ADAPTING AND TESTING PSYCHOLINGUISTIC TOOLBOXES FOR TURKISH VISUAL WORD RECOGNITION STUDIES

A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF INFORMATICS INSTITUTE
OF
MIDDLE EAST TECHNICAL UNIVERSITY

BY

BEGÜM ERTEN

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF
MASTER OF SCIENCE
IN
THE DEPARTMENT OF COGNITIVE SCIENCE

SEPTEMBER 2013

ADAPTING AND TESTING PSYCHOLINGUISTIC TOOLBOXES FOR TURKISH VISUAL WORD RECOGNITION STUDIES

Submitted by **Begüm Erten** in partial fulfillment of the requirements for the degree of **Master of Science in Cognitive Science, Middle East Technical University** by,

Prof. Dr. Nazife Baykal ______________________
Director, **Informatics Institute**

Prof. Dr. Cem Bozşahin ______________________
Head of Department, **Cognitive Science**

Prof. Dr. Cem Bozşahin ______________________
Supervisor, **Cognitive Science, METU**

Prof. Dr. Deniz Zeyrek ______________________
Co-Supervisor, **Cognitive Science, METU**

Examining Committee Members:

Asst. Prof. Dr. Cengiz Acartürk ______________________
Cognitive Science, METU

Prof. Dr. Cem Bozşahin ______________________
Cognitive Science, METU

Prof. Dr. Deniz Zeyrek ______________________
Cognitive Science, METU

Asst. Prof. Dr. Murat Perit Çakır ______________________
Cognitive Science, METU

Dr. Ruket Çakıcı ______________________
Computer Engineering, METU

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Name, Last Name: BEGÜM ERTEN

Signature :

ABSTRACT

ADAPTING AND TESTING PSYCHOLINGUISTIC TOOLBOXES FOR TURKISH VISUAL WORD RECOGNITION STUDIES

Erten, Begüm
M.Sc., Department of Cognitive Science
Supervisor : Prof. Dr. Cem Bozşahin
Co-Supervisor : Prof. Dr. Deniz Zeyrek

September 2013, 68 pages

This study presents two different software programs to be used in Turkish visual word recognition studies: *KelimetriK* and *Wuggy with a Turkish plug-in extension*. KelimetriK is a query-based software program developed as part of this thesis. KelimetriK provides word and bi-gram/tri-gram frequencies, orthographic neighborhood (ON), orthographic relatedness (transposed letter similarity and subset/superset similarity) and OLD20 (orthographic Levensthein Distance 20) scores. Wuggy is a pseudoword (i.e. wordlike non-words) generator that is adaptable to multiple languages (Keuleers, & Brstbaert, 2010). The pseudowords are generated from one or more user defined reference words by considering the sub-syllabic elements' bi-gram frequencies (the onset-nucleus-coda pattern). Developing a Turkish plug-in for the Wuggy software was also part of this thesis. KelimetriK and the Turkish plug-in for Wuggy were used on a lexical decision visual word recognition study. The aim of the study was to investigate the specific influence of word-frequency, ON, OLD20 and imageability on response time and accuracy scores . It was hypothesized that all the tested linguistic variables will influence the Turkish visual word recognition data.

Keywords: visual word recognition, pseudoword generation, lexical decision, linguistic variables, Turkish words and pseudowords

ÖZ

PSİKOLİNGUİSTİK YAZILIMLARIN GÖRSEL KELİME TANIMLAMA ÇALIŞMALARI İÇİN TÜRKÇE'YE UYARLANMASI VE TEST EDİLMESİ

Erten, Begüm
Yüksek Lisans, Bilişsel Bilimler Bölümü
Tez Yöneticisi : PROF. DR. CEM BOZŞAHİN
Ortak Tez Yöneticisi : PROF DR. DENİZ ZEYREK

Eylül 2013, 68 sayfa

Bu çalışmada Türkçe görsel kelime tanıma çalışmalarında kullanılabilecek olan iki yazılım sunulmaktadır: KelimetriK ve Türkçe eklenti paketiyle Wuggy. KelimetriK bu tez çalışması için geliştirilmiş sorgu tabanlı bir yazılımdır. Yazılımın çıktı olarak verdiği dilbilimsel parametreler: kelime ve bi-gram frekans değerleri, ortografik komşuluk, ortografik ilişkilendirme, (altküme/üstküme benzerliği) ve OLD20 (ortografik Levenshtein uzaklık 20) değerleridir. Wuggy birçok dile adapte olabilen kelimemsi üreticisidir. Kelimemsiler dilde bulunmayan fakat dilin kuralları açısından doğru olan kelimelerdir. Wuggy yazılımında kelimemsiler kullanıcı tarafından belirtilmiş olan bir veya daha fazla kelimenin hece altı elemenlarının (açış, çekirdek ve bitiş) bi-gram geçiş frekansını dikkate alarak üretilir. Bu tez çalışmasının bir parçasi olarak Wuggy yazılımına bir Türkçe eklenti paketi geliştirilmiştir. KelimetriK ve Türkçe eklenti paketiyle Wuggy yazılımları sözcüksel karar testi uygulanan bir kelime tanıma çalışmasında kullanılmıştır. Bu çalışmanın amacı kelime frekansı, ON, OLD20 ve imgelenebilirlik değerinin tepki süresi ve doğruluk değerlerine olan etkisini incelemektir. Hipotezimiz, test edilen bütün dilbilimsel değişkenlerin Türkçe görsel kelime tanımlama verilerini etkilemesi yönündedir.

Anahtar Kelimeler: görsel kelime tanımlama, sözcüksel karar, kelimemsi üretimi, dilbilimsel değişkenler, Türkçe kelime ve kelimemsiler

to my husband Umut Uyumaz,
my father Münir Erten
and
my mother Müjgan Nurgül Erten

ACKNOWLEDGMENTS

I would like to thank Cem Bozşahin and Deniz Zeyrek for their support, encouragement and guidance while conducting my studies of the thesis project. I would like to thank Deniz Zeyrek again for her helpful comments and encouragements that highly motivated me to pursue my studies on the field of linguistics.

I would also like to thank Ruket Çakıcı for her motivating lecture on the statistical methods in natural language processing which helped me a lot while developing the Turkish psycholinguistic software programs.

I would also like to thank Cem Bozşahin and his students for their extra work to obtain frequency values of the words in the Turkish stem list. I would also like to thank Murat Perit Çakır and Cengiz Acartürk for their effort on reviewing my thesis work.

Finally, I would like to thank my husband Umut Uyumaz for his support and encouragement since the beginning of my thesis studies, I am especially grateful for his helps on the programming part of the work. I would also like to thank my parents Müjgan Nurgül and Münir Erten who were always with me and helped me to stay persistent on my academic goals.

TABLE OF CONTENTS

List of Tables

List of Figures

CHAPTER 1

Introduction

This thesis is about developing, adapting and testing psycholinguistic tools for Turkish visual word recognition studies. The developed software programs are *KelimetriK* and *Wuggy with Turkish plug-in extension*. These software programs were developed as a helpful source to the experimenters who are planning to prepare Turkish psycholinguistic studies. A lexical decision task was conducted on a group of subjects, several hypothesis testing models were provided to the users.[1]

KelimetriK is a query-based software program designed to derive several outputs concerning the lexical properties of a queried word. These lexical properties are word-frequency, bi-gram/tri-gram frequency (decomposing the word into number of elements), orthographic neighborhood (ON: number of words that can be obtained by substituting a single element), orthographic relatedness (transposed letter similarity (e.g. ”corny” and ”crony”), subset/superset similarity (e.g. ”old” and ”cold”) and OLD20 (orthographic Levenshtein distance 20) scores.[2]

Wuggy is a pseudoword generator software in which the non-words are generated in a short duration by calculating the transition frequencies of words' sub-syllabic elements. The software can be adapted to other languages by providing a language's syllabified word list with frequency values and defining an orthographic rules of the words' sub-syllabic pattern. Wuggy is currently available for Basque, Dutch, English, French, German, Serbian, Spanish and recently for Turkish words.

The stimuli set of the lexical decision task was prepared using the KelimetriK and the Turkish version of Wuggy. The stimuli set consists of 250 Turkish words and 250 non-words. The words varied in the dimension of word-frequency (per-million), ON, OLD20 and imageability.[3] Word-frequency, ON and OLD20 scores were obtained using the KelimetriK software. The pseudowords were generated with the Wuggy software with Turkish plug-in, each pseudoword was derived from a word stimulus which was the best possible candidate from the 10 non-

[1] In lexical decision task, subjects decide whether a presented letter string is a word or a pseudoword.

[2] OLD20 is the average of most closest 20 words in the unit of Levenshtein distance which is the transformation of one word to the other with minimum letter change.

[3] Imageability of a word is the extent that it arises a mental image.

words. The following section provides a detailed explanation of pseudowords.

1.1 Pseudowords: the "Wordlike" Non-words

Phonemes, the smallest sound units in a language, are not randomly conjoined in a word, rather they are governed by a set of rules (Lass, 1984). For example, the open class words in English such as nouns and adjectives includes at least a stressed vowel. *Pseudowords* (non-words) are pronounceable string of letters that conforms the *orthographic* and *phonologic* rules of a language without having any semantic representation (Grainger et al., 2003; Keuleers and Brysbaert, 2010).[4] In other words, even though pseudowords does not have any semantic representation, all the sounds and combination of letters are legal according to the language rules. Pseudowords should not be confused with illegal conjunction of strings such as "hsyfh" which cannot be pronounced with a language's phonetic rules.

Pseudowords are useful tools for psycholinguistic tasks. They can be used to investigate human lexical processing, and also many different types of human cognitive processes such as phonological reading performance (Good and Kaminski, 2002; Simos et al., 2000). Pseudowords can be used to assess the phonological decoding process which is deficient in dyslexic patients (Grainger et al., 2003). Furthermore, pseudowords can also be used to measure reading proficiency of healthy individuals (Good and Kaminski, 2002). Readers who acquired general knowledge of phonology rules of a language will also be good at detecting pseudowords.

Pseudowords can also be used to test second language learners' proficiency level. A study on Japanese university students has shown that overestimation score and false alarm rates for English non-words were much more lower for advanced learners then beginners (Stubbe, 2012).

