition, where the exact physical shape of the letters does matter. However, "the discrepancy between uppercase and lowercase letters. . . ceases to matter at an early stage in the visual stream" (Dehaene, 2009).

The accuracy data obtained from the two experiments supports Dehaene's hypothesis, suggesting that letter shape does not have a visible effect on accuracy rates in a lexical decision task, which measures the end result of a process that goes well beyond the very early stages of visual processing.

5.1.8 Decomposition route is slow

Table 34 shows the global average response times and accuracies of all stimuli in the HIGH and LOW conditions of Experiment 1 (simple nouns) and Experiment 2 (novel complex nouns).

Table 34. Global Average RTs and Accuracies in Exp. 1 and Exp. 2

	Mean RT	Mean accuracy
Simple (Exp. 1)	647	97.2%
Complex (Exp. 2)	1290	97.2%

As can be seen in Table 34, there is a two-fold difference between the mean RTs of simple nouns and novel complex nouns. Mean accuracies, on the other hand, are identical.

In Experiment 1, we can be sure that the subjects used the full-listing route exclusively, because there is nothing to be decomposed there. In Experiment 2, on the other hand, we can be sure that the subjects used some form of the decomposition route, because all stimuli are novel root+suffix combinations never encountered before, and have to be parsed into at least two parts (root + suffixes), even if the entire suffix sequence has a single full-listed entry of its own.

In other words, the two-fold difference between the mean RT in Experiment 1 (647 ms) and the mean RT in Experiment 2 (1,290 ms) gives us a rough idea about the relative speeds of the full-listing and decomposition routes: The suffix sequence at the end of the word seems to be adding an immense workload to the processing of the simple root. Obviously, this is just an informal, preliminary observation, and several additional experiments specifically addressing this issue are needed.

5.2 Main findings

This section reports those findings that are strictly related to the three research questions and hypotheses defined at the beginning.

5.2.1 Frequency effect in bare nouns

Experiment 1 provides clear statistical evidence that more frequent simple words are processed faster than less frequent simple words. Although this is an important and basic finding, there is not much to say about it, since the existence of this effect has already been demonstrated innumerable times in the literature, for many typologically unrelated languages.

The importance of this finding lies more in the fact that it has been obtained in the first large-scale online psycholinguistic experiment on Turkish. The clear statistical results, which have been replicated by all eight versions of Experiment 1, suggest that the methodology developed in this study has produced valid and reliable results.

5.2.2 Left-to-right processing

Results from the PR and RP conditions of Experiment 2 provide some evidence that complex words are parsed from left to right, as suggested by Hankamer (1989).

Non-words that contain a pseudo root and real suffixes such as *gansien+lerinizin* (the PR condition) were rejected significantly faster than non-words that contain a real root and pseudo suffixes such as *bıldırcın+ganıluf* (the RP condition).[42] Critically, this happened despite the fact that the more rapidly processed PR stimuli are more complex than the more slowly processed RP stimuli (*gansien+lerinizin* contains four "morphemes", while *bıldırcın+ganıluf* contains only two).

The finding that complex Turkish words are processed from left to right would have an important implication: If the root is recognized before the suffixes, this means that the orthographic, phonological, syntactic, paradigmatic and semantic representations of the root can be activated before the suffixes have been recognized. If there are mechanisms in the human brain that are fast enough to communicate this multi-level information to the "morphological parser/disambiguator" before it has begun processing the suffixes, this means that the parser has a significant advantage at the outset.

For example, consider the ambiguous string *kalemim* 'my pen / I'm a pen'. If the parser already has the semantic information that the root refers to an inanimate object that can be owned by persons, it might use this information to disambiguate the *-im* suffix more efficiently, and decide that the correct parse is *kalem+P1sg*, rather than *kalem+Pres+A1sg*.

A similar mechanism could be at work at the orthographic/phonological level: If the parser already has the information that the last vowel of the root is an */e/*, it might recognize the suffix *-im* more efficiently, since the combination */e/+/i/* obeys the rules of vowel harmony.

[42] In the four versions that used names of tools as roots, the difference between the RP and PR conditions was not statistically significant, but the RT difference was in the expected direction (PR < RP) also in those versions.

To summarize, the fact that the root has already been recognized as the parser begins to process the suffixes would have important implications at the morphological, semantic, phonological and orthographic levels. Of course, these are informal suggestions without an empirical basis.

5.2.3 Mental representation of suffix sequences

Results from the HIGH and LOW conditions of Experiment 2 provide clear statistical evidence that the processing speed of a suffix sequence is sensitive to the frequency of that suffix sequence, everything else being equal. This evidence supports the hypothesis that the human brain can use suffix sequences like *P1pl+Gen+Rel* (i.e. *-(I)mIzInki*, as in the word-form *çocuğumuzunki* 'the one that belongs to our child') as effective processing units in the recognition of complex words.

As discussed in Section 2.1, existing literature on morphological processing is centered around the "full-listing vs. decomposition" dichotomy. But in all existing models, the thing that is being fully-listed is either an entire root/stem (as in *çocuk* 'child', *çocukluk* 'childhood' or *çocuksuluk* 'childishness'), or an entire complex word-form (as in *çocuğumuzunki* 'the one that belongs to our child'). None of the visual word recognition models in the language processing literature allow for an undecomposed mental entry for a suffix sequence like *P1pl+Gen+Rel*. However, this is exactly what the results of Experiment 2 suggest.

Let us repeat the suggestion in Frauenfelder & Schreuder (1992):

> . . . it is possible that combinations of roots and affixes that a listener encounters frequently could get a separate access representation. Consequently, a single word form might be recognized through the cooperative efforts of the direct [i.e. full-listing] route and the parser [i.e. the decomposition route]. The *frequently co-occurring roots plus affixes* [emphasis added] would be recognized by the direct route, and the *rest of the*

> *word* [emphasis added] by the parser that combines the results of the direct route with the *remaining morphemes* [emphasis added] to be parsed. (Frauenfelder & Schreuder, 1992, p. 180)

Note that Frauenfelder & Schreuder (1992) are not repeating the predictions of a standard "race" version of the dual-route model here. More specifically, they are not simply claiming that the word-form *tamamlananlar* 'the ones that have been completed', for example, is represented both as an unanalyzed chunk (*tamamlananlar*) and also in fully-decomposed form (*tamam+la+n+an+lar*), where the two representations engage in a "race" in which the faster route prevails. Rather, their claim predicts that the initial, four-morpheme part of the word (*tamamlanan* 'the one that has been completed') is listed as an unanalyzed chunk because it is a frequently used complex-form, and is thus accessed through the direct route, while "the rest of the word" (i.e. "the remaining morphemes", which in this case only consists of the plural marker *-lAr*) is analyzed by the parser, and finally, the full-listed complex-form *tamamlanan* and the morpheme *-lar* are combined, to arrive at the complex form *tamamlananlar*. In other words, the root and the first three suffixes are accessed by the direct route in a single step, while the last suffix is accessed in decomposed form.

This hypothesis is supported by three empirical findings of this study: (a) frequent suffix sequences may have their own undecomposed mental representation, (b) 99.9975% of possible suffix sequences are not used even once in a 283-million-word corpus, and (c) those sequences that are used have an extremely uneven distribution, where the most frequent few hundred sequences account for the majority of the word-forms encountered.

Let us demonstrate the resulting picture with a concrete example: The suffix template *+P3sg+Loc+Rel* occurs 982,105 times in the BOUN Corpus (3,470 per million), making it the 23rd most frequent suffix-template of Turkish. Being highly frequent, this suffix-template is a good candidate for having an undecomposed mental representation of its own. Secondly, the bare noun *haşema* refers to a type of swimsuit worn by conservative Muslim women, and is a rarely used root that occurs only 98 times in the BOUN Corpus. Finally, the combination *haşemasındaki* ‘the one in her swimsuit’ is a grammatical word-form, but does not occur in the BOUN Corpus.[43]

When encountered with the zero-frequency complex-form *haşemasındaki*, the full-listing route is of no help, since this combination has never been encountered before and cannot have an undecomposed entry of its own. Thus, the decomposition route analyzes the word into two parts, *haşema*- and -*sındaki* (rather than decomposing it into the four parts *haşema-, -sı, -da, and -ki*, as proposed in existing models). The mental representations of the root *haşema* and the suffix template *+P3sg+Loc+Rel* presumably both contain orthographic, phonological, syntactic, paradigmatic and semantic information. In the final synthesis step, the information coming from the root is combined with the information coming from the suffix template, and the word as a whole is recognized.

Let us go one step further, and apply the suggestion in Frauenfelder & Schreuder (1992) to an even more complex form: The complex-form *haşemasındakilerin* ‘of those in her swimsuit’ contains the suffix-template *+P3sg+Loc+Rel+A3pl+Gen*, which occurs only 2,683 times (9 per million) in the BOUN Corpus. Not being among the most frequent few hundred suffix-templates, *+P3sg+Loc+Rel+A3pl+Gen* probably does not have a separate mental representation of its own. The combination of the last two suffixes, on the other hand

[43] As of the date of writing, *haşemasındaki* also did not produce any results on the Google search engine, which is presumably larger than the BOUN Corpus by several orders of magnitude.

(*A3pl+Gen*) occurs 2,342,113 times in the BOUN Corpus (8,276 per million), making it the 13th most frequent suffix-template of Turkish. Being highly frequent, *A3pl+Gen* is another good candidate for having an undecomposed mental representation of its own.

As before, when encountered with the zero-frequency complex-form *haşemasındakilerin*, the full-listing route is of no help. Thus, the decomposition route analyzes the word into three parts, *haşema-*, *-sındaki, and -lerin*. The information coming from the root is combined with the information coming from the two suffix templates *+P3sg+Loc+Rel* and *+A3pl+Gen*, and the word as a whole is recognized.[44]

[44] Moreover, if the left-to-right parsing hypothesis is correct, and if there exist sufficiently fast communication mechanisms between the brain regions that process the root and the suffixes, the parser already "knows" that the root refers to a type swimsuit and that its last vowel is /a/, thus starting the suffix-recognition process, and the final semantic synthesis with a significant advantage.

CHAPTER 6

CONCLUSION AND FUTURE WORK

The quantitative work on corpus data and the results of the two experiments seem to have provided an interesting picture, which in some cases supports well-established findings in the literature, and in some other cases represents a departure from existing models of language processing.

First of all, the importance of frequency in the mental representation and processing of language has been reaffirmed. The results also support the well-known phenomena that frequency distributions are extremely uneven, that different frequency measures are highly correlated, and that linguistic data is characterized by sparsity.

The suggestion that suffix sequences might have their own mental representation, on the other hand, is a slight but important departure from existing models of visual word recognition. In fact, this is a refinement of existing models based on decomposition, in that it allows for the representation of suffix sequences, in addition to the representation of individual suffixes. In other words, the only change is that the "unit of decomposition" is a chunk of several suffixes instead of a single suffix, while the analysis and synthesis mechanisms remain the same.

Finally, Sections 6.1 to 6.4 present some ideas for future work in the field.

6.1 Continue quantification

The quantification and modelling effort in Chapter 3 seems to have paid off. General-purpose linguistic datasets such as the frequency distributions of letters, suffixes, suffix templates and their ngrams have been produced and reported. These datasets seem to have significantly contributed to the clear statistical results obtained from the experiments.