1.2 Pseudowords in Visual Word Recognition Tasks

Pseudowords are useful tools for investigating the underlying mechanism of cognitive lexical processing; the process beginning from pattern recognition of letter conjunctions to accessing its semantic representation (Cibelli, 2012; Mainy et al., 2007). It is necessary to pass a sequence of steps before accessing the mental representation of any recognized word. The processing starts from visual recognition of the letters and then proceeds to recognizing the word's forms and the letter strings. Then, a transformation happens from graphemic word form to corresponding phonological word form, and finally gets access to its corresponding semantic representations (Mainy et al., 2007). Thus, the complete process happens to be on four different layered dimensions of phonological, lexical, grammatical and semantic (Cibelli, 2012).

Pseudowords can be used to manipulate the dimensions of lexical processing on any kind of experimental settings, a contrast between the semantic dimension and rest of the other dimen-

[4] *Orthography* is concerned with how graphemes are combined; *phonology* is concerned with how phonemes are combined.

sions. For example, if the phonology dimension is prior to the orthographic dimension, this can be detected with comparing the recognition duration between orthographic and homophonic pseudowords.

Visual word recognition tasks are ideal experimental paradigms for comparison of lexical access times among words and pseudowords. In visual word recognition tasks subjects are expected to respond to each visually presented letter as quickly and accurately as possible. Rest of the procedure is same with LDT. Two most common types of visual word recognition tasks are lexical decision task (LDT) and speeded naming task (SNT) (Balota et al., 2007). In LDT, subjects are expected decide whether a visually presented string of letters are either a word or a pseudoword. In SNT, subjects are expected to name a visually presented sequence of letters as quickly as possible. The time duration recorded in SNT is assumed to be devoted for grapheme-to-phoneme mental transformation.

There is also a priming version of LDT in which a priming word is presented for a very short duration (like 50 milliseconds) just before the presentation of a target word. Priming research designs are useful to investigate how a given set of neighbor words activate each other (e.g. competition or facilitation). However, they do not provide any information about the influences of orthographically similar words on a single word (Andrews, 1997; Bowers et al., 2005). In other words, priming studies are useful only for studying the behavioral activation patterns on a given stimuli set but not for testing on a single word.

There are different types of behavioral patterns between different types of pseudowords. In one of the previous studies, response times of legal and illegal non-words were compared on a lexical decision task (Gibbs and Van Orden, 1998). The results showed that response time for *illegal pseudowords* such as ”ldfa” were shortest (496 msec) when compared with *legal (orthographic) pseudowords* such as ”dilt” (588 msec) followed by *pseudohomophonic pseudowords* such as ”durt” (698 msec).[5]

1.3 Theories on Pseudoword Processing

There are two models on pseudoword processing: parallel-distributed (*connectionist*) model and dual-route model (Peereman et al., 1998). According to the parallel-distributed model, the processing pathways of words and pseudowords are mutual, the difference arises in the activation amount of orthographic, phonological and semantic routes (Seidenberg and McClelland, 1989). Thus, entirely the same routes are involved for words and pseudowords. However, the amount of activation on these routes are increased for words when compared with pseudowords. According to the dual-route model, different pathways are involved in lexical processing between words and pseudowords, the former traces a lexical route and the latter a sub-lexical route (Peereman et al., 1998).

Neuroimaging and electrophysiological recording technology on recent studies were used to

[5] A pseudohomophonic pseudowords have the same pronunciation with a real word.

investigate the underlying mechanism of lexical and sub-lexical neural pathways of lexical processing (Cibelli, 2012; Hauk et al., 2006; Price et al., 1996). If there are different pathways as the dual-route model predicts, different neural regions should be involved on the processing of words and pseudowords. However, if the lexical processing pathways are mutual as the parallel-distributed modal predicts, the same neural regions should be involved both for words and pseudowords. A recent neuroimaging study compared the brain activation between real words and pseudowords and concluded that only real words activated left posterior middle temporal and angular gyri which is assumed to be involved in lexical semantic processing (Raettig and Kotz, 2008). On the other hand, another study used electrophysiological intracranial technology to compare the neural activation between auditory processed words and pseudowords on a repetition research paradigm (repeating each verbal stimuli after heard), however they could not detect any difference among words and pseudowords as predicted by the connectionist model (Cibelli, 2012).

1.4 Aim of the Behavioral Study

There were two specific goals for conducting the behavioral studies: to provide hypothesis-testing models to psycholinguistic experimenters and to investigate the behavioral effects of several linguistic variables for Turkish words. To be more specific, the first goal was to use the recently developed Turkish psycholinguistic tools on a visual word recognition study. The second goal was to provide answers to specific research questions concerning the effect of several linguistic variables on behavioral results for Turkish words and then comparing these results with other languages.

Some specific research questions were developed after an extensive research on the general literature of visual word recognition studies;

- What is the overall difference on the response time and error rate scores for words and non-words? The response duration of words are predicted to be lower for words than non-words. The error rates are also expected to be lower for words than non-words.

- Is there a specific influence of word-frequency, ON, OLD20 and imageability variables for the behavioral scores for Turkish words? Each of these variables are expected to have a specific influence on the behavioral data.

- Which one of the lexical variables of ON and OLD20 has a stronger influence on words' lexical detection duration. Based on previous studies, both of the variables are expected to influence the response times as observed in in English visual word recognition studies, whereas the effect is expected to be stronger for the OLD20 scores (Andrews, 1997; Yarkoni et al., 2008). To be more specific, words having low ON scores and high OLD20 scores are expected to facilitate the lexical detection process and result in smaller response times.

- What is the interaction of word-frequency with OLD20 and imageability? Based on previous studies, it is expected that word-frequency will interact with OLD20 but not with imageability (McMullen and Bryden, 1987; Yarkoni et al., 2008). For OLD20, the response times are expected to be smaller for words in the high-OLD20 and high-frequency condition.

1.5 Overview of the Chapters

- **Chapter 1** reviews the literature on the common usage of pseudowords. Pseudowords are mostly used in the lexical processing studies. The chapter also gives explanation of the aim of the present thesis work.
- **Chapter 2** reviews some of the common linguistic variables which have influence on the behavioral results. These linguistic variables can be due to the words' orthographic representation (e.g. word-frequency, ON, OLD20) or from other behavioral variables (e.g. age of acquisition).
- **Chapter 3** introduces the KelimetriK software, its interface and what it is used for. The chapter also explains the software's algorithm, and its corpus.
- **Chapter 4** introduces the Wuggy software, its adjustable features. The chapter also reviews the general literature on the alternative methods for generating pseudowords and the other psycholinguistic studies that used the software. Moreover, the chapters also provides a detailed explanation of the algorithm and the recently developed Turkish plug-in.
- **Chapter 5** describes the methods of the behavioral study: selection of the words and the non-words, the subject group and details of the lexical decision task.
- **Chapter 6** reports the general results of the study: descriptive statistics of lexical properties of the stimuli set, the statistical analysis conducted on behavioral scores and the control studies.
- **Chapter 7** discusses the findings of the study in detail, the limitations of the study and some possible future studies.

1.6 Summary

Pesudowords are lexically meaningless wordlike non-words composed according to a set of phonetic and orthographic rules. Pseudowords are useful tools for lexical processing studies and can also be used for testing the proficiency level of beginner readers. Visual word recognition tasks (lexical decision and speed naming task) are suitable experimental research paradigms to test the lexical processing difference of words and pseudowords.

The two contrary theories of pseudoword processing are parallel-distributed (connectionist) and dual-route model. According to the parallel-distributed model, words and pseudowords are processed in the same functional pathways. On the other hand, the dual-route model claims that there are different functional pathways for word and pseudoword processing. There are pieces of evidence for both theories in the existing literature.

Aim of this thesis study was to develop two psycholinguistic tools and provide several hypothesis-testing models to the users by conducting a visual word recognition study. The task of the study was a lexical decision task and its aim was to investigate the effects of several linguistic variables on the behavioral results.

CHAPTER 2

Orthographic and Psycholinguistic Dimension of Words

A word stimulus in a psycholinguistic study should only vary in one dimension (e.g. word versus non-word) and all the other variables should be kept equal across conditions. In other words, all of the linguistic variables beside the linguistic variable to be manipulated should be kept as similar as possible. Otherwise, some unwanted confounding variables could make the behavioral results unreliable. Most of the linguistic variables are orthographic variables some of which are *word frequency, word length, orthographic neighborhood size, orthographic Levenshtein distance 20, orthographic relatedness, bi-gram* and *tri-gram frequency measures*. The other psycholinguistic variables are *age of acquisition, imageability, concreteness* and *cognate facilitation effect*.

2.1 Word-Frequency

Word-frequency is the rate of encountering a word in common language usage; if this rate is high, the word will have a high-frequency value, otherwise it will have a low-frequency value. High-frequency words are preferred to be used more than the low-frequency words. The effect of frequency is prominent on psycholinguistic studies (Forster and Chambers, 1973). After successive repetition of similar research paradigms, a consistent logarithmic function was detected when plotting word-frequency as a function of response time (Carroll and White, 1973; Davis, 2005). Moreover, the impact of the effect is higher for smaller frequency values (e.g. between 10 to 20 per-million) than for higher frequency values (e.g.between 100 to 110 per-million). Forster and Chambers (1973) compared the response times of low-frequency words, high-frequency words and non-words on naming and lexical decision tasks and found that reaction time for high frequency words are significantly lower than both low-frequency words and pseudowords. Nevertheless, the difference between low-frequency words and pseudowords was not significant. This study demonstrated that when a word is recognized, accessing to its lexical meaning is prior to decoding the orthographic and phonetic rules.

2.2 Standardized Word-frequency Values

Standardized word frequency scores should be used in psycholinguistic studies for reliable behavioral results. Kučera and Francis (1967) database is one of the popular word-frequency database for American English which was used in more than 200 published studies (Brysbaert and New, 2009). Nonetheless, Kučera & Francis's database is being criticized by the researchers because its size is based on a corpus of 1.014 million words that the number is small to determine the frequency scores of infrequent words (Burgess et al., 1998). Size of a corpus should be large, as large as 16 million words for a reliable word frequency scores (Keuleers et al., 2010a). The British National Corpus (Burnard, 1995) and the CELEX English Linguistic Database (Baayen et al., 1993) are some of the other standardized word frequency databases. The British National Corpus is an 89.7 million word-frequency list based on 3.261 written texts (Burnard, 1995). CELEX English Linguistic Database (Baayen et al., 1993) is based on a corpora of 17.9 million words (16.6 million from written and 1.3 million from spoken) and both lemma and word-form frequency values are included.

Lemma and *word-form frequencies* are the two different kinds of word frequency counts that are being used in common linguistic literature (Brysbaert and New, 2009). *Lemma frequency* is the total frequency of a word regardless of its type (e.g. adjective, noun and verb and inflected forms). The lemma frequency of the word ”book” is obtained by summing the word category forms of ”book (verb)” and ”book (noun)”, as well as summing the inflected forms of ”books”, ”booked”. *Word-form frequency* is obtained by counting the words having different categories. For example, there are two different word-form frequency values for the word ”damage” in a corpus: ”to damage” as verb and ”damage” as noun.

Psycholinguistic tools such as *Wordgen* (Duyck et al., 2004) and *Wuggy* (Keuleers and Brysbaert, 2010)*pseudoword generators* prefer lemma over word-form frequency values because the length of database is smaller for lemma frequency values and using this results in more efficient search through the database in a lesser time (Duyck et al., 2004). Moreover, the non-inflected forms of words are also preferred on psycholinguistic studies because they are good candidates for a homogeneous stimuli set. However, considering only the lemma frequencies in visual word recognition studies may not be the best solution for valid and reliable results. Kresse et al. (2012) observed the individual effect of the word-form and lemma frequencies on a lexical decision and a word naming task for German words. The impact of lemma and word-form frequency was different on the behavioral data; the response time scores were higher for the lemma frequencies. The authors concluded that both lemma and word-form frequency types should be considered in psycholinguistic studies.

In addition to a corpus's size, its source is also another factor that determines the quality of word-frequency scores (Brysbaert and New, 2009; Keuleers et al., 2010a; Vega et al., 2011). Traditionally, the linguistic researchers were using textbooks and newspapers as a corpus source which are not the true representative of the common word usage (Brysbaert and New, 2009). These traditional sources are usually edited texts in which the language is being corrected multiple times before going to press. Recently, experimenters preferred using some uncorrected text such as the internet websites and movie subtitles as a corpus source and obtained better results (Adelman et al., 2006; Keuleers et al., 2010a). Adelman et al. (2006) counted the appearance of a word in different number of small texts and compared the reliability of these word-frequency scores with the scores obtained by counting the repetition of a word in a large text on a visual word recognition and a lexical decision task. Their study showed that considering the *contextual diversity* on word frequency counts predicts lexical decision and word naming scores more than the traditional way does. In a visual word recognition study, the contextual diversity of a word was compared with the classical word-frequency measures using movie subtitles as the corpus for Dutch words and named this corpus as SUBTLEX (Keuleers et al., 2010a). The results show that, SUBTLEX explains more variance (up to 10%) on lexical decision and word naming data than traditional word frequency counts (extra 1%-3% more variance).

2.3 Word Length

Word length is another important linguistic variable to be considered on a psycholinguistic study stimuli set. There is a consistent parallel pattern between the length of a word and reaction time on a lexical decision and visual naming task (Chumbley and Balota, 1984; Davis, 2005; Forster and Chambers, 1973). Forster and Chambers (1973) compared response time scores of words with varying length (four to six letters) and varying syllable number (one versus two syllables) on a lexical decision and visual naming task and observed a significant effect of word-length on the reaction time, but not any effect of syllable. New et al. (2006) examined this length effect more thoroughly on a lexical decision task and the results pointed out a consistent pattern regardless of variation in word frequency, number of syllables and orthographic neighborhood size. Words with letter length of 3 to 5 facilitates its detection, the effect is null for words with letter length of 5 to 8, and the detection is inhibited with letter length of 8 to 13.

Because there is a prominent effect of letter length on the visual word recognition data, it is advisable for psycholinguistic experimenters to keep the length equal in their word stimuli set. This would prevent the data from a confounding factor and would increase the reliability of the behavioral scores. For example, Chumbley and Balota (1984) compared words and pseudowords on a lexical decision task in one of their experiments and the reaction times were smaller for the pseudowords. Later than they realized that their verbal stimuli set was not totally homogeneous the mean length of the non-words were one letter shorter than the words. This situation crated an unwanted artifact on the behavioral data and resulted on average of 13 milliseconds shorter reaction times for the non-words.

2.4 Orthographic Neighborhood Size

Orthographic neighborhood (ON) size (also called the Coltheart's N metric) refers to the number of words that can be obtained by substituting a single letter in a word list (Coltheart et al., 1977). For example, the word ”song” has six orthographic neighborhoods which are ”long”, ”sung”, ”pong”, ”gong”, ”sing”, ”tong” in CELEX English database (Duyck et al., 2004). ON size is another important linguistic variable to be considered for visual word recognition studies. ON size of a non-word is also a strong indication of how that non-word is orthographically distant from a word (Duyck et al., 2004). For example, the English non-word ”durt” has an ON size above ”0” and it sounds more like a word than the letter string ”hzva”.

The behavioral effect of ON is present in various reading performance tasks such as lexical decision, word-naming, perceptual identification and semantic categorization tasks (Andrews, 1997). The existing literature on ON size points out a *facilitation effect* on the reaction time in a lexical decision task which means that the higher ON a word has, the shorter its response time is (Andrews, 1997). However, the situation changes when considering the frequency values of words that the two variables interact and causes an extra facilitation effect for low-frequency words, and an inhibition effect for high-frequency words (Yarkoni et al., 2008). Detection of a high-frequency and high ON size word causes an inhibition effect because the orthographic neighbor words compete with the target word while accessing to its semantic meaning.

Baayen et al. (1993) inquired the relationship between ON size, word-frequency and word length and analyzed these relationships in English and Dutch language using the CELEX database. Results have shown that, the length of a word has a negative correlation with both word-frequency (WF) and ON. To be more specific, four-letter-words had an average of 100.3 WF score and 7.2 ON size, five-letter-words on had an average of 34.2 WF score and 2.4 ON size and six-letter-words had an average of 16.5 WF scores and 1.1 ON size. It is not very surprising for the ON value to be higher for four-letter-words because shorter length sizes increase the chance of orthographic letter combination (Baayen et al., 1993).

2.5 Orthographic Relatedness

Words having similar orthographic representation are also orthographically related to each other. For example, the words ”contract” and ”contrast” have have one different letter from each other hence they are orthographically related (Duyck et al., 2004). The behavioral effect of orthographically related items on visual word recognition study is contradictory, some studies indicate a facilitation effect, other studies indicate an inhibition effect on the behavioral scores. However, the evidence is mostly on the side of a facilitation (Andrews, 1997). The existing literature on orthographic relatedness brings up two different types of orthographic similarity: subset/ superset similarity and transposed letter similarity (Bowers et al., 2005; Davis, 2005).

Subset/Superset similarity: This similarity is a special case of orthographic relatedness and occurs when there is an embedded word in another word (Bowers et al., 2005; Davis, 2005). For example, the word ”utter” is the subset of the word ”butter” and the word ”butter” is the superset of ”utter”. Studies show that the presence of an embedded word in a visual word stimuli decreases the response time and increases the accuracy scores in visual word recognition tasks (Bowers et al., 2005). Thus, psycholinguistic experimenters should be careful on their stimuli set about the unsystematic presence of an embedded word, such words can decrease the words' response duration irrelevant of the study's research hypothesis.

Transposed letter (TL) similarity: TL similarity is another special case of orthographic relatedness. This similarity occurs when one word can be transformed to the other by the replacement of two letters. There are two different types of transposed letter similarity: adjacent and non-adjacent letter similarity (Perea et al., 2008). As its name suggests, the adjacent similarity occurs when two adjacent letters change place with each other (e.g. the word ”cosMEtic” and the non-word ”cosEMtic”). Non-adjacent similarity is the case when there are one or more intervening words between two transposed letters. For example, the word ”sentence” has non-adjacent transposed letter similarity with one intervening letter in the non-word (e.g. cosTeMic) and two intervening letters in the non-word ”coImetSc”. Perea et al. (2008) compared the response times between adjacent and non-adjacent TL similarity words for Spanish words in a priming lexical decision task. The results have shown that response times and accuracy scores were lowest for adjacent letters and gradually increased among one and two intervening words. Thus, the effect was still present in two intervening letter non-adjacent similarity with (the word ”chocolate” and pseudoword ”choaolcte”).

In addition to adjacency, whether the transposed letter represents a vowel or a consonant may also influence the behavioral scores. Studies conducted on English and Spanish words demonstrate that, the behavioral effect was only present when the replacement is only present for the consonant letters (C-C transposition) but not for vowel letters (V-V transposition) on a priming and normal lexical decision task (Lupker et al., 2008; Perea and Lupker, 2004). An example for the C-C transposition is the similarity between the word ”casino” and the pseudoword ”caniso”, and V-V transposition is the similarity between the word ”animal” and ”anamil”. Thus, there are different kinds of TL similarity, and psycholinguistic experimenters should check for the unsystematic presence of each type of similarity in their stimuli sets for reliable behavioral results.

2.6 Orthographic Levenshtein Distance 20

Orthographic Levenshtein Distance 20 (OLD20) is a recently developed measure for orthographic similarity as an alternative to ON size (see Section 2.4). The parameter is derived from one of the standardized computer science string metrics called Levenshtein distance (LD). LD is an algorithm (explained in Section 3.3) used to assess the distance between two letter strings[1]. It is a continuous metric that indicates the minimum number of insertion, deletion or substitution operation required for turning one string to another (Levenshtein, 1966). For

[1] http://en.wikipedia.org/wiki/Levenshtein_distance

example, passing from the string ”foul” to ”fuel” requires two operations; one substitution and one addition operation. OLD20 is the average of 20 closest words on a lexical database in the unit of LD. Table 2.1 shows the OLD20 values for words ”condition” and ”pistacho”. The word ”condition” has a low OLD20 value (2.4) and its neighbor words are closely spaced. On the other hand, the word ”pistacho” has a high OLD20 value (4.3) and its neighbor words are distantly spaced.