As mentioned in Section 2.2.1, the properties of words can be quantified at a surprisingly large number of levels. Several metrics are waiting to be defined and measured, including the following: letter features, uniqueness point, number of phonemes, number of syllables, polarity, emotional content, imageability, family size, family frequency, age of acquisition, number of synonyms, number of senses, collocates, visual recognition times, visual recognition accuracies.

6.2 Publish datasets

In the spirit of the "copy-left" movement, all linguistic datasets as well as the raw data of the experiments should be made publicly (and electronically) available without any copyright restrictions. This will prevent researchers from reinventing the wheel every time they form a testable hypothesis or design an experiment, and will allow more in-depth analysis of existing experimental results and facilitate the design of replication studies.

6.3 Analyze existing data

Large amounts of data, including the demographic data of subjects, have been collected in the two experiments. Although the statistical analysis of the demographic data has not been included here, several interesting questions can be asked:

How do age, gender, handedness and education level affect visual word recognition? The data on handedness is especially interesting because the experiments had a factorial design where the choice of keys intentionally forced half of the subjects to press the ‘yes’ button with their left hand, and the other half with their right hand. Moreover, handedness data is available for each subject. In other words, we know if a subject used his/her dominant hand or non-dominant hand in any given response. Anecdotal evidence from the pilot studies indicates that the subjects prefer to press the ‘yes’ button with their dominant hand. How does this parameter affect response times and recognition accuracies?

Another analysis that is possible with existing data concerns the training and fatigue effects described in Section 4.2.6. How does the order of presentation of the stimuli affect the subjects’ performance? When does the training effect come to an end? When does the fatigue effect kick in? These analyses would have important implications for future experiment design.

Finally, existing experimental data allows us to examine the relationships between the following pairs of variables: word length vs. RT and error rate, number of syllables vs. RT and error rate, bigram frequency vs. RT and error rate, neighborhood size and neighborhood frequency vs. RT and error rate (see McGinnies, Comer, & Lacey, 1952, p .67; Schiepers, 1980, p. 75), the correlation between RT and error rate (see Taft, 1979).

6.4 Additional experiments

The JavaScript-based online platform can be used to quickly implement any type of lexical decision experiment at almost zero cost. The following factorial design, for example, would answer several questions in a single experiment:

Design a 4 x 4 factorial lexical decision experiment for complex-forms where the root is either a high-frequency real word, a low-frequency real word, a phonotactically plausible non-word, or a phonotactically implausible non-word, while the suffix-template is either a high-frequency template, a low-frequency template, a zero-frequency template (i.e. grammatical but unattested), or an ungrammatical template (e.g. an impossible suffix sequence).

APPENDIX A

SUFFIXES OF TURKISH

#	Tag	Description	Suffix Type*	Vowel Type**	Surface Forms
00	A1pl	1st person plural	Infl.	I	*-(y)Iz*
01	A1sg	1st person singular	Infl.	I	*-(y)Im*
02	A2pl	2nd person plural	Infl.	I	*-(y)InIz*
03	A2sg	2nd person singular	Infl.	I	*-(y)In*
04	A3pl	3rd person plural	Infl.	A	*-lAr*
05	A3sg	3rd person singular	Infl.	I	*Ø*
06	Abbr	Abbreviation	-	-	*Ø*
07	Able	Abilitative	Infl.	A	*-Abil*
08	Abl	Ablative	Infl.	A	*-Dan*
09	Acc	Accusative	Infl.	I	*-(y)I*
10	Acquire		Drv.	A	*-lAn*
11	Acro	Acronym	-	-	*Ø*
12	Adj	Adjective	POS	-	*Ø*
13	Adv	Adverb	POS	-	*Ø*
14	AfterDoingSo	-(y)Ip Gerund	Infl.	I	*-(y)Ip*
15	Agt	Agentive	Drv.	I	*-CI*
16	Aor	Aorist	Infl.	Irr.	*-Ar, -Ir*
17	Apos	Apostrophe	-	-	*Ø*
18	AsIf		Drv.	I	*-CasınA*
19	AsLongAs		Infl.	I	*-DıkçA*
20	Become		Drv.	A	*-lAş*
21	ByDoing So	-(y)ArAk Gerund	Infl.	A	*-(y)ArAk*
22	Caus	Causative	Drv.	I	*-Dır*
23	Cond	Conditional	Infl.	A	*-sA*
24	Cop	Copula	Infl.	I	*-DI*
25	Dat	Dative	Infl.	A	*-(y)A*
26	Desr		Infl.	A	*-sA,*
27	Dim	Diminutive	Drv.	I	*-Cık*
28	Equ	Equative	Infl.	A	*-CA*
29	EverSince		Infl.	A	*-(y)Agel*
30	FeelLike		Infl.	A	*-(y)AsI*
31	FitFor		Drv.	I	*-lIk*

#	Tag	Description	Suffix Type*	Vowel Type**	Surface Forms
32	Fut	Future Tense	Infl.	A	*-(y)AcAk*
33	FutPart		Infl.	A	*-(y)AcAğI*
34	Gen	Genitive	Infl.	I	*-(n)In*
35	Hastily		Infl.	I	*-(y)Iver*
37	Imp	Imperative	Infl.	-	*Ø*
38	Inf1	Nominalizer 1	Infl.	A	*-mAk*
39	Inf2	Nominalizer 2	Infl.	A	*-mA*
36	Inf3	Nominalizer 3	Infl.	I	*-(y)Iş*
40	Ins		Infl.	A	*-(y)lA*
41	Loc	Locative	Infl.	A	*-DA*
42	Narr	Narrative	Infl.	I	*-mIş*
43	NarrNess		Drv.	I	*-mIşlIk*
44	Neces		Infl.	A	*-mAlI*
45	Neg	Negative	Infl.	A	*-mA*
46	Ness		Drv.	I	*-lIk*
47	NoHats		-	-	*Ø*
48	Nom	Nominative	Infl.	-	*Ø*
49	NotState		Drv.	A	*-mAzlIk*
50	Noun	Noun	POS	-	*Ø*
51	Opt	Optative	Infl.	A	*-AlIm*
52	P1pl	1st person pl. possessive	Infl.	I	*-(I)mIz*
53	P1sg	1st person sing. possessive	Infl.	I	*-(I)m*
54	P2pl	2nd person pl. possessive	Infl.	I	*-(I)mIz*
55	P2sg	2nd person sing. possessive	Infl.	I	*-(I)n*
56	P3pl	3rd person pl. possessive	Infl.	A	*-lArI*
57	P3sg	3rd person sing. possessive	Infl.	I	*-(s)I*
59	Pass	Passive	Drv.	I	*-Il*
58	Past	Past Tense	Infl.	I	*-DI*
60	PastPart	Nominalizer	Infl.	I	*-DIğI*
61	Pnon	No Possessive Marker	Infl.	-	*Ø*
62	Pos	Positive	Infl.	-	*Ø*
63	Pres	Relative-Clause Marker	Infl.	A	*-(y)An*
64	PresPart		Infl.	A	*-(y)An*
65	Prog1	Progressive	Infl.	I	*-Iyor*
66	Prog2	Progressive	Infl.	A	*-mAktA*
67	Pron	Pronoun	POS	-	*Ø*

#	Tag	Description	Suffix Type*	Vowel Type**	Surface Forms
68	Prop	Proper Name	POS	-	*Ø*
69	Rel	Relativizer	Infl.	Inv.	*-ki,*
70	Repeat		Infl.	A	*-Adur*
71	Since		Infl.	A	*-(y)AlI*
72	SinceDoingSo		Infl.	A	*-(y)AlI*
73	Verb	Verb	POS	-	*Ø*
74	When		Infl.	I	*-(y)IncA*
75	While		Infl.	Inv.	*-(y)ken,*
76	With	Instrumental/Comitative	Drv.	I	*-lI*
77	Without	Abessive/Privative	Drv.	I	*-sIz*
78	WBATHDS[45]		Infl.	A	*-(y)AmAdAn*
79	WHDS[46]		Infl.	A	*-mAksIzIn*

* Infl. = inflectional, Drv. = derivational, POS = main part of speech

** I = I-type vowel, A = A-type vowel, Inv. = invariant, Irr. = irregular

[45] WBATHDS: WithoutBeingAbleToHaveDoneSo
[46] WHDS: WithoutHavingDoneSo

APPENDIX B

MOST FREQUENT LETTERS OF TURKISH ACCORDING TO BOUN CORPUS

WORD-INITIAL		WORD-FINAL		ANYWHERE	
Letter	Frequency	Letter	Frequency	Letter	Frequency
b	44,090,273	n	51,290,297	a	260,112,729
k	30,734,003	e	44,283,345	e	187,851,722
d	29,722,725	a	40,295,016	i	181,751,723
a	28,505,182	i	36,811,773	n	156,278,766
y	25,878,252	r	34,392,772	r	146,219,968
i	23,071,464	ı	28,379,563	l	146,131,138
o	17,752,022	k	22,863,560	ı	105,261,356
g	16,740,931	u	14,362,462	k	99,454,330
v	15,014,283	m	10,511,513	d	93,472,768
h	13,488,226	l	8,132,453	y	71,975,642
s	13,341,906	z	8,003,251	m	67,682,549
e	7,354,131	t	6,503,245	u	67,104,224
ö	7,194,989	ş	4,487,439	t	62,472,970
p	6,612,897	ü	3,385,040	s	58,926,177
ç	6,205,912	p	2,541,234	b	56,353,705
f	5,660,025	ç	1,968,650	o	53,107,713
ş	4,998,813	y	1,826,598	ü	37,400,349
ü	4,699,078	s	1,819,216	ş	35,287,648
m	4,006,470	o	1,478,418	z	31,211,288
u	3,874,624	f	952,823	g	25,384,097
r	3,649,792	d	604,248	v	23,825,184
t	3,264,524	h	594,838	ğ	22,579,397
c	3,151,614	b	353,756	h	21,452,381
z	2,780,477	v	342,677	c	20,918,799
l	2,029,066	g	251,404	ç	20,702,437
n	1,867,444	ğ	144,846	p	18,681,422

ı	806,469	j	111,616	ö	17,830,438
j	335,857	c	91,021	f	11,173,306
ğ	0	w	43,562	j	1,491,305
q	0	ö	3,670	w	86,649
w	0	q	1,143	q	9,619
x	0	x	0	x	0
Σ	326,831,449	Σ	326,831,449	Σ	2,102,191,799

APPENDIX C

MOST FREQUENT LETTER BIGRAMS OF TURKISH

ACCORDING TO BOUN CORPUS

Note that ‘#’ marks the beginning of a word, and ‘|’ marks the end. For example, the bigram *#b* refers to the letter *b* occurring at the beginning of a word, while the bigram *n*| refers to the letter *n* occurring at the end of a word.