Table 2.1: OLD20 values for the words ”condition” and ”pistacho” and their 20 LD closest words (Adapted from Yarkoni et al. (2008))

CONDITION		PISTACHIO	
Neighbor Words	Levenshtein Distance	Neighbor Words	Levenshtein Distance
conditions	1	distraction	4
coalition	2	hibachi	4
cognition	2	mustache	4
conditional	2	mustached	4
conditioned	2	mustaches	4
conditioner	2	pigtail	4
conduction	2	pistil	4
contrition	2	pitch	4
conviction	2	pitched	4
recondition	2	pitcher	4
rendition	2	pitches	4
addition	3	pitching	4
audition	3	psychic	4
collation	3	psycho	4
collision	3	abstain	5
commotion	3	abstraction	5
conception	3	antacid	5
concoction	3	attach	5
concretion	3	attache	5
conditioners	3	attached	5
OLD20: 2.4		OLD20: 4.3	

OLD20 was developed by Yarkoni and his colleagues in 2008 as a continuous parameter for the assessment of orthographic similarity for psycholinguistic studies. The number 20 was chosen because it is the number that maximally correlates with response times in a lexical decision task (Yarkoni et al., 2008). The authors claim that OLD20 is a better measure for orthographic similarity than the ON because it is more comprehensive in terms of detecting orthographic relatedness between the two words. ON is only concerned with substitution of a single letter and cannot detect the similarity between orthographically related words such as ”carny” and

”candy” or ”word” and ”world”. The correlation of OLD20 scores with behavioral measures (response times and accuracy scores) are better than ON in regression analysis that OLD20 has 1%-2% more variance on behavioral scores (Yarkoni et al., 2008). Additionally, word length is positively correlated with OLD20 but not with ON (Yarkoni et al., 2008).

2.7 Bi-gram and Tri-gram Frequencies

Decomposing a string of tokens (e.g. words) into sequences of ”n” number of neighboring elements are called n-grams (Manning et al., 1999). It is very common of this ”n” to be two and are called bi-grams, if ”n” is three they are called tri-grams. For example, the Turkish word ”araba” (car in English) consist of four bi-grams [ar, ra, ab, ba] and three tri-grams [ara, rab, aba]. There are two different types of bi-gram/tri-gram frequency measures: position dependent and position independent (Davis, 2005). Type frequency is a position dependent frequency measurement, the location of the bi-gram in a word is considered while counting the number of bi-grams in the lexical list. For example the word ”edge” has four four-letter words that has the bi-gram ”ed” in the first position according to CELEX database [eddy, edge, edgy, edit] (Davis, 2005). Token frequency is the sum of word-frequency values of each word in type frequency list. Average bi-gram/tri-gram token frequency is token frequency value divided by the number of words in type frequency list. Summed type bi-gram/tri-gram frequency is a position independent frequency measurement obtained by summing the type frequencies. Summed type bi-gram frequency is an indication of word-likeliness for a pseudoword, the higher this value, the more ”wordlikely” a pseudoword is (Duyck et al., 2004). Summed log bi-gram frequency is the sum of the logarithms of its token frequency value for each bi-gram/tri-gram.

There is not a consistent effect of bi-gram frequency on visual word recognition tasks however, some of the studies observed a facilitation effect when tested on a lexical decision task (Andrews, 1997). Moreover, bi-gram frequency scores of a word seem to be positive correlated with its pronounceability speed and accuracy scores (Massaro et al., 1979).

2.8 Other Psycholinguistic Variables

There are also some behavioral linguistic variables that has an effect on the behavioral results. *Age of acquisition* is one of these variables and psycholinguistic studies point out an interaction of acquisition age and frequency values on the behavioral scores (Bird et al., 2001). Earlier acquired words have higher frequency scores and result in smaller response times when tested on lexical decision and speeded naming tasks (Carroll and White, 1973).

Imageability and *concreteness* are other variables which have influence on the behavioral scores. Imageability of a word refers to the extent that it arises a sensory mental image; concreteness refers to the extent it can be represented with the senses (Paivio et al., 1968). Imageability and concreteness scores are available for some of the words in the MRC database (Coltheart, 1981). MRC (Medical Research Council) is a 98,538-word psycholinguistic lexical database of American English. The imageability and concreteness scores were derived from the Paivio et al. (1968) study. According to the MRC database, the word ”against” has a low imageability score; the word ”blonde” has a high imageability score; the word ”rule” has a low concreteness score; the word ”ankle” has a high concreteness score. Visual word recognition data is positively correlated with a word’s imageability level but not with its concreteness level (Marcel and Patterson, 1978; Paivio et al., 1968).

The *cognate facilitation effect* is another linguistic variable that needs to be considered while preparing a stimuli set. Cognate words share the same origin with some other word(s) which can be in another language (Peeters et al., 2013). For example, the word ”fruit” in English and the word ”fruto” in Spanish are cognates of each other because both of the words were derived from the same latin root. In a lexical decision study, response times to cognate words were smaller and more accurate than non-cognate words for Dutch-English bilingual subjects (Dijkstra et al., 2010). Moreover, the strength of ”cognate facilitation effect” is parallel to words that have the most similar orthographic representation (Dijkstra et al., 2010). Thus, a psycholinguistic experimenter should be aware of the ”cognate” words in their word stimuli set for more reliable and valid results.

2.9 Summary

In psycholinguistic studies, the stimuli set should be homogeneous for a reliable behavioral results. A visual word data-set varies on several aspects among conditions and such variations should be carefully controlled by the psycholinguistic experimenters. These variations can be on the dimension of orthographic as well as behavioral. A word is conjunction of several letter strings based on some rules. Thus, a word is orthographically consists of several number of letters (word-length), and has a frequency value (word-frequency), some words can be obtained by substituting a single letter from it (orthographic neighborhood), has a neighborhood set and this can be dense or sparse (the OLD20 value), can be orthographically very similar with other words (orthographic relatedness), can be decomposed into two or more string of letters as the subsets (n-grams).

Behaviorally, each word has a semantic mental representation and several factors may have influence this retrieval process. Each word was acquired in a time and this could be either late or early (age of acquisition). A word also has an *imageability level*, which is higher for countable noun words such as ”apple” and a *concreteness level*, which is higher for words that refer to solid objects. Moreover, words can be a *cognate* of another word and share same or similar orthographic representations.

CHAPTER 3

The KelimetriK Software: A psycholinguistic tool for Turkish Visual Word Recognition Studies

3.1 Introduction

Selection of appropriate words for a fully controlled word stimuli set is an essential component for preparing an appropriate visual word recognition task (Bowers et al., 2005; Perea and Pollatsek, 1998). For example, if the stimuli set of a visual word recognition study is full of high frequency words, this may create a bias on the behavioral scores, hence would lead to incorrect inferences about the hypothesis. Thus, the experimenters who are intended to work with any kind of verbal stimuli should consider psycholignustic variables to prepare valid behavioral assessments.

KelimetriK is a query-based software designed to calculate several orthographic statistics. As shown in Figure 3.1, the software has a user-friendly interface developed to be a helpful source to psycholinguistic experimenters. The software provides information about the linguistic variables of word-frequency, neighborhood size, orthographic similarity and relatedness. KelimetriK's counterparts in other languages are N-watch in ENglish (Davis, 2005) and BuscaPalabras in Spanish (Davis and Perea, 2005).

3.2 The Lexical Variables in KelimetriK

The lexical variables provided in the KelimetriK are word-frequency, bi-gram and tri-gram frequency, orthographic neighborhood size (ON), orthographic Levenshtein distance 20 (OLD20), transposed letter and subset/superset similarity.

Word frequency: is a value that describes of how many times a word occurred in a given text. Research shows that there is a consistent logarithmic relationship between a word's frequency score and its reaction time; the impact of effect is the highest on low-frequency words and gets smaller for the words having high-frequency scores (Davis, 2005).

Bi-gram and Tri-gram Frequencies: Bi-grams and tri-grams are obtained by decomposing a

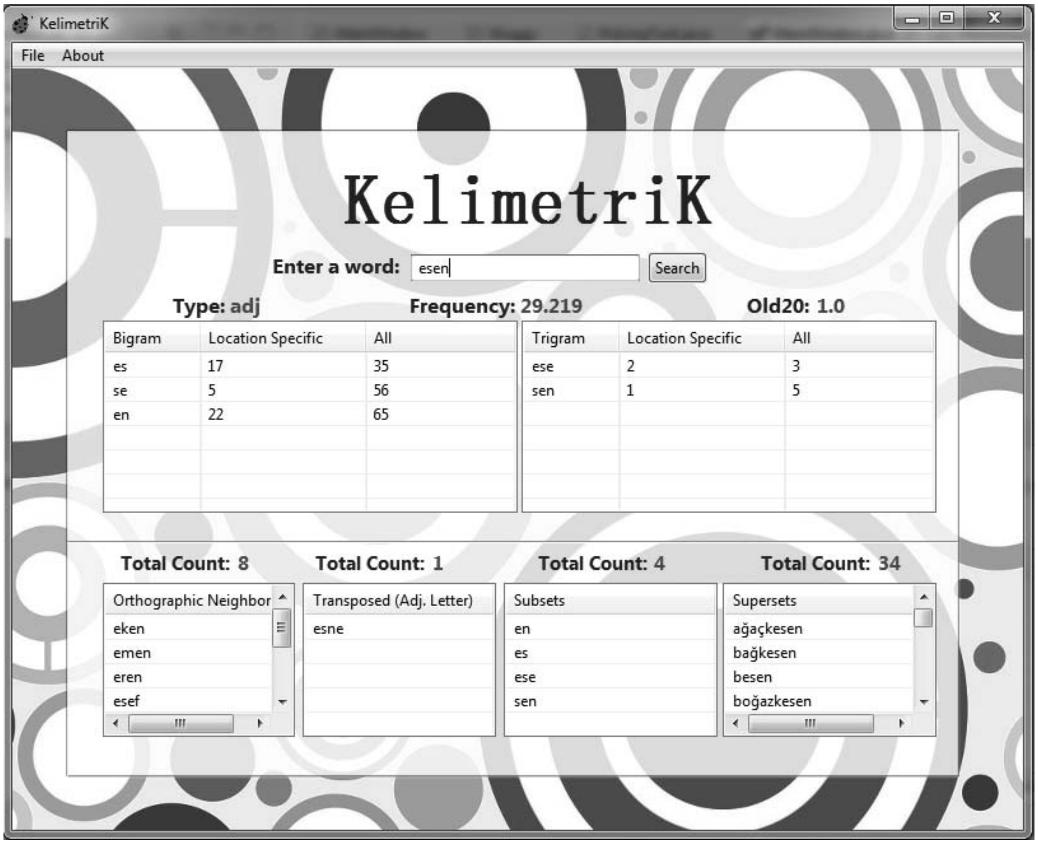

Figure 3.1: *The interface of the KelimetriK software.*

word (string of tokens) into sequences of two and three number of neighboring elements (Manning et al., 1999). For example, the Turkish word ”kule” (tower in English) can be decomposed into three different bi-gram sets (”ku”, ”ul”, ”le”). Bi-gram (or tri-gram) frequencies are obtained by counting how many words (four in this case) words starts with first bi-gram set (e.g. ”ku”), how many words have second bi-gram set in the middle (e.g. ”ul”), and how many words ends with the last bi-gram set (”le”) in a given lexical word database. The tri-grams for the word ”kule” are ”kul” and ”ule” respectively. Average bi-gram (or tri-gram frequency) is obtained by dividing a word’s bi-gram frequency by the total number of bi-grams of that word.

Orthographic neighborhood (ON) size: Refers to the number of words that can be obtained from a lexical database by substituting a single letter of a word (Coltheart et al., 1977). For example, the ON size of the word ”kule” is 9 (”şule”, ”kula”, ”kulp”, ”fule”, ”kale”, ”köle”, ”kele”, ”kile” and ”kuşe”) if searched on KelimetriK. There is evidence that the ON size influences word processing performances in the visual word recognition tasks of lexical decision, naming, perceptual identification, and semantic categorization (Perea and Pollatsek, 1998).

Orthographic Levenshtein distance 20 (OLD20): Refers to the the average of the 20 closest words in the unit of Levenshtein distance (Yarkoni et al., 2008). Levenshtein distance between the two strings of letters is obtained by counting the minimum number of operations (substitution, deletion or insertion) required while passing from one letter string to the other (Levenshtein, 1966). Behavioral studies in English show that, OLD20 is negatively correlated with orthographic neighborhood size (r=-0.561) and positively correlated with word-length (r=0.868) (Yarkoni et al., 2008). Moreover, OLD20 explains more variance on visual word recognition scores than ON and word length (Yarkoni et al., 2008).

Orthographic Similarity: Two words are orthographically similar if they are neighbors of each other like the words *alın* ("forehead" in English) and *alan* ("area" in English). Adjacent transposed letter (ATL) and subset/superset similarities are the two most common orthographic similarities in the existing literature (Davis, 2005). ATL similarity is the case when two letters differ from each other based on a single pair of adjacent letters such as the Turkish words *esen* (blustery) and *esne* (yawn) (See Section2.5 for other types of transposed letter similarity). Studies have shown that ATL similarity may facilitate word detection performances on naming and lexical decision task (Andrews, 1997). Subset/Superset similarity occurs when there is an embedded word in a given input word such as *süt* (subset: "milk" in Turkish) *sütun* (superset: "pillar" in Turkish). Presence of a subset in a stimuli word may influence a subject's reading performance, which could result in a confounding factor on the behavioral results (Bowers et al., 2005).

3.3 The algorithm of the KelimetriK

KelimetriK uses the Bozşahin et al. (2012) Turkish unique stem list to calculate the orthographic scores (See Section 4.4 for details). Output of these calculations depicted on the interface with an organized layout. The user interface and the output format can be seen in Figure 3.1, the outputs are explained in the text below.

Word Type: It is obtained from the corpus.

Word Frequency: It is obtained from the corpus.

Old20: Calculates the average of the most closest 20 words in the unit of Levenshtein distance. As explained in Section 3.3, the Levenshtein distance between the two strings of letters is calculated by counting the number of minimum required operations (substitution, deletion or insertion) while passing from one string into to the other (Levenshtein, 1966). Its mathematical formula can be seen in Formula 3.1

$$
lev_{x,y}(i,j) = \begin{cases} max(i,j) & if\ min(i,j) = 0 \\ min \begin{cases} lev_{x,y}(i-1,j)+1 \\ lev_{x,y}(i,j-1)+1 \\ lev_{x,y}(i-1,j-1)+[x[i] \neq y[i]] \end{cases} & otherwise \end{cases} \tag{3.1}
$$

In the mathematical formula, the first element in the minimum clause corresponds to character deletion, insertion and the match or mismatch of characters respectively.

Bi-gram Counts: For a given query word, bi-gram frequency is calculated for all the words in the corpus having the letter same length. There are two types of bi-gram frequency counts: position dependent and position independent. The position of bi-grams should also considered match in the former case, and not considered in the latter case.

Tri-gram Counts: The calculation method is the same with the bi-gram frequency calculation.

Orthographic Neighborhood Count: For a given query word, the number of differences in the strings' characters are calculated for the words in the corpus having same length. If the total character difference is one, then it is added into the ON word-list.

Counting Adjacently Transposed Words: All the other words that have the same length with the query are iterated one-by-one. For each matching word, all the two consecutive characters (letters of the word) are swapped one-by-one and compared to the query. If the comparison of the two match, then the word is added into the transposed word-list.

Counting Subset Words: A regular expression is used for the algorithm that checks whether a word is a subset of a queried word. Assume that the queried word is <query> and the word that the algorithm is searching for is <word>. If <query> matches with the regular expression ".*<word>.*", then it is added into the subset word list. This regular expression is a formal way of saying that if you can construct a <query> by prepending and appending characters to a <word>, then the <word> is in the subset of <query>.

Superset Words Count: Finding the superset of a <query> is similar to finding its subset. The only difference is that <query> and <word> are switched position within the regular expression this time. If <word> matches to the regular expression ".*<query>.*", it means that it is a superset of the query and included into the superset word list.

3.4 Sample Usage of the KelimetriK

It is assumed that four high frequency words are needed for a hypothetical visual word recognition study and the experimenter has already prepared a candidate list of 6 words to query for their lexical properties on KelimetriK. The experimenter used KelimetriK's word list while preparing this candidate word-list which contains information about the word-frequency, ON and OLD20 scores. The word's in the stem list are obtained from the Turkish stem list of

Bozşahin et al. (2012) study (See Section 4.4 for details). Figure 3.2 describes how to access to the KelimetriK's corpus from the software's interface. If the "File" button positioned on left above is pressed, it is in the section of "Open Word List". The list is sorted according to alphabetical order of words.

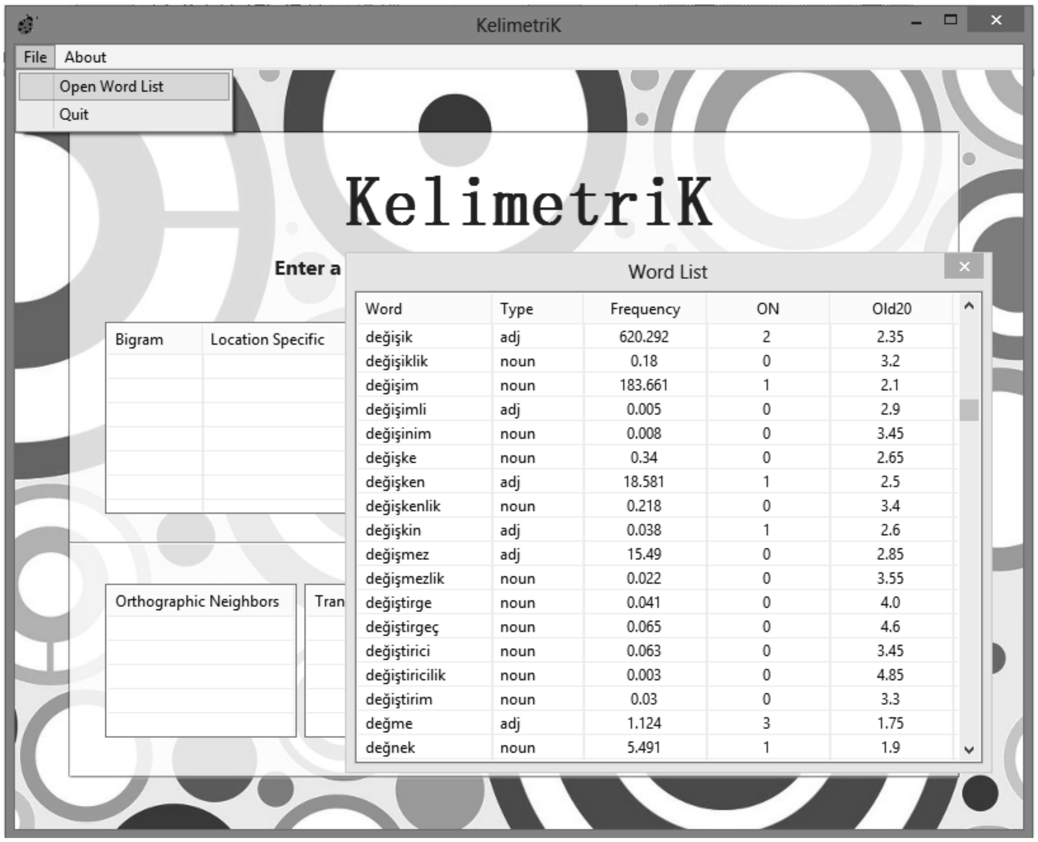

Figure 3.2: *Screenshot of the KelimetriK word list.*

Table 3.1 shows a possible candidate word list, their English translation and some of the orthographic scores. The scores were obtained by querying each word one at a time on KelimetriK. Each word's orthographic score should be examined carefully before including it to the stimuli set. For example, the word *uyum* ("coherence" in English) should not be used on the stimuli set because its frequency score is too low when compared to the other words. The word *ırak* ("away" in English) is also not a very good candidate to be in the stimuli set because it has a very little ON scores compared to the other words. Thus, the response duration of this word might be longer when compared with the other words regardless of an effect of frequency (Andrews, 1997). Moreover, the word *alan* ("area" in English) has much more supersets than the other words which might cause an uncontrolled facilitation effect on response times (Bowers et al., 2005).