MIDDLE		INITIAL		FINAL	
bigram	frequency	bigram	frequency	bigram	frequency
ar	40,968,010	#b	44,090,273	n|	51,286,906
la	39,915,859	#k	30,734,003	e|	44,283,345
an	37,557,802	#d	29,722,725	a|	40,294,977
er	33,383,695	#a	28,505,182	i|	36,810,927
in	31,224,527	#y	25,878,252	r|	34,223,197
le	31,191,572	#i	23,071,464	ı|	28,379,563
de	27,034,519	#o	17,752,022	k|	22,810,860
ın	25,790,373	#g	16,740,931	u|	14,362,440
da	24,380,520	#v	15,014,283	m|	10,509,129
en	24,095,506	#h	13,488,226	l|	8,130,942
ya	23,109,256	#s	13,341,906	z|	7,986,238
ir	21,892,623	#e	7,354,131	t|	6,495,846
il	21,135,179	#ö	7,194,989	ş|	4,460,013
ka	21,132,649	#p	6,612,897	ü|	3,385,040
ma	19,913,850	#ç	6,205,912	p|	2,540,547
nd	19,077,058	#f	5,660,025	ç|	1,942,684
bi	18,927,254	#ş	4,998,813	y|	1,819,238
ra	18,709,191	#ü	4,699,078	s|	1,789,257
al	18,203,574	#m	4,006,470	o|	1,478,418
ak	17,828,963	#u	3,874,624	f|	839,960

APPENDIX D

MOST FREQUENT LETTER TRIGRAMS OF TURKISH ACCORDING TO BOUN CORPUS

Note that ‘#’ marks the beginning of a word, and ‘|’ marks the end. For example, the trigram *#bi* refers to the letter bigram *bi* occurring at the beginning of a word, while the trigram *an|* refers to the letter bigram *an* occurring at the end of a word.

MIDDLE		INITIAL		FINAL	
trigram	frequency	trigram	frequency	trigram	frequency
lar	17,840,775	#bi	14,923,284	an|	14,090,177
ler	13,563,353	#ya	12,925,211	da|	12,573,360
eri	12,504,179	#ka	12,816,174	in|	11,564,726
arı	11,298,223	#ve	12,020,801	ir|	11,556,602
bir	10,971,873	#de	10,416,458	en|	10,786,749
ara	9,241,543	#ol	10,320,346	de|	10,705,252
nda	8,758,887	#bu	9,265,853	ve|	9,040,772
ile	7,535,094	#ba	9,259,101	ın|	8,551,320
ası	7,134,179	#ha	7,245,574	ar|	7,928,454
lan	7,032,130	#ge	6,576,875	ak|	7,214,981
ini	6,819,644	#da	6,303,411	er|	6,793,134
ını	6,752,163	#be	5,427,026	le|	6,209,786
rin	6,341,343	#il	5,086,397	nı|	5,697,230
yor	6,244,097	#ko	5,034,178	ni|	5,161,005
ınd	6,222,435	#al	4,751,984	bu|	5,129,891
nde	5,742,409	#ar	4,247,296	ri|	4,538,774
anı	5,733,537	#ye	3,953,526	ne|	4,474,776
ama	5,686,917	#iç	3,934,643	na|	4,399,122
rın	5,521,342	#is	3,868,457	ki|	4,387,558
ola	5,513,024	#ku	3,760,266	ek|	4,267,612

APPENDIX E

MOST FREQUENT SUFFIXES OF TURKISH ACCORDING TO BOUN CORPUS

TEMPLATE-INITIAL		TEMPLATE-FINAL		ANYWHERE	
Morpheme	Frequency	Morpheme	Frequency	Morpheme	Frequency
P3sg	37,531,526	P3sg	29,565,419	P3sg	61,851,909
A3pl	22,780,373	Loc	17,614,011	A3pl	27,504,964
Pass	10,402,113	Dat	16,554,996	Loc	20,157,607
Loc	8,522,658	Gen	16,499,069	Dat	16,571,996
Gen	7,358,236	Acc	11,896,321	Gen	16,540,615
Inf2	7,280,436	Postp	10,729,668	Pass	12,129,429
Dat	6,908,506	PresPart	8,846,461	Acc	11,896,571
PastPart	6,746,480	Past	8,727,693	Inf2	11,044,901
Past	6,423,379	A3pl	7,497,970	Past	10,411,050
PresPart	6,113,048	Abl	6,428,026	PresPart	9,613,180
Caus	4,766,224	Ins	4,423,279	PastPart	8,891,756
With	4,590,190	With	4,363,149	Abl	6,456,381
Prog1	4,142,779	Prog1	3,577,605	Prog1	5,970,774
Neg	3,139,148	Cop	3,290,851	Caus	5,723,564
Aor	2,750,310	A1sg	3,186,572	Aor	5,184,027
BDS	2,490,351	BDS	2,988,599	Neg	4,904,699
Abl	2,467,746	Aor	2,651,071	With	4,873,814
Ness	2,230,889	Inf2	2,633,883	Ins	4,426,759
Inf1	2,218,927	Rel	2,527,737	A1sg	4,001,305
Able	2,084,949	Inf1	2,469,923	Cop	3,333,858
Fut	1,877,973	Fut	2,274,837	Fut	3,249,452
Narr	1,825,715	A1pl	1,944,972	Narr	3,232,754
Acc	1,702,255	Narr	1,942,025	BDS	2,988,842
Ins	1,677,114	Cond	1,562,579	Able	2,858,709
Pres	1,607,567	A2pl	1,318,598	Inf1	2,827,749
A1sg	1,370,696	P1pl	1,012,695	Rel	2,602,953

Agt	1,350,042	Ness	1,006,754	A1pl	2,591,640
P1pl	1,093,301	ADS	973,602	Pres	2,541,703
Inf3	1,073,004	P1sg	969,591	Ness	2,401,229
P1sg	1,064,183	A2sg	839,391	P1pl	2,168,476
Cond	1,051,620	While	834,525	Cond	1,866,292
A1pl	981,085	P3pl	652,394	P1sg	1,811,347
Imp	979,586	Agt	635,911	FutPart	1,613,010
FutPart	835,203	Without	499,614	A2pl	1,564,829
Become	818,196	PastPart	488,299	Agt	1,403,042
ADS	781,396	Inf3	460,785	Imp	1,311,749
Acquire	772,957	P2pl	454,622	Inf3	1,156,009
Without	566,544	Imp	407,537	P2pl	1,045,165
P2pl	469,500	Desr	320,397	P3pl	978,504
Prog2	417,019	Equ	311,570	A2sg	974,714
Desr	390,183	When	276,326	ADS	973,615
Rel	343,244	Pres	268,923	While	834,532
Opt	313,180	Prog2	214,074	Become	834,523
A2pl	310,003	P2sg	126,734	Acquire	773,523
P2sg	262,838	AsIf	111,338	Prog2	681,233
A2sg	223,245	AsLongAs	71,268	Without	570,933
When	223,117	Opt	60,657	Desr	547,534
Neces	132,626	WHDS	31,312	Opt	378,377
AsIf	91,414	Dim	9,066	Equ	311,570
P3pl	52,635	SDS	7,083	P2sg	299,716
AsLongAs	47,675	WBATHDS	6,558	When	297,415
While	44,677	FeelLike	3,706	Neces	237,551
Equ	42,624	NarrNess	3,249	AsIf	111,338
WHDS	21,977	NotState	1,581	AsLongAs	71,269
Dim	21,070	Since	1,254	WHDS	31,312
Hastily	17,017	FitFor	963	Dim	21,207
Recip	7,910			Hastily	20,234
NotState	7,069			Narrness	10,495
WBATHDS	5,545			NotState	9,014

SDS	5,434			Recip	7,910
NarrNess	5,138			SDS	7,096
FeelLike	4,772			FeelLike	7,001
EverSince	4,282			WBATHDS	6,558
Stay	2,405			EverSince	6,323
FitFor	1,666			Stay	2,407
Since	1,254			FitFor	1,666
Repeat	830			Since	1,254
Start	286			Repeat	966
Almost	85			Start	286
				Almost	92
Σ	176,419,110	Σ	186,577,093	Σ	450,671,965

APPENDIX F

MOST FREQUENT SUFFIX BIGRAMS OF TURKISH ACCORDING TO BOUN CORPUS

Note that ‘#’ marks the beginning of a suffix template, and ‘|’ marks the end. For example, *#+P3sg* refers to the suffix *P3sg* occurring immediately after the root, while *Loc+|* refers to the suffix *Loc* occurring at the very end.

MIDDLE		INITIAL		FINAL	
bigram	frequency	bigram	frequency	bigram	frequency
A3pl+P3sg	11,125,524	#+P3sg	37,531,526	P3sg+|	29,565,419
P3sg+Loc	9,181,147	#+A3pl	22,780,373	Loc+|	17,614,011
P3sg+Acc	8,319,809	#+Pass	10,402,113	Dat+|	16,554,996
PastPart+P3sg	6,272,174	#+Loc	8,522,658	Gen+|	16,499,069
P3sg+Dat	5,823,687	#+Gen	7,358,236	Acc+|	11,896,321
P3sg+Gen	4,516,591	#+Inf2	7,280,436	PresPart+|	8,846,461
Inf2+P3sg	4,336,171	#+Dat	6,908,506	Past+|	8,727,693
A3pl+Gen	2,975,743	#+PastPart	6,746,480	A3pl+|	7,497,970
Pass+PresPart	2,423,869	#+Past	6,423,379	Abl+|	6,428,026
P3sg+Abl	2,306,425	#+PresPart	6,113,048	Ins+|	4,423,279
Loc+Rel	2,241,198	#+Caus	4,766,224	With+|	4,363,149
Pass+Inf2	2,162,245	#+With	4,590,190	Prog1+|	3,577,605
Pres+Cop	1,832,847	#+Prog1	4,142,779	Cop+|	3,290,851
P3sg+Ins	1,660,370	#+Neg	3,139,148	A1sg+|	3,186,572
Caus+Pass	1,606,985	#+Aor	2,750,310	ByDoingSo+|	2,988,599
A3pl+Dat	1,541,957	#+ByDoingSo	2,490,351	Aor+|	2,651,071
A3pl+Loc	1,366,901	#+Abl	2,467,746	Inf2+|	2,633,883
Pass+Past	1,288,424	#+Ness	2,230,889	Rel+|	2,527,737
FutPart+P3sg	1,265,524	#+Inf1	2,218,927	Inf1+|	2,469,923
Inf2+A3pl	1,086,718	#+Able	2,084,949	Fut+|	2,274,837

APPENDIX G

MOST FREQUENT SUFFIX TRIGRAMS OF TURKISH ACCORDING TO BOUN CORPUS

Note that ‘#’ marks the beginning of a suffix template, and ‘|’ marks the end. For example, *#+A3pl+P3sg* refers to the bigram *A3pl+P3sg* occurring immediately after the root, while *P3sg+Acc+*|refers to the bigram *P3sg+Acc* occurring at the very end.