Table 3.1: A possible Turkish candidate word-list, their English translations and some of their orthographic statistic scores obtained using the KelimetriK software.

Words	Frequency (Per-Million)	ON	OLD20	# of Supersets
bura (”here”)	841.085	16	1.0	13
ırak (”away”)	855.922	4	1.45	32
uyum (”harmony”)	245.885	7	1.45	11
sıra (”order”)	1180.975	14	1.0	21
fark (”difference”)	1175.313	14	1.05	5
alan (”area”)	997.853	19	1.0	74

3.5 Limitations of the KelimetriK Software

There are some limitations of KelimetriK that could restrict researcher’s options while preparing their stimuli set. The first limitation is on searching for the words that have the same orthographic representations but has different meanings. One example of these kinds of words in Turkish is *gül* which means ”rose” when used as a noun and ”to smile” when used as a verb. However, this word would be described as a verb if queried on KelimetriK and its noun version would be disregarded. This kind of limitation is not caused by the software’s algorithm but because of the word list. KelimetriK uses a unique stem list (Bozşahin et al., 2012) that provides only one category per-word. Thus, it is not possible to overcome this limitation unless using another word list as a corpus. However, no such a replacement is currently possible because an alternative Turkish lexical list does not also contains the word-frequency scores.

The second limitation of the KelimetriK is that multiple words cannot be searched at the same time. However, the orthographic scores of word-frequency, ON and OLD20 scores are also available on the KelimetriK’s corpus. Accessing to this word list is available from the interface (See Figure 3.2 to access to the KelimetriK’s word-list).

3.6 Summary

KelimetriK is a software program designed provide information about several orthographic scores of a queried word. These orthographic scores are word-frequency, bi-gram and tri-gram frequency, orthographic neighborhood size (ON), orthographic Levenshtein distance 20 (OLD20), transposed letter and subset/superset similarity. The orthographic calculations of the software are based on a Turkish stem-list and the software only gives output to words having the same length with the queried word. Querying for words that have the same orthographic representations but has different meanings is limited in KelimetriK. It is also not possible to query more than one word at the same time.

CHAPTER 4

The Wuggy Software: A Multilingual Pseudoword Generator

4.1 Introduction

Wuggy is a pseudoword generator software program developed by Keuleers and Brysbaert (2010). Wuggy's algorithm uses a novel method for pseudoword generations that the non-words are generated in a short duration from one or more template words which are defined by the users. Wuggy software was developed with the Phyton programming language (Van Rossum and Drake Jr, 1995) and the source code is available on the software's website.[1] The algorithm is flexible in terms of getting adapted to other languages if a syllabified word list of the language, a language's orthographic rules for defining each word's nucleus and word-frequency scores are provided. The software is available in Basque, Dutch, English, French, German, Serbian, Spanish and recently for Turkish.

Figure 4.1 shows the interface of Wuggy for the generated pseudo-words and their orthographic statistics for the template words "swim" and "bend". The available orthographic statistics are OLD20, ON size and deviation statistics (See Chapter 2 for detailed information of these orthographic statistics).

The underlying calculations behind pseudoword generation is explained in detail in Section 4.3 and a brief summary is also provided below. The syllabified version of the template words are provided to the software from the developers in as a corpus and the algorithm decomposes each of these syllables into the sub-syllabic elements of onset, nucleus and coda. A bi-gram frequency chain is created among these sub-syllabic elements. The bi-grams of two consequent sub-syllabic elements (e.g onset and nucleus; nucleus and coda) are calculated by summing the frequency of the words which have the same transitions. The non-words are generated by substituting syllabic elements of the template words with elements that have similar bi-gram frequencies. These elements are detected using the "concentric search" feature of the algorithm and the bi-gram frequency ranges are increased gradually if words with specified frequency ranges are not found.

[1] `http://crr.ugent.be/programs-data/wuggy`

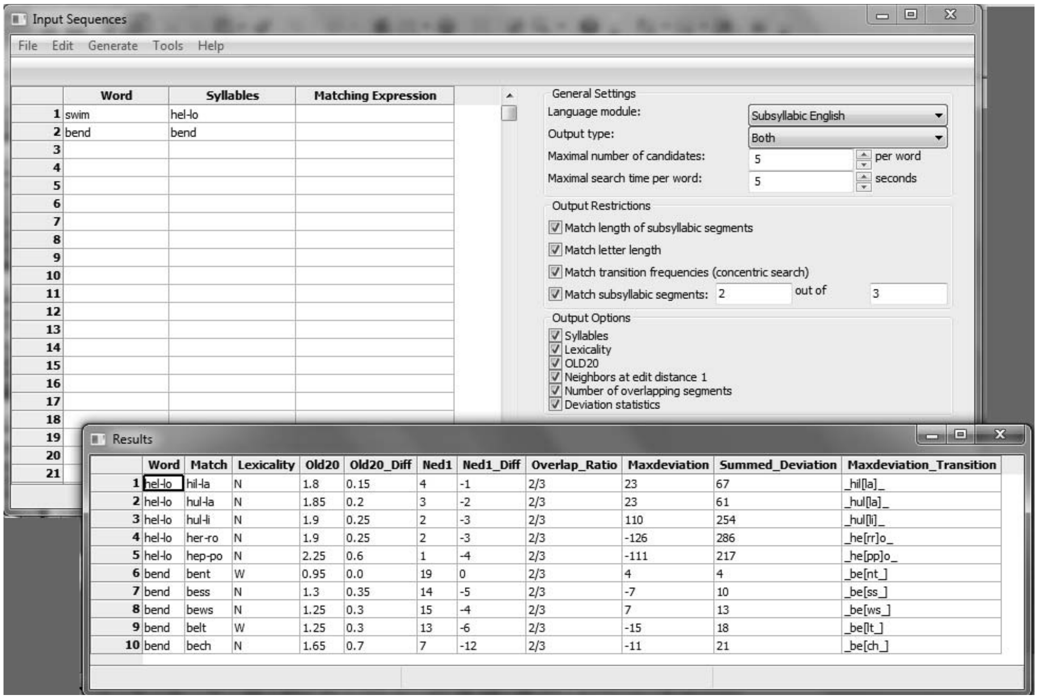

Figure 4.1: *The interface of the Wuggy Software.*

4.2 Methods for Obtaining Pseudowords

Manual methods: It is a common habit of researchers to obtain pseudowords by substituting real words' one or two letters with the other letters legal to the language's rules (Balota et al., 2007; Forster and Chambers, 1973; Keuleers and Brysbaert, 2010). One way for doing this is changing some of the letters of a word and then applying some constrictions specific to that particular language's common string conjunction patterns. For example, it is regular for the English words to start with ”pi-” and ”ma-” and end with ”-rk” and ”-lp” (Keuleers and Brysbaert, 2010). Rearranging letters of real words is another manual method for creating pseudowords. This method was used in Forster and Chambers (1973) study to obtain pronounceable English pseudowords such as ”vero” (derived from *over*) and ”tino” (derived from *into*).

Obtaining pseudowords by using the manual methods may not be an optimal solution for reliable behavioral results. Researchers implicit knowledge about the research's hypothesis may create a bias over word selections such as having individualistic preferences on modification of certain letters (Balota et al., 2007). Forster (2000) conducted a lexical decision study in order to investigate whether psycholinguistic experimenters have preferences for selecting items that may create bias over the results. The task was to select the word in the stimuli set which would result in smallest response time in a pair of frequently matched word set. As a result, the subjects' accuracy performances were above chance and the performance levels of the task increased as a function of the subject's research experience on the psycholinguistic field. The results supported the notion that the experimenter subjects have an intuitive knowledge toward the verbal stimuli that would support their research hypothesis. Thus, the manual methods should be avoided to prevent the behavioral study from a possible experimenter bias.

There are also some other methods alternative to the manual methods such as using one of the standardized databases or using one of the pseudoword generator programs. The second option can be a better method for pseudoword generation because psycholinguistic tools are flexible; the pseudowords can be generated according to the specific requirements of a behavioral study. Moreover, these software programs allows the users to filter out the options to obtain pseudowords on desired linguistic variables (See Chapter 2 for detailed explanation of the linguistic variables).

Standardized databases: There are many different standardized databases available in the existing literature. The English Lexicon Project (ELP) is one of the popular psycholinguistic database that contains the behavioral scores of reponse time and accuracy scores for 40,481 words and 40,481 non-words obtained from a large set of subjects (1,260 participants). The British Lexicon Project (BLP), another standardized database in English, provides psycholinguistic information for 14,365 words and 14,365 non-words (Keuleers et al., 2012).

ARC is also a standardized non-word database that stores 358,534 legal monosyllabic non-words, 48,534 pseudo-homophones and 310,000 non-pseudohomophonic non-words (Rastle et al., 2002). The search engine of ARC allows the users to specify range of the orthographic variables (word-length, orthographic neighborhoods and word frequency). However, all of the pseudowords in ARC database is monosyllabic which is inefficient for the experimenters who are going to use polysyllabic words in their stimuli set.

Pseudoword generation software programs: These software programs can generate a number of words and legal non-words in a minute and they are flexible in terms of defining the range of orthographic properties (e.g. word-frequency). WordGen is one of these software programs that the pseudowords are generated after application of a serial search process (Duyck et al., 2004). The serial search process is started with selecting an entry in the word list database.

The algorithm of Wordgen starts a serial search by selecting an entry in the word list database on its search space. Then, the pseudowords are generated and confirmed that they are pseudowords by a lexicality check process. This is checked by comparing the pseudoword with all the words in the lexical database and confirms that it does not match with any other word. If the word is proved to be a pseudoword, then the algorithm checks for whether it fits with the user-defined criteria within the limits of specified search time. The algorithm can pursue two different strategies while generating the pseudowords which are stringing letters randomly or substituting a single letter from one of the words in the lexical database. The users can constrain the pseudoword generation according to its letter length, neighborhood size, word-frequency, summed type bi-gram frequency, maximum legal bi-gram/tri-gram frequency. The pseudowords can also be specified with customizing the letter patterns such as "*b***".

As mentioned above, WordGen generates pseudowords with a serial search method and this causes a limitation while generating lengthy words because less pseudowords are generated within a specified time if they are lengthy. In addition, the search space is not so flexible to generate non-words at once with too many specified criteria. For example, if the users want to obtain pseudowords with the bi-gram frequency ranges of 3000 and 4000 and the ON size range of 7 and 9, they have to run the software program more than once for each restriction.

McWord is another available source for English visual word recognition studies (Medler and Binder, 2005). The software is not only a pseudoword generator, but also an orthographic lexical database. The database contains information of lexical (neighborhood statistics and word-frequency) and sub-lexical properties (letter combinations) of words. Moreover, the software permits the users to generate pseudowords with specified orthographic criteria.

LINGUA (the Language-Independent Neighborhood Generator of the University of Alberta) is a pseudoword generator software which provides information about words' orthographic scores such as ON, word and bi-gram frequency (Westbury et al., 2007). Unlike the other pseudoword generators, LINGUA is a language independent toolbox which can work on any user defined corpus. The software has an additional program for defining a corpus which requires a large text to generate non-words.

Wuggy is also one the available software programs to generate monosyllabic and polysyllabic pseudowords in a short duration (Keuleers et al., 2010a). Wuggy's developers claim for the software to be the most efficient pseudoword generator in the existing literature. Wuggy generates pseudowords by combining the words' legal sub-syllabic elements in its corpus while considering the transition-frequencies of template words which are defined by the users (See section 4.3 for a detailed explanation).

4.3 The Algorithm of Wuggy

The traditional way to generate pseudowords is based on combining sub-syllabic elements that are legal in a language. There can be billions of pseudowords generated by using this method. However, finding some particular pseudowords with specified criteria becomes inefficient due to the large amount of pseudowords. Wuggy overcomes this problem by proposing a new method to narrow down the search space.

Before the algorithm applies its filtering methods, a dictionary of bi-gram chains are constructed for words of a target language. First, each word in the corpus is syllabified and the algorithm uses the syllabified version of the words. The sub-syllabic segments of each syllabified word are decomposed by using onset-nucleus-coda (ONC) pattern of the language. Then, a tuple is constructed consisting of four components of the position of the segment in the word, the number of syllables in the word from which it is originated, the length of the subsyllabic segment, the letters of the sub-syllabic segment. The bi-gram chains are constructed by linking these tuples to each other. Each link can only be constructed once to the bi-gram chain dictionary. If the same link is encountered again, its frequency value is updated (either added one or its frequency values in the corpus see Algorithm 1). For example, if the algorithm constructs bi-gram chains for the Turkish word *kitap* (”book” in English), it obtains the syllabified version of the word as a first step which is ”ki-tap”. Next, the sub-syllabic segments are constructed as ”onset: k, nucleus: i, coda: _” and ”onset: t, nucleus: a, coda: p”. Then, a tuple for each sub-syllabic segment is constructed. In this example the tuple chain is demonstrated for two consecutive segments of ”k” and ”i”. The tuple for sub-syllabic segment ’k’ is (2, 1,’k’, 1) and for sub-syllabic segment ’i’ is (2, 1, ’i’, 2). A chain between these two segment is created with a frequency value 1 (assumed as it is the first occurrence of this chain) which is [(1,2,1,’k’), (2, 2, 1, ’i’)]: 1.

Algorithm 1 Constructing bi-gram chains.

1. **Inputs:**
   1) The lexicon L of the language.
   2) An onset-nucleus-coda pattern of the language.

2. **Output:** The bi-gram chains for sub-syllabic (onset-nucleus-coda) segments and their transition frequencies.

3. The bi-gram chains and transition frequencies for the lexicon L are determined by the following procedure:

   (a) Each word W in the lexicon L is syllabified and W_s is generated.

   (b) For each W_s sub-syllabic segments S are constructed using ONC pattern.

   (c) For each S_i in S, a tuple T_i is generated with the following information:
   T_i = (# of syllabuses in word W_s, length of the segment S_i, the segment S_i itself)

   (d) For each tuple T_i a bi-gram chain $bichain_{T_i - T_{i+1}}$ is constructed using T_{i+1}. If the bi-gram chain exists before, it its frequency is updated either by adding one or its frequency value in the corpus.
   $bichain_{T_i - T_{i+1}} = (T_i, T_{i+1}, freq_{T_i - T_{i+1}})$

These constructed bi-grams chains can be used in two different ways to restrict candidate pseudowords. The first way is the *segment length criterion*. A reference word such as *bridge* can be divided to its sub-syllabic segments as *br-i-dge* and the length of these segments are 2-1-3 respectively. If this restriction is applied, the generated pseudowords will have exactly the same sub-syllabic structure as the reference word hence the search space will be narrowed down. The second way is by *concentric search*, which restricts the pseudowords according to bi-gram chain transition frequency values for sub-syllabic segments. For example, if the bi-gram chains [_,*br*], [*br,i*], [*i,dge*], and [*dge,*_] have frequencies of 125, 25, 4, and 29 respectively, the pseudowords that do not fall within the given deviation of frequency would be filtered out. The deviation from the original chain frequency starts with ± 2 and increases as the powers of two (e.g., ± 4, ± 8, ± 16, ...) until a suitable bi-gram chain within the range of deviation is found.

Wuggy software is ready to generate pseudowords when the bi-gram chains for the selected language are constructed. The user can enter the reference words or the bi-gram chain is constructed for the entire words in the lexical list. Moreover, the user can also select several output restrictions to obtain pseudowords with desired features.

These restrictions are;

- **Sub-syllabic segment length:** This option limits the output word to have the same sub-syllabic structure as the reference word.
- **Letter length:** This option generates pseudowords with the same number of letters as the reference word.
- **Transition frequencies (concentric search):** This option applies the concentric search method as described above.
- **Sub-syllabic segments:** This options ensures that the generated pseudowords and the reference word have a particular ratio of overlapping sub-syllabic segments.

Besides the restriction options, the user can also specify some output options. These output options are number of syllables, lexicality, old20, ned1, overlap ratio of sub-syllabic segments, and deviation statistics.[2] The Wuggy algorithm is described in Algorithm 2.

Algorithm 2 The Wuggy algorithm.

1. **Inputs:**
   1) The lexicon *L* of the language.
   2) Reference words
2. **Outputs:**
   1) The generated pseudowords.
   2) Selected output options (e.g., Old20, Ned1, ...).
3. The user selects a targeted language and constructs bi-gram chains for the lexicon *L* (see Algorithm 1).
4. The reference words are syllabified using the lexicon.
5. Pseudowords are generated with selected restriction criterias.
6. The selected output options are calculated.

4.4 Turkish Plug-in for Wuggy

The Wuggy software is flexible and gets adapted to other languages easily if three pieces of lexical information of a language are provided. These are a lexical list as a corpus with word

[2] Ned1 is the number of number of words that can be obtained by a single operation of substituting, inserting or deleting of a single letter.

frequency values, a syllabified version of each words in the list and a sub-syllabic pattern of words (the onset-nucleus-coda pattern). The Turkish plug-in was included to the software as a part of the present thesis work and a unique stem list with frequency values was used as the lexical list.

The unique stem list: The list was obtained from Bozşahin et al. (2012) consisting of 24,414 Turkish stem words and it is called a ”unique” list because each word occurs only once even for the words that have multiple meaning. Each of the stems in the list was obtained after a morphological parsing process where the words are parsed such as ”gör-me-sin-de”. The morphological properties of words were disambiguated after the collection of a large amount of ”supertag” sets of the language. These ”supertag” sets not only include the morphological marks (labels such as stem categories and affixes) but also semantic information of the items. Semantic information was obtained by considering the context of each word and used to disambiguate the words in ambiguous contexts. For example, the word ”çizmeleri” can be disambiguated in two ways as”çizme-leri (”çizme[N]-POSS3P)” and ”çiz-me-le-ri (çiz[V]-INF-PERS3P)”. The disambiguated morphological parses were constructed by building a language morpho-model for Turkish words to predict the sequence of supertags by training the model. Overall, the model tags five million words with 94.2% accuracy using 5,917 stems.

The frequency values were also obtained from Bozşahin et al. (2012) study. The authors extracted the frequency values from the BOUN corpus using the words in stem list. The BOUN corpus is a Turkish web corpus obtained from four sub-corpora: three news websites (Milliyet, NTV and Radikal) and one general sampling of Turkish web pages (Sak et al., 2011). The corpus contains 423 million words, 491 million tokens (number of individual words) and 4.1 million types (number of distinct word forms).

Syllabified version of words: The words in the Turkish unique stem list taken from the Bozşahin et al. (2012) study and then syllabified using the hyphenation algorithm that is designed for TeX type setting system (MacKay, 1988).

Defining sub-syllabic pattern of words: The third piece of information for the Turkish plug-in is the onset-nucleus-coda pattern. The Wuggy needs to decompose words into the sub-syllabic elements to construct bi-gram chains as described with the Algorithm 1 (See Section 4.3). Decomposing a syllable into its sub-syllabic elements requires the knowledge of the ONC pattern of a language. Defining the nucleus pattern for a language is sufficient for the Wuggy algorithm which can divide the syllable to its onset, nucleus and coda. Since the syllables in Turkish words have a consonant-vowel-consonant (CVC) pattern, the nucleus can only be vowels of a, e, ı, i, o, ö, u, and ü.

4.5 Behavioral Studies Using the Wuggy Software

Even though only three years was passed from the development of Wuggy, it was used in many different psycholinguistic studies. These studies can be categorized under five different titles of ”Investigating the effect of linguistic variables on lexical processing”, ”Investigating the cognitive mechanisms of lexical processing", "Psycholinguistic lexical databases", "Verbal developmental studies" and "Human lexical decision simulation algorithms".

Studies investigating the effect of linguistic variables on lexical processing: Radanović and Milin (2011) used the Serbian version of Wuggy and compered word detection duration to investigate whether the morphological relatedness influences the cognitive system. In Serbian language, the animate and inanimate entities are distinctively marked that some animate words have a sibling and are morphologically related such as "lav" (lion) and "lavica" (lioness). There are also some other inanimate words that do not have a sibling hence, they are morphologically unrelated such as "otac" (father) and "majka" (mother). A lexical decision task was conducted on 168 Serbian nouns and pseudonouns and the results showed that animate nouns with a sibling were processed faster than the animate nouns without a sibling.