MIDDLE		INITIAL		FINAL	
trigram	frequency	trigram	frequency	trigram	frequency
PastPart+P3sg+Acc	2,165,802	#+A3pl+P3sg	9,544,305	P3sg+Acc+|	8,319,604
A3pl+P3sg+Acc	1,805,035	#+P3sg+Loc	7,799,760	P3sg+Loc+|	7,850,643
Pass+Inf2+P3sg	1,727,563	#+PastPart+P3sg	4,583,666	P3sg+Dat+|	5,819,860
A3pl+P3sg+Gen	1,315,406	#+P3sg+Dat	3,912,347	A3pl+P3sg+|	5,153,399
P3sg+Loc+Rel	1,137,732	#+P3sg+Acc	3,079,146	P3sg+Gen+|	4,513,539
A3pl+P3sg+Dat	1,010,340	#+P3sg+Gen	2,704,439	PastPart+P3sg+|	3,499,092
A3pl+P3sg+Loc	913,214	#+A3pl+Gen	2,578,768	A3pl+Gen+|	2,972,157
Pass+PastPart+P3sg	789,047	#+Pass+PresPart	2,130,266	Inf2+P3sg+|	2,556,278
Neg+PastPart+P3sg	605,338	#+Inf2+P3sg	2,092,255	Pass+PresPart+|	2,320,360
Inf2+A3pl+P3sg	603,229	#+Pass+Inf2	1,729,727	P3sg+Abl+|	2,298,768
FutPart+P3sg+Acc	561,148	#+P3sg+Abl	1,505,106	Loc+Rel+|	2,196,511
Inf2+P3sg+Acc	539,945	#+A3pl+Dat	1,332,418	Pres+Cop+|	1,812,872
A3pl+P3sg+Abl	538,501	#+A3pl+Loc	1,284,056	P3sg+Ins+|	1,659,845
Inf2+P3sg+Dat	478,531	#+Caus+Pass	1,270,248	A3pl+Dat+|	1,541,736
PastPart+A3pl+P3sg	453,444	#+P3sg+Ins	1,155,107	A3pl+Loc+|	1,247,082
Caus+Pass+Inf2	425,156	#+Pres+Cop	1,115,565	Pass+Past+|	1,242,546
A3pl+P3sg+Ins	332,299	#+Pass+Past	1,032,413	Inf2+Dat+|	992,226
Inf2+P3sg+Gen	284,802	#+Inf2+A3pl	901,621	Past+A1sg+|	981,006
Able+Neg+Aor	277,174	#+Loc+Rel	822,412	Prog1+A1sg+|	838,053
P3sg+Pres+Cop	266,564	#+Pass+PastPart	784,309	Aor+While+|	733,472

APPENDIX H

MOST FREQUENT 200 NOMINAL SUFFIX TEMPLATES OF TURKISH ACCORDING TO BOUN CORPUS

Rank	Suffix template	Frequency		Rank	Suffix template	Frequency
1.	Noun	67,395,919		61.	Noun+Acquire+Caus+Inf2	62,999
2.	Noun+P3sg	15,846,993		62.	Noun+Agt+P3sg	59,754
3.	Noun+Loc	6,719,144		63.	Noun+A3pl+P1pl+Gen	56,522
4.	Noun+Gen	6,682,436		64.	Noun+A3pl+P1sg	54,296
5.	Noun+P3sg+Loc	6,438,347		65.	Noun+A3pl+P1pl+Acc	53,859
6.	Noun+Dat	6,340,183		66.	Noun+Agt+Gen	51,371
7.	Noun+A3pl	5,098,469		67.	Noun+Agt+A3pl+P3sg	49,857
8.	Noun+A3pl+P3sg	4,491,411		68.	Noun+P1sg+Ins	48,061
9.	Noun+With	4,023,238		69.	Noun+P3sg+Equ	46,235
10.	Noun+P3sg+Dat	3,645,429		70.	Noun+Acquire+Inf2	45,282
11.	Noun+P3sg+Acc	2,774,868		71.	Noun+P2pl+Loc	44,990
12.	Noun+P3sg+Gen	2,553,663		72.	Noun+With+P3sg	44,240
13.	Noun+A3pl+Gen	2,342,113		73.	Noun+P2pl+Gen	43,554
14.	Noun+Abl	2,132,159		74.	Noun+P3sg+Loc+Pres+Cop	43,306
15.	Noun+Ins	1,605,902		75.	Noun+Equ	42,526
16.	Noun+P3sg+Abl	1,335,324		76.	Noun+With+A3pl	40,998
17.	Noun+A3pl+Dat	1,216,820		77.	Noun+A3pl+P2pl+Acc	39,804
18.	Noun+Acc	1,168,631		78.	Noun+P1sg+Abl	39,737
19.	Noun+A3pl+P3sg+Acc	1,168,590		79.	Noun+P3sg+Past	39,607
20.	Noun+A3pl+P3sg+Gen	1,163,978		80.	Noun+Pres+A1pl	37,607
21.	Noun+A3pl+Loc	1,147,513		81.	Noun+With+Ness	37,263
22.	Noun+P3sg+Ins	1,109,168		82.	Noun+Acquire+Caus+Pass+Inf2+P3sg	33,980
23.	Noun+P3sg+Loc+Rel	982,105		83.	Noun+A3pl+P1pl+Dat	33,974
24.	Noun+A3pl+P3sg+Dat	859,904		84.	Noun+A3pl+P2pl+Dat	33,264
25.	Noun+Loc+Rel	793,974		85.	Noun+A3pl+P2pl	32,857
26.	Noun+A3pl+P3sg+Loc	731,473		86.	Noun+A3pl+P1sg+Acc	32,133
27.	Noun+Pres+Cop	595,410		87.	Noun+Ness+P3sg+Acc	32,016
28.	Noun+A3pl+Abl	536,835		88.	Noun+Ness+P3sg+Dat	31,476
29.	Noun+A3pl+Ins	526,525		89.	Noun+Agt+A3pl+Dat	30,608
30.	Noun+P1sg	518,644		90.	Noun+P1pl+Abl	30,273

31.	Noun+Without	491,522		91.	Noun+A3pl+Equ	29,247
32.	Noun+A3pl+P3sg+Abl	464,901		92.	Noun+P3sg+Loc+Past	25,126
33.	Noun+Ness	463,592		93.	Noun+Become+Inf2	24,370
34.	Noun+Agt	401,099		94.	Noun+Cond	24,359
35.	Noun+P1pl	363,791		95.	Noun+Ness+Dat	24,040
36.	Noun+Rel	326,758		96.	Noun+Loc+Pres+Cop	24,020
37.	Noun+A3pl+P3sg+Ins	298,408		97.	Noun+Become+Caus+Pass+PresPart	23,799
38.	Noun+P3sg+Pres+Cop	153,970		98.	Noun+Ness+Gen	23,728
39.	Noun+P1pl+Loc	149,957		99.	Noun+P3sg+Loc+Pres+A1pl	23,592
40.	Noun+P1pl+Dat	143,185		100.	Noun+With+Past	23,312
41.	Noun+P2pl	140,720		101.	Noun+Become+Past	21,938
42.	Noun+P1pl+Gen	138,904		102.	Noun+P3sg+Loc+Pres+A1sg	21,566
43.	Noun+P1sg+Acc	124,485		103.	Noun+P2sg+Dat	21,278
44.	Noun+Agt+A3pl	121,465		104.	Noun+P2sg+Loc	21,165
45.	Noun+P1sg+Dat	110,737		105.	Noun+Acquire+Caus+Inf1	20,534
46.	Noun+A3pl+Loc+Rel	109,199		106.	Noun+Agt+Ness	20,382
47.	Noun+A3pl+P1pl	108,625		107.	Noun+A3pl+P3sg+Equ	20,190
48.	Noun+Ness+P3sg	108,580		108.	Noun+Acquire+Caus+Past	20,036
49.	Noun+P1pl+Acc	108,315		109.	Noun+With+A3pl+Gen	19,210
50.	Noun+P1pl+Loc+Rel	105,413		110.	Noun+Acquire+Caus+PresPart	18,763
51.	Noun+A3pl+P3sg+Loc+Rel	104,048		111.	Noun+Acquire+Caus+Pass+PresPart	18,636
52.	Noun+Past	100,203		112.	Noun+A3pl+P3sg+Pres+Cop	18,437
53.	Noun+P2pl+Acc	98,630		113.	Noun+With+A3pl+P3sg	18,157
54.	Noun+P2sg	97,754		114.	Noun+P3pl+Gen	17,963
55.	Noun+P1sg+Loc	94,006		115.	Noun+Acquire+Inf2+P3sg	17,261
56.	Noun+A3pl+Pres+Cop	90,967		116.	Noun+A3pl+P1pl+Loc	17,213
57.	Noun+P2pl+Dat	85,521		117.	Noun+Acquire+PresPart	17,179
58.	Noun+With+Pres+Cop	81,112		118.	Noun+Become+PresPart	17,147
59.	Noun+Agt+A3pl+Gen	75,817		119.	Noun+A3pl+P1sg+Ins	17,075
60.	Noun+P1sg+Gen	71,508		120.	Noun+P2sg+Acc	17,030

Rank	Suffix template	Frequency		Rank	Suffix template	Frequency
121.	Noun+A3pl+P1sg+Dat	16,520		161.	Noun+With+Pres+A1pl	10,493
122.	Noun+P2sg+Equ	16,312		162.	Noun+Become+Caus+PastPart+P3sg	10,400
123.	Noun+With+Gen	16,283		163.	Noun+Acquire+Prog1	10,286
124.	Noun+With+Ness+P3sg	15,588		164.	Noun+Pres+A2sg	10,070
125.	Noun+P2pl+Abl	15,526		165.	Noun+Acquire+Narr	10,030
126.	Noun+P2pl+Ins	15,446		166.	Noun+Acquire+Caus+Inf2+Dat	9,955
127.	Noun+P1pl+Ins	15,258		167.	Noun+Ness+Abl	9,843
128.	Noun+Ness+P3sg+Gen	15,161		168.	Noun+Loc+Pres+A1sg	9,843
129.	Noun+P2sg+Ins	15,068		169.	Noun+Acquire+Caus+Pass+Fut	9,748
130.	Noun+Ness+P3sg+Loc	15,030		170.	Noun+Without+Pres+Cop	9,732
131.	Noun+While	14,913		171.	Noun+Become+Narr	9,590
132.	Noun+Agt+Dat	14,775		172.	Noun+Without+A3pl	9,357
133.	Noun+Loc+Pres+A1pl	14,590		173.	Noun+P2sg+Gen	9,301
134.	Noun+Loc+Past	14,587		174.	Noun+Agt+A3pl+P3sg+Abl	9,275
135.	Noun+P3pl+Loc	14,389		175.	Noun+Acquire+Inf1	9,225
136.	Noun+Agt+Acc	14,386		176.	Noun+Acquire+ByDoingSo	9,220
137.	Noun+Become+Inf2+P3sg	14,278		177.	Noun+A3pl+P1sg+Loc	9,101
138.	Noun+P1sg+Loc+Rel	14,054		178.	Noun+Pres+A1sg	8,993
139.	Noun+Agt+A3pl+Abl	14,020		179.	Noun+Acquire+Inf2+Dat	8,964
140.	Noun+Become+Caus+Past	13,772		180.	Noun+Ness+A3pl	8,918
141.	Noun+A3pl+P1sg+Gen	13,734		181.	Noun+With+Dat	8,808
142.	Noun+With+Pres+A1sg	13,655		182.	Noun+Acquire+Caus+Pass+Aor	8,680
143.	Noun+Acquire+Past	13,311		183.	Noun+Become+Fut	8,643
144.	Noun+P3sg+Loc+While	13,213		184.	Noun+Acquire+Caus+Inf2+Loc	8,615
145.	Noun+P3sg+Pres+A1sg	13,050		185.	Noun+Ness+Loc	8,613
146.	Noun+A3pl+P2pl+Gen	12,659		186.	Noun+P3pl+Dat	8,478
147.	Noun+A3pl+P1pl+Abl	12,504		187.	Noun+Dim	8,324
148.	Noun+Acquire+Caus+Pass+Narr	12,446		188.	Noun+Gen+Pres+Cop	8,168
149.	Noun+Ness+A3pl+P3sg	12,268		189.	Noun+Acquire+Caus+Pass+Past	8,155
150.	Noun+Agt+A3pl+Ins	12,197		190.	Noun+With+A3pl+Dat	7,789
151.	Noun+Agt+A3pl+P3sg+Gen	12,175		191.	Noun+Without+Ness	7,749
152.	Noun+A3pl+Past	12,088		192.	Noun+Become+Caus+PresPart	7,726
153.	Noun+Acquire+Caus+Inf2+P3sg	11,917		193.	Noun+Become+Caus+Inf1	7,596
154.	Noun+Become+Caus+Pass+Fut	11,828		194.	Noun+Agt+A3pl+P3sg+Dat	7,543
155.	Noun+Narr	11,824		195.	Noun+A3pl+P1sg+Abl	7,487
156.	Noun+A3pl+P1pl+Ins	11,733		196.	Noun+Pres+A2pl	7,467

Rank	Suffix template	Frequency		Rank	Suffix template	Frequency
157.	Noun+Become+Caus+Pass+Past	11,334		197.	Noun+Become+Caus+Pass+Inf2+P3sg	7,415
158.	Noun+Loc+While	10,931		198.	Noun+With+Ness+P3sg+Acc	7,327
159.	Noun+Agt+P1sg	10,695		199.	Noun+With+Pres+A2sg	7,301
160.	Noun+P2pl+Loc+Rel	10,681		200.	Noun+Loc+Rel+A3pl	7,232

APPENDIX I

NON-WORD STIMULI USED IN EXPERIMENT 1

STIMULUS SET 1			STIMULUS SET 2		
Wuggy	Wuggy+suffix	Random	Wuggy	Wuggy+suffix	Random
mene	kusesiz	çöiyş	fekotan	kürütlük	hdggıüok
tezt	gırganlı	möpty	fordat	estlitalık	bbmp
netaet	pasagınçlı	jıvr	febne	kürazarlık	gtyjorg
konze	mümabazıcı	nrk	kiliz	keneylli	şchglsho
filovüs	gerükçü	kçhcpku	danak	meznuplu	lüiçaa
fentre	civetmek	mügvg	maşen	layılı	apgm
sıragra	kazonsuz	jhgv	cara	ütuybacı	gypbdj
cata	şimlik	kjounk	aşbe	ahrörcü	öpfcimf
telyih	goraflık	saio	tese	çödancı	tsaaivb
itrah	yıltozmak	prfd	halla	famocasız	üemsc
koltekt	gosınlık	afeucsof	teşkuraf	ifriksiz	çicrğ
uyta	yırahursuz	üld	kireyar	pahşassız	yhnjcgk
buban	ankleçocu	rehcö	nüt	teşsemek	tmuç
bantü	sarbuçlu	ütıek	ıbek	teşritemek	jeö
roçle	zaşutmak	ratfndgblö	gaya	bemenmek	şıcjc

APPENDIX J

REAL-WORD STIMULI USED IN EXPERIMENT 1

STIMULUS SET 1		STIMULUS SET 2	
Low Frequency	High Frequency	Low Frequency	High Frequency
lokma	duman	amfi	komisyon
münazara	hamam	kement	model
paranoya	marka	ütopya	yastık
çim	tel	basen	damar
bando	sınav	yel	iplik
ozan	yıldız	pala	teknoloji
yorgan	kanun	tekke	aktör
tespih	motor	entrika	desen
pire	travma	akçe	darbe
ülser	cam	kiriş	fırsat
gravür	israf	şehzade	lise
olta	öfke	külot	kavga
sarnıç	çete	fitne	telgraf
lama	tepki	kravat	mezhep
teyp	enfeksiyon	şemsiye	tüccar
yosun	kokteyl	meze	tehlike
ruble	küme	çıban	bel
antrepo	cemaat	paçavra	maraton
filozof	damat	çıra	ağrı
peruk	niyet	iblis	ışık
sivilce	bellek	yayla	fosil
bere	pasaport	fenomen	dava
korse	müze	paspas	cadı
zabıt	filtre	kiremit	not
küpe	çorap	apse	fatura

APPENDIX K

THREE SETS OF NOUN ROOTS USED IN EXPERIMENT 2

Animals		Plants		Tools	
Turkish	English	Turkish	English	Turkish	English
ahtapot	octopus	*şeftali*	peach	*törpü*	nail-file
akrep	scorpion	*avokado*	avocado	*poşet*	bag
antilop	antelope	*bezelye*	greenpeas	*zincir*	chain
bülbül	nightingale	*buğday*	wheat	*battaniye*	blanket
bıldırcın	quail	*domates*	tomato	*kereste*	lumber
civciv	chick	*fıstık*	pistachio	*pergel*	compass (drawing)
gergedan	rhino	*fasulye*	beans	*gönye*	set-square
goril	gorilla	*ıspanak*	spinach	*kerpeten*	pliers
jaguar	jaguar	*kaktüs*	cactus	*tornavida*	screwdriver
kelebek	butterfly	*kayısı*	apricot	*kiremit*	roof tile
köstebek	mole	*lahana*	cabbage	*zımba*	stapler
kanguru	kangaroo	*maydanoz*	parsley	*cetvel*	ruler
kunduz	beaver	*mercimek*	lentils	*defter*	notebook
kurbağa	frog	*nohut*	chickpeas	*iskemle*	chair
leopar	leopard	*pırasa*	leek	*sehpa*	coffee-table
orangutan	orangutan	*patates*	potato	*anahtar*	key
porsuk	badger	*patlıcan*	eggplant	*cüzdan*	wallet
salyangoz	snail	*portakal*	orange	*telefon*	telephone
serçe	sparrow	*zencefil*	ginger	*pusula*	compass (navigation)
penguen	penguin	*zeytin*	olives	*sürahi*	jug

APPENDIX L

SUFFIX TEMPLATES, WORD-FORMS AND NON-WORDS USED IN EXPERIMENT 2

SUFFIX TEMPLATE	COND.	ANIMALS	PLANTS	TOOLS
+P3sg+Pres+A1pl	HIGH	*ahtapotuyuz*	*domatesiyiz*	*defteriyiz*
+Acquire+ByDoingSo	HIGH	*kunduzlanarak*	*kaktüslenerek*	*cüzdanlanarak*
+Acquire+Inf2+Dat	HIGH	*porsuklanmaya*	*zencefillenmeye*	*telefonlanmaya*
+Loc+Pres+A1sg	HIGH	*köstebekteyim*	*kayısıdayım*	*kerestedeyim*
+Become+Caus+Past	HIGH	*gergedanlaştırdı*	*ıspanaklaştırdı*	*battaniyeleştirdi*
+Pres+A1pl	HIGH	*orangutanız*	*mercimeğiz*	*pergeliz*
+A3pl+P1sg+Abl	HIGH	*serçelerimden*	*fasulyelerimden*	*iskemlelerimden*
+P3sg+Equ	HIGH	*kurbağasınca*	*bezelyesince*	*zımbasınca*
+A3pl+P2pl+Abl	HIGH	*salyangozlarınızdan*	*zeytinlerinizden*	*törpülerinizden*
+P3sg+Loc+While	HIGH	*kangurusundayken*	*şeftalisindeyken*	*pusulasındayken*
+Acquire+Inf2+Gen	HIGH	*jaguarlanmanın*	*lahanalanmanın*	*sürahilenmenin*
+P2pl+Abl	HIGH	*akrebinizden*	*portakalınızdan*	*anahtarınızdan*
+Become+Inf2+Dat	HIGH	*antiloplaşmaya*	*maydanozlaşmaya*	*kerpetenleşmeye*
+Acquire+Caus+ByDoingSo	HIGH	*gorillendirerek*	*nohutlandırarak*	*sehpalandırarak*
+Pres+Cop+A3pl	HIGH	*kelebektirler*	*pırasadırlar*	*zincirdirler*
+A3pl+P2pl+Gen	HIGH	*penguenlerinizin*	*avokadolarınızın*	*cetvellerinizin*
+A3pl+Past	HIGH	*bülbüllerdi*	*patlıcanlardı*	*gönyelerdi*
+P1sg+Equ	HIGH	*civcivimce*	*patatesimce*	*poşetimce*
+P2sg+Ins	HIGH	*leoparınla*	*fıstığınla*	*kiremitinle*
+Acquire+Caus+Inf1	HIGH	*bıldırcınlandırmak*	*buğdaylandırmak*	*tornavidalandırmak*
+A3pl+Narr	LOW	*serçelermiş*	*avokadolarmış*	*zincirlermiş*
+Acquire+Inf2+Abl	LOW	*kunduzlanmadan*	*kaktüslenmeden*	*kiremitlenmeden*
+Become+Inf1	LOW	*gergedanlaşmak*	*ıspanaklaşmak*	*kerpetenleşmek*
+Gen+Pron+Rel+Abl	LOW	*bülbülünkinden*	*patlıcanınkinden*	*tornavidanınkinden*
+Past+A3pl	LOW	*penguendiler*	*kayısıydılar*	*pergeldiler*
+P3sg+While	LOW	*kurbağasıyken*	*şeftalisiyken*	*pusulasıyken*
+Acquire+Pass+Past	LOW	*jaguarlanıldı*	*nohutlanıldı*	*cüzdanlanıldı*

+P1sg+Past	LOW	*porsuğumdu*	*fıstığımdı*	*anahtarımdı*
+P2pl+Equ	LOW	*akrebinizce*	*patatesinizce*	*törpünüzce*
+Become+Caus+AfterDoingSo	LOW	*antiloplaştırıp*	*zencefilleştirip*	*battaniyeleştirip*
+Acquire+Pass+Aor	LOW	*bıldırcınlanılır*	*lahanalanılır*	*telefonlanılır*
+P2pl+Pres+A1sg	LOW	*salyangozunuzum*	*mercimeğinizim*	*poşetinizim*
+Become+ByDoingSo	LOW	*civcivleşerek*	*buğdaylaşarak*	*sürahileşerek*
+Loc+Adj+Rel+Abl	LOW	*orangutandakinden*	*zeytindekinden*	*iskemledekinden*
+Dat+Pres+Cop	LOW	*kelebeğedir*	*pırasayadır*	*defteredir*
+Become+Aor+While	LOW	*gorilleşirken*	*maydanozlaşırken*	*sehpalaşırken*
+Loc+Pres+A2pl	LOW	*köstebektesiniz*	*portakaldasınız*	*kerestedesiniz*
+A3pl+P2sg+Ins	LOW	*leoparlarınla*	*fasulyelerinle*	*cetvellerinle*
+P3sg+Loc+Narr	LOW	*kangurusundaymış*	*domatesindeymiş*	*zımbasındaymış*
+P3sg+Pres+A2sg	LOW	*ahtapotusun*	*bezelyesisin*	*gönyesisin*
	HNW1	*ahtazonüyif*	*tomanesimıf*	*değleriyul*
	HNW1	*kınsolmanalak*	*kasyuzleneray*	*lolbanmanarak*
	HNW1	*porlürlatmara*	*kayusıgamif*	*keressegemız*
	HNW1	*körtebalgeyif*	*merfimefal*	*yerçelil*
	HNW1	*senyelerisnan*	*favekselerifçen*	*ilhurlelerircen*
	HNW2	*orangutanıf*	*şeftalifindeykan*	*zımbasısya*
	HNW2	*jaguargancanin*	*maydanozranlası*	*sürahituyrenin*
	HNW2	*antilopranlası*	*pırasavirlas*	*sehpalasyırabak*
	HNW2	*leoparirma*	*avokadolericizat*	*gönyeliryi*
	HNW2	*bıldırcınganfirmaş*	*patlıcantismı*	*poşetakte*
	HNW3	*aşdebinizden*	*beduryesince*	*celefirlanmaya*
	HNW3	*kelesaldırlar*	*yahatalanmanın*	*nastülerinizden*
	HNW3	*gansienlerinizin*	*siltakalınızdan*	*anortanınızdan*
	HNW3	*bülpıklardı*	*cuvetlendirerek*	*kinpetirleşmeye*
	HNW3	*kelesallarlec*	*bunrehlendirmek*	*celhurlarınızın*
	LNW1	*serfelennış*	*kasyuzlenmeven*	*perdumbiler*
	LNW1	*kındünmatgadan*	*kakıdulkılas*	*loğdanlanulmı*
	LNW1	*pornuğezku*	*miksimehinibam*	*ilhurledekifçen*
	LNW1	*kistemorfesitiz*	*favekkeleyesre*	*deşlerezır*

LNW1	*attazonuvin*	*somateğincazmif*	*kenustekaginif*
LNW2	*jaguarpanıkfı*	*avokadolercıs*	*poşetimivaz*
LNW2	*antilopganlısıs*	*patlıcanünmindel*	*sürahitevekak*
LNW2	*bıldırcınganıluf*	*şeftalisitban*	*sehpalaşırfan*
LNW2	*orangutanfadingen*	*pırasayazir*	*zımbasündakmıl*
LNW2	*leoparsariyka*	*maydanozdabarcen*	*gönyesuvin*
LNW3	*pülpığınkinden*	*cuvetlenildi*	*kinpetirleşmek*
LNW3	*gansiendiler*	*yahatalanılır*	*anortanımdı*
LNW3	*aşdebinizce*	*bunrehleşerek*	*nastünüzce*
LNW3	*naçgivleşerek*	*siltakaldasınız*	*celefirlanılır*
LNW3	*kelesaladır*	*beduryesisin*	*celhurlarınla*

APPENDIX M

INSTRUCTIONS USED IN EXPERIMENT 2

Turkish: “Ekranın ortasında önce bir * işareti, sonra da yanyana 10-15 harf göreceksiniz. Bu harflerin (ne kadar tuhaf ve uzun da olsa) Türkçe bir kelime olup olamayacağına karar vermeniz gerekiyor.

Örneğin ‘bilgilendirmeliyiz’ ve ‘ağaçlandırmalıyız’ gayet mantıklı ve ara sıra kullanılabilecek kelimeler. Oysa tam aynı yapıda olan ‘orangutanlandırmalıyız’ son derece tuhaf bir kelime. Ama yine de Türkçenin kök-ek birleşim kurallarına uygun, ve mecbur kalınırsa kullanılabilecek bir kelime. Yani sonuçta böyle bir kelime imkansız değil. Öte yandan ‘yengeçasdfasfd’ veya ‘asfdasdflendirmeliyiz’ Türkçe birer kelime değil, hiçbir şart altında da olamaz.

İşte sizden neyin kelime olduğuna, neyin kelime olmadığına bu şekilde karar vermenizi istiyoruz. Ekranda gördüğünüz şey (çok uzun ve tuhaf da olsa) bir kelime ise “2” tuşuna, tamamen imkansız bir harf yığını ise "9" tuşuna basın.

Deney, sabrınızı zorlamamak için sadece 70 kelimeden oluşuyor ve genellikle 5-6 dakika sürüyor. 10 kelimelik alıştırmaya başlamak için lütfen herhangi bir tuşa basın.”

English: “In the middle of the screen, you will first see an asterisk (*) and then a string of 10-15 letters. You must decide whether or not this letter string can be a Turkish word (even if a very strange and long one).

For instance, bilgilendirmeliyiz ‘we should inform’ and ağaçlandırmalıyız ‘we should plant (trees)’ are perfectly logical words that can be used from time to time. However, the identically formed orangutanlandırmalıyız ‘we should orangutanize’ is an extremely strange word. Still, it is a valid word that complies with the root-suffix combination rules of Turkish, and can be used when the need arises. In other words, such a word is not impossible. On the other hand, orangutanasdfasfd ‘orangutanasdfasfd’ or asfdasdflendirmeliyiz ‘we should asfdasdf’ are not Turkish words, and can never be.

This is how you should decide if a string is a possible word or not. Please press “2” if what you see can be a word (even if a very long and strange one), and press “9” if it’s a completely impossible bunch of letters.

To avoid testing your patience, we have limited the experiment to only 70 words. It usually takes 5-6 minutes to finish. Please press any key to start the 10-word training section.”

APPENDIX N

DESCRIPTIVE STATISTICS FOR EXPERIMENT 1

Descriptive Statistics for Version 1 of Experiment 1

(n = 48)	LOW	HIGH	WUGGY	SUFFIX	RANDOM
Mean	668	615	809	823	635
Standard Error	15,93	14,75	27,66	27,02	16,34
Median	660	589	810	786	608
Standard Deviation	110,37	102,24	191,63	187,26	113,21
Kurtosis	-0,0007	-0,5251	0,1424	-0,1468	-0,0546
Skewness	0,6892	0,5928	0,7470	0,7126	0,5930
Minimum	488	450	554	576	448
Maximum	945	862	1376	1326	964

Descriptive Statistics for Version 2 of Experiment 1

(n = 79)	LOW	HIGH	WUGGY	SUFFIX	RANDOM
Mean	662	607	758	801	626
Standard Error	12,16	12,00	16,55	20,88	11,97
Median	636	592	710	746	597
Standard Deviation	108,10	106,67	147,12	185,55	106,42
Kurtosis	0,53	2,47	0,37	1,06	0,53
Skewness	0,87	1,28	0,99	1,14	0,94
Minimum	458	420	544	515	463
Maximum	953	1007	1184	1399	934

Descriptive Statistics for Version 3 of Experiment 1

(n = 54)	LOW	HIGH	WUGGY	SUFFIX	RANDOM
Mean	684	618	838	910	643
Standard Error	13,23	11,10	19,91	26,57	13,73
Median	681	613	824	896	620
Standard Deviation	97,21	81,56	146,28	195,25	100,92
Kurtosis	0,01	-0,29	-0,70	-0,81	1,10
Skewness	0,45	0,51	0,27	0,40	1,19
Minimum	478	469	571	633	492
Maximum	929	806	1153	1387	924

Descriptive Statistics for Version 4 of Experiment 1

(n = 53)	LOW	HIGH	WUGGY	SUFFIX	RANDOM
Mean	693	632	849	917	644
Standard Error	14,44	14,12	22,60	31,09	14,82
Median	691	598	842	900	623
Standard Deviation	105,15	102,77	164,55	226,31	107,86
Kurtosis	-0,69	-0,79	-0,65	-0,40	0,79
Skewness	0,21	0,25	0,53	0,71	0,96
Minimum	488	416	606	612	473
Maximum	903	826	1225	1466	947

Descriptive Statistics for Version 5 of Experiment 1

(n = 63)	LOW	HIGH	WUGGY	SUFFIX	RANDOM
Mean	682	631	801	848	646
Standard Error	14,94	13,78	21,52	23,15	15,73
Median	663	603	768	810	621
Standard Deviation	118,56	109,39	170,84	183,73	124,83
Kurtosis	0,42	1,46	0,60	-0,35	0,65
Skewness	0,84	1,20	0,92	0,79	0,95
Minimum	479	460	547	565	435
Maximum	1027	982	1336	1292	1040

Descriptive Statistics for Version 6 of Experiment 1

(n = 47)	LOW	HIGH	WUGGY	SUFFIX	RANDOM
Mean	649	604	799	849	642
Standard Error	14,57	13,38	23,51	26,48	17,08
Median	639	585	778	827	634
Standard Deviation	99,89	91,71	161,20	181,56	117,07
Kurtosis	1,38	0,52	-0,48	-0,77	0,85
Skewness	0,99	0,80	0,45	0,38	0,78
Minimum	500	459	544	567	428
Maximum	988	881	1158	1245	1004

Descriptive Statistics for Version 7 of Experiment 1

(n = 61)	LOW	HIGH	WUGGY	SUFFIX	RANDOM
Mean	691	634	836	894	640
Standard Error	15,68	14,37	24,42	28,60	14,43
Median	675	622	803	852	631
Standard Deviation	122,45	112,26	190,71	223,41	112,67
Kurtosis	-0,24	0,03	0,81	-0,59	-0,05
Skewness	0,60	0,55	0,92	0,49	0,42
Minimum	469	426	486	493	386
Maximum	1007	925	1382	1405	929

Descriptive Statistics for Version 8 of Experiment 1

(n = 50)	LOW	HIGH	WUGGY	SUFFIX	RANDOM
Mean	673	611	808	859	627
Standard Error	15,05	12,48	17,38	22,12	13,10
Median	648,5	597	782	835,5	600,5
Standard Deviation	106,41	88,22	122,93	156,38	92,67
Kurtosis	0,43	-0,20	0,64	0,13	0,18
Skewness	0,94	0,57	1,00	0,89	0,78
Minimum	528	447	630	641	495
Maximum	982	831	1175	1268	893

APPENDIX O

DESCRIPTIVE STATISTICS FOR EXPERIMENT 2

Descriptive Statistics for Version 1 of Experiment 2

(n = 119)	HIGH	LOW	PP	RP	PR
Mean	1238	1305	1396	1485	1294
Standard Error	35,06	35,17	44,62	40,67	40,69
Median	1201	1330	1378	1423	1214
Standard Deviation	382,53	383,75	486,83	443,68	443,98
Kurtosis	-0,68	-0,68	-0,89	-0,63	0,51
Skewness	0,40	0,15	0,33	0,25	0,89
Minimum	573	548	594	650	604
Maximum	2100	2164	2445	2545	2682

Descriptive Statistics for Version 2 of Experiment 2

(n = 109)	HIGH	LOW	PP	RP	PR
Mean	1196	1272	1341	1447	1266
Standard Error	29,26	30,44	36,87	33,62	35,25
Median	1171	1239	1338	1431	1272
Standard Deviation	305,55	317,90	385,00	351,07	368,10
Kurtosis	-0,45	-0,52	-0,61	-0,67	-0,06
Skewness	0,36	0,30	0,25	0,02	0,48
Minimum	613	643	661	699	648
Maximum	2028	2020	2204	2247	2432

Descriptive Statistics for Version 3 of Experiment 2

(n = 144)	HIGH	LOW	PP	RP	PR
Mean	1145	1218	1318	1463	1386
Standard Error	26,45	27,75	35,66	33,95	35,01
Median	1116	1201,5	1256	1434,5	1350,5
Standard Deviation	317,49	333,11	428,03	407,45	420,18
Kurtosis	0,13	-0,34	-0,71	-0,56	-0,83
Skewness	0,53	0,16	0,38	0,20	0,29
Minimum	519	518	594	603	642
Maximum	2139	2087	2401	2554	2340

Descriptive Statistics for Version 4 of Experiment 2

(n = 114)	HIGH	LOW	PP	RP	PR
Mean	1251	1300	1392	1546	1433
Standard Error	32,53	32,15	41,59	39,03	41,17
Median	1190	1286	1370	1560	1388,5
Standard Deviation	347,41	343,34	444,10	416,76	439,65
Kurtosis	-0,54	-0,57	-0,18	-0,98	-0,50
Skewness	0,49	0,38	0,55	0,07	0,48
Minimum	684	677	605	755	702
Maximum	2134	2195	2558	2350	2561

Descriptive Statistics for Version 5 of Experiment 2

(n = 98)	HIGH	LOW	PP	RP	PR
Mean	1327	1375	1382	1579	1547
Standard Error	39,88	38,10	44,79	45,57	50,15
Median	1262	1311,5	1302	1566	1513,5
Standard Deviation	394,85	377,20	443,49	451,20	496,48
Kurtosis	-0,45	-0,61	-0,44	-0,63	-0,82
Skewness	0,42	0,23	0,50	0,11	0,27
Minimum	611	589	631	645	642
Maximum	2315	2229	2543	2599	2737

Descriptive Statistics for Version 6 of Experiment 2

(n = 109)	HIGH	LOW	PP	RP	PR
Mean	1139	1215	1228	1436	1395
Standard Error	31,00	33,68	41,78	44,15	42,94
Median	1092	1203	1131	1418	1325
Standard Deviation	323,65	351,66	436,22	460,95	448,37
Kurtosis	-0,24	-0,69	-0,01	-0,15	-0,68
Skewness	0,58	0,36	0,78	0,53	0,51
Minimum	602	621	569	683	677
Maximum	2104	1981	2562	2886	2471

Descriptive Statistics for Version 7 of Experiment 2

(n = 101)	HIGH	LOW	PP	RP	PR
Mean	1327	1370	1520	1556	1405
Standard Error	35,64	36,86	49,07	44,56	46,99
Median	1348	1398	1457	1527	1386
Standard Deviation	358,19	370,50	493,15	447,86	472,28
Kurtosis	-0,65	-0,96	-1,10	-0,90	-0,65
Skewness	0,17	0,03	0,17	0,05	0,39
Minimum	653	652	676	688	671
Maximum	2272	2106	2549	2582	2582

Descriptive Statistics for Version 8 of Experiment 2

(n = 134)	HIGH	LOW	PP	RP	PR
Mean	1301	1365	1512	1546	1353
Standard Error	28,33	30,18	40,69	35,62	35,83
Median	1297	1358,5	1499,5	1535	1313
Standard Deviation	328,00	349,37	471,05	412,36	414,82
Kurtosis	-0,55	-0,92	-0,12	-0,75	-0,43
Skewness	0,28	0,10	0,46	0,18	0,39
Minimum	713	657	692	781	637
Maximum	2160	2148	2937	2543	2636

Descriptive Statistics for Version 9 of Experiment 2

(n = 116)	HIGH	LOW	PP	RP	PR
Mean	1204	1254	1414	1565	1478
Standard Error	29,45	30,35	40,57	39,24	40,98
Median	1184	1280	1384	1591,5	1453
Standard Deviation	317,19	326,89	436,95	422,73	441,36
Kurtosis	0,04	0,15	-0,87	-0,78	-0,67
Skewness	0,37	0,39	0,22	0,10	0,27
Minimum	629	627	652	789	734
Maximum	2215	2337	2495	2632	2794

Descriptive Statistics for Version 10 of Experiment 2

(n = 112)	HIGH	LOW	PP	RP	PR
Mean	1280	1342	1461	1681	1562
Standard Error	25,03	24,56	34,44	33,51	33,50
Median	1263	1344	1482	1736	1532
Standard Deviation	264,90	259,88	364,52	354,62	354,56
Kurtosis	-0,26	0,19	-0,44	-0,37	-0,61
Skewness	0,34	0,23	0,16	-0,31	0,35
Minimum	735	753	712	869	837
Maximum	2054	2113	2405	2455	2395

Descriptive Statistics for Version 11 of Experiment 2

(n = 101)	HIGH	LOW	PP	RP	PR
Mean	1330	1401	1384	1556	1530
Standard Error	30,05	29,88	39,18	38,49	38,30
Median	1299	1356	1290	1499	1535
Standard Deviation	301,96	300,31	393,74	386,86	384,91
Kurtosis	-0,68	-0,67	-0,76	-0,22	-0,97
Skewness	0,39	0,23	0,43	0,47	0,24
Minimum	831	882	715	816	806
Maximum	2034	2075	2225	2695	2471

Descriptive Statistics for Version 12 of Experiment 2

(n = 90)	HIGH	LOW	PP	RP	PR
Mean	1375	1446	1446	1576	1546
Standard Error	40,47	40,77	47,24	42,57	48,34
Median	1372	1445	1387	1604	1563
Standard Deviation	383,89	386,74	448,13	403,90	458,60
Kurtosis	-0,02	-0,46	-0,86	-0,85	-0,95
Skewness	0,51	0,14	0,21	-0,25	0,08
Minimum	731	665	652	685	767
Maximum	2401	2326	2408	2349	2588

REFERENCES

Aksan, Y., Aksan, M., Koltuksuz, A., Sezer, T., Mersinli, Ü., Demirhan, U. U., ... & Kurtoglu, Ö. (2012, May). Construction of the Turkish National Corpus (TNC). In *LREC* (pp. 3223-3227).

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

Amenta, S., & Crepaldi, D. (2012). Morphological processing as we know it: An analytical review of morphological effects in visual word identification.

Anderson, S. R. (1988). Morphology as a parsing problem. *Linguistics*, 26(4), 521-544.

Andrews, S. (1992). Frequency and neighborhood effects on lexical access: Lexical similarity or orthographic redundancy? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 18(2), 234.

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, R. H. (2001). *Word frequency distributions*. Dordrecht: Kluwer Academic Publishers.

Baayen, R. H. (2010). A real experiment is a factorial experiment. *The Mental Lexicon*, 5(1), 149-157.

Baayen, R. H. (2014). Experimental and psycholinguistic approaches. In R. Lieber & P. Stekauer (Eds.). *The Oxford handbook of derivational morphology*. Oxford: Oxford University Press, 95-117.

Baayen, R. H., Milin, P., & Ramscar, M. (2016). Frequency in lexical processing. *Aphasiology*, *30*(11), 1174-1220.

Balota, D. A., & Chumbley, J. I. (1984). Are lexical decisions a good measure of lexical access? The role of word frequency in the neglected decision stage. *Journal of Experimental Psychology: Human Perception and Performance*, 10(3), 340.

Balota, D. A., Pilotti, M., & Cortese, M. J. (2001). Subjective frequency estimates for 2,938 monosyllabic words. *Memory & Cognition*, 29(4), 639-647.

Banguoğlu, T. (1974). *Türkçenin grameri*. Istanbul: Baha Matbaası.

Broadbent, D. E. (1967). Word-frequency effect and response bias. *Psychological Review*, 74(1), 1.

Burani, C., & Caramazza, A. (1987). Representation and processing of derived words. *Language and Cognitive Processes*, 2(3-4), 217-227.

Butterworth, B. (1983). Lexical representation. In. B. Butterworth (Ed.), *Language production, Vol. 2: Development, writing, and other language processes* (pp. 257-294). London: Academic Press

Bybee, J. L. (2011). Usage-based theory and grammaticalization. In B. Heine & H. Narrog (Eds.), *The Oxford Handbook of Grammaticalization*. Oxford: Oxford University Press, pp. 69.78

Caramazza, A., Laudanna, A., & Romani, C. (1988). Lexical access and inflectional morphology. *Cognition*, 28(3), 297-332.

Cattell, J. M. (1886). The time it takes to see and name objects. *Mind*, 11(41), 63-65.

Clark, H. H. (1973). The language-as-fixed-effect fallacy: A critique of language statistics in psychological research. *Journal of Verbal Learning and Verbal Behavior*, 12(4), 335-359.

Coleman, E. B. (1964). Generalizing to a language population. *Psychological Reports*, 14(1), 219-226.

Coltheart, M., Davelaar, E., Jonasson, T., & Besner, D. (1977). Access to the internal lexicon. In S. Dornič (Ed.), *Attention and Performance VI: Proceedings of the Sixth International Symposium on Attention and Performance,* Vol. 6.

Cutler, A. (1981). Making up materials is a confounded nuisance, or: Will we able to run any psycholinguistic experiments at all in 1990?. *Cognition*, 10(1-3), 65-70.

Davis, C. J., Perea, M., & Acha, J. (2009). Re(de)fining the orthographic neighborhood: the role of addition and deletion neighbors in lexical decision and reading. *Journal of Experimental Psychology: Human Perception and Performance*, 35(5), 1550.

Dehaene, S. (2009). *Reading in the brain: The new science of how we read*. London: Penguin.

de Leeuw, J. R. (2015). jsPsych: A JavaScript library for creating behavioral experiments in a Web browser. *Behavior Research Methods*, 47(1), 1-12.

de Leeuw, J. R., & Motz, B. A. (2016). Psychophysics in a Web browser? Comparing response times collected with JavaScript and Psychophysics Toolbox in a visual search task. *Behavior Research Methods*, 48(1), 1-12.

del Prado Martín, F., Bertram, R., Häikiö, T., Schreuder, R., & Baayen, R. H. (2004). Morphological family size in a morphologically rich language: the case of

Finnish compared with Dutch and Hebrew. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 30(6), 1271.

del Prado Martín, F. M., Deutsch, A., Frost, R., Schreuder, R., De Jong, N. H., & Baayen, R. H. (2005). Changing places: A cross-language perspective on frequency and family size in Dutch and Hebrew. *Journal of Memory and Language*, 53(4), 496-512.

Donders, F. C. (1969). On the speed of mental processes. *Acta Psychologica*, 30, 412-431.

Durgunoğlu, A. Y. (2003). Recognizing morphologically complex words in Turkish. In *Reading complex words* (pp. 81-92). New York, Springer US.

Durrant, P. (2013). Formulaicity in an agglutinating language: the case of Turkish. *Corpus Linguistics and Linguistic Theory*, 9(1), 1-38.

Eren, E. (2014). *Frequency and length effects of words on fixation durations in Turkish sentence reading* (Unpublished master's thesis). Middle East Technical University, Ankara, Turkey. Retrieved from http://etd.lib.metu.edu.tr/upload/12617750/index.pdf

Erten, B., Bozşahin, C., Zeyrek, D. (2014). Turkish resources for visual word recognition. *LREC 2014, The 9th edition of the Language Resources and Evaluation Conference*, 26-31 May, Reykjavik, Iceland, pp. 2106-2110.

Evans, V., & Green, M. (2006). *Cognitive linguistics: An introduction*. Edinburgh, Lawrence Erlbaum Associates Publishers.

Ferrand, L., New, B., Brysbaert, M., Keuleers, E., Bonin, P., Méot, A., ... & Pallier, C. (2010). The French Lexicon Project: Lexical decision data for 38,840 French words and 38,840 pseudowords. *Behavior Research Methods*, 42(2), 488-496.

Ford, M. A., Marslen-Wilson, W. D., & Davis, M. H. (2003). Morphology and frequency: Contrasting methodologies. *Trends in Linguistics Studies and Monographs*, 151, 89-124.

Ford, M. A., Davis, M. H., & Marslen-Wilson, W. D. (2010). Derivational morphology and base morpheme frequency. *Journal of Memory and Language*, 63(1), 117-130.

Forster, K. I., & Chambers, S. M. (1973). Lexical access and naming time. *Journal of Verbal Learning and Verbal Behavior*, 12(6), 627-635.

Forster, K. I. (2000). The potential for experimenter bias effects in word recognition experiments. *Memory & Cognition*, 28(7), 1109-1115.

Frauenfelder, U. H., & Schreuder, R. (1992). Constraining psycholinguistic models of morphological processing and representation: The role of productivity. In *Yearbook of Morphology 1991* (pp. 165-183). Dordrecht, Springer Netherlands.

Friedenberg, J., & Silverman, G. (2011). *Cognitive science: An introduction to the study of mind*. London, Sage.

Gacan, P. (2014). *The morphological processing of derived words in L1 Turkish and L2 English* (Unpublished master's thesis). Middle East Technical University, Ankara, Turkey. Retrieved from http://etd.lib.metu.edu.tr/upload/12617429/index.pdf

Gernsbacher, M. A. (1984). Resolving 20 years of inconsistent interactions between lexical familiarity and orthography, concreteness, and polysemy. *Journal of Experimental Psychology*: 113(2), 256.

Giraudo, H., & Grainger, J. (2000). Effects of prime word frequency and cumulative root frequency in masked morphological priming. *Language and Cognitive Processes*, 15(4-5), 421-444.

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

Göksel, A., & Kerslake, C. (2005). *Turkish: A comprehensive grammar*. London, Routledge.

Grainger, J. (1990). Word frequency and neighborhood frequency effects in lexical decision and naming. *Journal of Memory and Language*, 29(2), 228-244.

Güngör, T. (2003). Lexical and morphological statistics for Turkish. *Proceedings of TAINN*, 409-412.

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

Gürel, A. & Uygun, S. (2013). Representation of multimorphemic words in the mental lexicon: Implications for second language acquisition of morphology. S. Baiz, N. Goldman, & R. Hawkes (Eds). *Proceedings of the 37th Annual Conference on Language Development* (pp. 122-133). Somerville: Cascadilla Press.

Hankamer, J. (1989). Morphological parsing and the lexicon. In *Lexical representation and process* (pp. 392-408). Cambridge, MIT Press.

Hasson, U., & Giora, R. (2007). Experimental methods for studying the mental representation of language. *Methods in Cognitive Linguistics*. Amsterdam: John Benjamins, 302-322.

Hayasi, T. (1996). The dual status of possessive compounds in Modern Turkish. *Symbolae Turcologicae. Studies in honor of Lars Johanson on the occasion of his sixtieth birthday*, 6, 119-129.

Howes, D. H., & Solomon, R. L. (1951). Visual duration threshold as a function of word-probability. *Journal of Experimental Psychology*, 41(6), 401.

Howes, D. (1954). On the interpretation of word frequency as a variable affecting speed of recognition. *Journal of Experimental Psychology*, 48(2), 106.

Keuleers, E., Brysbaert, M., & New, B. (2010). SUBTLEX-NL: A new measure for Dutch word frequency based on film subtitles. *Behavior Research Methods*, 42(3), 643-650.

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

Khmaladze, E. V. (1988). The statistical analysis of a large number of rare events. *Stichting Mathematisch Centrum*, (R 8804), 1-21.

Kırkıcı, B., & Clahsen, H. (2013). Inflection and derivation in native and non-native language processing: Masked priming experiments on Turkish. Bilingualism: *Language and Cognition*, 16(04), 776-791.

Laudanna, A., & Burani, C. (1985). Address mechanisms to decomposed lexical entries. *Linguistics*, 23(5), 775-792.

Lewis, G. (1999). *The Turkish language reform: A catastrophic success*. Oxford, Oxford University Press.

MacKay, D. G. (1978). Derivational rules and the internal lexicon. *Journal of Verbal Learning and Verbal Behavior*, 17(1), 61-71.

Markman, A. B. (2006). Psychology of knowledge representation. In Nadel, L. (Ed.). *Encyclopedia of cognitive science (Vol. 3)*. Retrieved August 17, 2016, from https://www.researchgate.net/publication/228053810
_Encyclopedia_of_Cognitive_Science

McGinnies, E., Comer, P. B., & Lacey, O. L. (1952). Visual-recognition thresholds as a function of word length and word frequency. *Journal of Experimental Psychology*, 44(2), 65.

Meunier, F., & Segui, J. (1999). Morphological priming effect: The role of surface frequency. *Brain and Language*, 68(1), 54-60.

Meyer, D. E., Osman, A. M., Irwin, D. E., & Yantis, S. (1988). Modern mental chronometry. *Biological Psychology*, 26(1-3), 3-67.

Murrell, G. A., & Morton, J. (1974). Word recognition and morphemic structure. *Journal of Experimental Psychology*, 102(6), 963.

Nisbett, R. E., & Wilson, T. D. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological Review*, 84(3), 231.

Niswander, E., Pollatsek, A., & Rayner, K. (2000). The processing of derived and inflected suffixed words during reading. *Language and Cognitive Processes*, 15(4-5), 389-420.

Oflazer, K., Say, B., Hakkani-Tür, D. Z., & Tür, G. (2003). Building a Turkish treebank. In *Treebanks* (pp. 261-277). Dordrecht, Springer Netherlands.

Özer, S. (2010). *Morphological priming in Turkish nominal compound processing* (Unpublished master's thesis). Middle East Technical University, Ankara, Turkey. Retrieved from https://etd.lib.metu.edu.tr/upload/12612631/index.pdf

Pierce, J. (1960). *A frequency count of Turkish words*. Ankara: Milli Eğitim Müdürlüğü.

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

Pinker, S., & Prince, A. (1994). Regular and irregular morphology and the psychological status of rules of grammar. In S. D. Lima, R. Corrigan & G. K. Iverson (Eds.), *The reality of linguistic rules,* pp. 321-351.

Popescu, I. I. (2009). *Word frequency studies (Vol. 64).* Berlin, Boston: De Gruyter Mouton.

Raaijmakers, J. G., Schrijnemakers, J. M., & Gremmen, F. (1999). How to deal with "the language-as-fixed-effect fallacy": Common misconceptions and alternative solutions. *Journal of Memory and Language*, 41(3), 416-426.

Rescorla, M. (2015). The Computational Theory of Mind. In E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy (winter 2015 edition)*. Retrieved August 17, 2016, from http://plato.stanford.edu/archives/win2015/entries/computational-mind

Rice, G. A., & Robinson, D. O. (1975). The role of bigram frequency in the perception of words and nonwords. *Memory & Cognition*, 3(5), 513-518.

Rubin, D. C. (1980). 51 properties of 125 words: A unit analysis of verbal behavior. *Journal of Verbal Learning and Verbal Behavior,* 19(6), 736-755.

Sak, H., Güngör, T., & Saraçlar, M. (2007, February). Morphological disambiguation of Turkish text with perceptron algorithm. In *International Conference on Intelligent Text Processing and Computational Linguistics* (pp. 107-118). Springer Berlin Heidelberg.

Sak, H., Güngör, T., & Saraçlar, M. (2008). Turkish language resources: Morphological parser, morphological disambiguator and web corpus. In

Advances in natural language processing (pp. 417-427). Springer Berlin Heidelberg.

Schiepers, C. (1980). Response latency and accuracy in visual word recogniton. *Perception & Psychophysics*, 27(1), 71-81.

Schreuder, R., Grendel, M., Poulisse, N., Roelofs, A., & van de Voort, M. (1990). Lexical processing, morphological complexity and reading. In D. Balota, G. B. Flores d'Arcais & K. Rayner (Eds.) *Comprehension processes in reading* (pp. 125-141). Hillsdale, NJ: Lawrence Erlbaum Associates Inc.

Schreuder, R., & Baayen, R. H. (1995). Modeling morphological processing. In L. B. Feldman (Ed.), *Morphological aspects of language processing*, Psychology Press, p. 131.

Schreuder, R., & Baayen, R. H. (1997). How complex simplex words can be. *Journal of Memory and Language*, 37(1), 118-139.

Sears, C. R., Hino, Y., & Lupker, S. J. (1995). Neighborhood size and neighborhood frequency effects in word recognition. *Journal of Experimental Psychology: Human Perception and Performance*, 21(4), 876.

Segui, J., Mehler, J., Frauenfelder, U., & Morton, J. (1982). The word frequency effect and lexical access. *Neuropsychologia*, 20(6), 615-627.

Seidenberg M. S. (1987) Sublexical structures in visual word recognition: access units or orthographic redundancy? In Coltheart M. (Ed.), *Attention and performance: the psychology of reading*. Vol. 7. Hillsdale, NJ: Erlbaum.

Seidenberg, M. S., & McClelland, J. L. (1989). A distributed, developmental model of word recognition and naming. *Psychological Review*, 96(4), 523.

Sezer, T. & Sezer, B. (2013). TS Corpus: Herkes için Türkçe derlem. In *Proceedings of the 27th National Linguistics Conference* (pp. 217-225).

Snodgrass, J. G., & Jarvella, R. J. (1972). Some linguistic determinants of word classification times. *Psychonomic Science*, 27(4), 220-222.

Solomyak, O., & Marantz, A. (2010). Evidence for early morphological decomposition in visual word recognition. *Journal of Cognitive Neuroscience*, 22(9), 2042-2057.

Stemberger, J. P., & MacWhinney, B. (1986). Frequency and the lexical storage of regularly inflected forms. *Memory & Cognition*, 14(1), 17-26.

Swift, L. B. (1963). *A reference grammar of modern Turkish (Vol. 19)*. Bloomington: Indiana University Publications.

Taft, M., & Forster, K. I. (1975). Lexical storage and retrieval of prefixed words. *Journal of Verbal Learning and Verbal Behavior,* 14(6), 638-647.

Taft, M. (1979). Recognition of affixed words and the word frequency effect. *Memory & Cognition*, 7(4), 263-272.

Taft, M. (2004). Morphological decomposition and the reverse base frequency effect. *Quarterly Journal of Experimental Psychology Section A*, 57(4), 745-765.

Tolgay, E. A. (2015). *A Turkish database for psycholinguistic studies: a corpus based study on frequency, age of acquisition, and imageability* (Doctoral dissertation). Middle East Technical University, Ankara, Turkey. Retrieved from http://etd.lib.metu.edu.tr/ upload/12618663/index.pdf

van Casteren, M., & Davis, M. H. (2007). Match: A program to assist in matching the conditions of factorial experiments. *Behavior Research Methods*, 39(4), 973-978.

Vannest, J., & Boland, J. E. (1999). Lexical morphology and lexical access. *Brain and Language*, 68(1), 324-332.

Vannest, J., Newport, E. L., Newman, A. J., & Bavelier, D. (2011). Interplay between morphology and frequency in lexical access: The case of the base frequency effect. *Brain Research*, 1373, 144-159.

Zipf, G. K. (1949). *Human behavior and the principle of least effort: An introduction to human ecology*. Cambridge, Mass.: Addison-Wesley Press, Inc.