German version of Wuggy was used in another study to investigate the specific effect of word-form frequency, lemma frequency and OLD20 separately on lexical processing (Kresse et al., 2012). The authors used s lexical decision and a speeded naming task to compare the response time and accuracy scores among low and high lemma frequencies, wordform frequencies and OLD20 scores on German nouns. As a result, both the lemma, wordform frequencies and OLD20 scores had an effect on the behavioral scores, but lemma frequency had more effect. Moreover, the response times for low OLD20 words were more than the words having high OLD20 scores.

Studies investigating the cognitive mechanism of lexical processing: Keuleers et al. (2010b) used the Dutch version of Wuggy to investigate the effect of prolonged practice on visual word recognition. A lexical decision task was used which lasted of 57 blocks of 500 trials. The practice effect was measured as the difference between response time and accuracy scores of the first and the last block. As a result, the effect of practice was very small, as small as 40 milliseconds of latency and 2% of accuracy.

The Dutch version of Wuggy was used and the behavioral effect of post-error-slowing (PES) was investigated this time (Dutilh et al., 2012). PES is a common tendency of subjects while performing a behavioral task that their response duration gets increased when they make an error on a previous trial. The study investigates the underlying mechanism of PES using a drift-diffusion model analysis. Thus, subject's accuracy scores are decomposed by taking into account the response times of each single trial. As a result, PES occurs because subjects get more cautious after making an error which slows down their reactions.

The English and French versions of Wuggy were used in another study to investigate how cognate words are processed in adult French-English bilinguals (Peeters et al., 2013). Cognate words have the same root or have similar orthographic representations with another word and can be in two languages (See Section 2.8 for details). The study was a neuro-recording study; the subjects' electrophysiological data was being recorded while they were performing a lexical decision task. The response time and accuracy scores were compared between high and low-frequency English and French words, high-frequency English, low-frequency French words and vice versa. Results showed that cognate words have a "facilitation effect" on word recognition process. Moreover, the electrophysiological data was consistent with the behavioral results; a N400 component was detected in trials displaying the cognate words. N400 component was a a negative component in the EEG waveform that peaks around approximately around 400 milliseconds and associated with lexical processing of high-frequency words (Rugg, 1990).

Psycholinguistic lexical databases: Wuggy was also used in some of the mega projects, the ones used for developing lexical psycholinguistic databases. One of these studies is the Dutch Lexicon Project which consists of 14,000 words (2,807 mono and 11,262 disyllabic) and 14,000 non-words (Keuleers et al., 2010b). The behavioral scores were obtained from 39 participants on a lexical decision task which was repeated over 58 blocks of 500 trials. The behavioral scores were validated by comparing the results with France (FLP) and the English Lexicon Project (ELP) by conducting some virtual experiments.

The Wuggy was also used in the SUBTLEX-NL project; a lexical word-frequency psycholinguistic database in Dutch language (Keuleers et al., 2010a). The SUBTLEX contains 43,729,424 words in the database derived from 8,443 movie subtitles. Unlike the classical word-frequency calculation methods (e.g. K-F word frequency database: See Chapter 2), contextual differences are also considered in the SUBTLEX which had in total of 8,070 different contexts. Reliability of the frequency scores were validated with running a lexical decision task on 39 subjects. As a result, SUBTLEX frequencies predicted both the response times and the accuracy scores more than the CELEX frequencies (Baayen et al., 1993).

The British Lexicon Project (BLP) was another psycholinguistic mega-study that used the Wuggy software (Keuleers et al., 2012). BLP consists of 14,365 words (8,010 mono and 20,720 disyllabic) and 14,365 non-words. The behavioral scores were obtained from 78 participants tested on a lexical decision task which was repeated over 57 blocks of 500 trials. Comparing the scores with ELP demonstrated a high correlation between the two scores. Because the procedure of the task was the same with the one used in DLP, validity of the assessments was confirmed when consistent results were obtained among ELP and DLP.

Verbal developmental studies: Psycholinguistic developmental studies are concerned with the developmental process of the young readers' learning process, the strategies they pursue while reading a text, and the difference between young reading disabled patients and normal developing children.

The study of Henry (2012) was a longitudinal developmental study that used the English version of Wuggy software to investigate the underlying cognitive mechanisms of morphological awareness development in reading disabled subjects. A lexical decision task was used to test the subjects' reading performances regularly between 14 to 17 years. The results showed that, the accuracy and response time scores improved significantly during the experiment duration when the effect of practice was controlled.

In another study, Wuggy was used again but this time to investigate the underlying cognitive mechanisms of the *regularity effect* in grade 3 and grade 4 English speaking children (Schmalz et al., 2013). This effect was observed when young readers apply letter-to-sound rules even to irregularly pronounced words such as "yatch". The results showed that, the regularity effect was observed only for the low-frequency words and the effect was more prominent for children who had poorer reading performances.

Human Lexical Decision Simulation Algorithms: The LD1NN algorithm was was developed by Keuleers and Brysbaert (2011) to test the quality of words and non-words in the lexical decision task. The algorithm simply predicts the word stimuli by calculating word and non-word probabilities of the previously presented trials. The algorithm considers only the orthographic properties of words while making word/non-word predictions that it does not have any information about the words' semantic properties. The method of nearest neighborhood classification and Levenshtein distance were used for calculating the probabilities. The LD1NN algorithm was tested on several psycholinguistic lexical databases and it was found that the non-words of the databases that used the Wuggy software were less prone to bias when compared with the others such as the ARC non-word database (Keuleers and Brysbaert, 2011).

4.6 Summary

Wuggy is a pseudoword generator software which is adaptable to multiple languages. The software is available for Basque, Dutch, English, French, German, Serbian, Spanish and recently for Turkish. Some of the methods for obtaining pseudowords are creating them manually such as modifying one or two letters of real words, using one of the standardized databases the English Lexicon Project, and using one of the available pseudoword generation software programs such as the Wuggy software.

The Wuggy software has an efficient algorithm for generating pseudowords. The algorithm decomposes the syllabified version of the words into its sub-syllabic components (onset, nucleus, coda) and then constructs the bi-gram frequency chains out of these elements. Many pseudowords are generated by different combinations of the sub-syllabic components however, only the pseudowords that have similar transition frequencies with its user-defined template word(s) are permitted to be in the candidate list (detected with the *concentric search* feature).

The Turkish plug-in was included to the algorithm by providing a syllabified lexical database with word-frequency values, and an onset-nucleus-coda pattern of the Turkish words. The corpus was obtained from a Turkish unique stem list which also contains the frequency values (Bozşahin et al., 2012).

This chapter also mentions some of the studies which used the Wuggy software to prepare their behavioral tasks. These studies are about investigating the effect of linguistic variables on lexical processing, investigating the effect of cognitive mechanisms of lexical processing, developing psycholinguistic lexical databases, investigating the process of verbal development and developing an algorithm that simulates the human lexical processing.

CHAPTER 5

Methods

Previous chapters were about the development of the software programs, this chapter explains the methodology of the behavioral study. The two recently developed Turkish psycholinguistic tools were tested by conducting a lexical decision task. The behavioral data was obtained for hypothesis testing. The stimuli set was prepared using the KelimetriK and Wuggy software programs. The former was for the preparation of the words in the stimuli set which varied in frequency (per-million), ON, OLD20 and imageability. Wuggy was used for generation of the pseudowords. The following section explains this preparation process in detail.

5.1 Stimuli

The stimuli consisted of 250 Turkish words and 250 pseudowords (non-words). The words were selected from a Turkish stem list which were derived after a morphological parsing process (Bozşahin et al. (2012); See the Section 4.4 for details). The pseudowords were obtained using the Wuggy software, each was derived from a word stimulus (See Section 4.3 for details). In other words, each pseudoword had the smallest transition frequency values from its temple word, hence it was the best possible candidate out of the other 10 candidate pseudowords.

Selection of Words: There were 250 words in total. The words varied in the dimension of word-frequency (per-million), orthographic neighborhood (ON), OLD20 and imageability (See Chapter 2 for details). Word-frequency scores were obtained from a 30.000-word Turkish stem list (Bozşahin et al., 2012). Range of the frequency values was between 0.021 and 3149.437 per-million (sdv= 339.78). ON scores were obtained using the algorithm of the *KelimetriK* software (See Chapter 3 for details). Range of the ON values was between 0 and 22 (sdv= 2.92). OLD20 scores were also obtained using *KelimetriK's* algorithm. Range of the OLD20 values was between 1 and 2.6 (sdv= 0.21). Imageability scores were obtained from the MRC database (Coltheart, 1981). Range of the imageability values was between 208 and 633 (sdv= 101.23).

All the 250 words in the stimuli set were selected from a candidate list which consisted of 5012 bi-syllablic five-letter-length words. The candidate list was the filtered version of the Turkish stem list, filtered out to control for the letter length and syllable number (Bozşahin et al., 2012). The words in the candidate list also contained the orthographic statistics of word-frequency, N and OLD20.

The 250 word stimuli were chosen randomly from the candidate word list after application of several procedures to make the variance of orthographic scores similar to the ones in the candidate list. Firstly, the variance was calculated for each of the three scores of word-frequency, N and OLD20. Then, these 500.000 different subsets of words were randomly chosen each contained 250 words. Each of the subset's variance its distance from the candidate list was calculated (see Formula 5.1 and 5.2)

$$\underset{S' \subset S, |S'|=250}{\arg\max} \quad dist \tag{5.1}$$

where

$$\begin{aligned} dist = {} & [Var(S_{freq}) - Var(S'_{freq})]^2 \\ & + [Var(S_N) - Var(S'_N)]^2 \\ & + [Var(S_{old20}) - Var(S'_{old20})]^2 \end{aligned} \tag{5.2}$$

In order to measure the distance, the squared differences of each of the three scores' variances were calculated separately and then summed together. The subset having minimum distance from the the 500.000 subsets was chosen as the word stimuli. Each of the words in the subset were translated into English to seek for their Imageability scores on the the MRC database (Coltheart, 1981). Approximately 25% of the words in the subset had to be replaced with the other words that had the most similar orthographic scores. Such replacement was necessary for the stimuli set either because their English version were not in the MRC database or they were a possible cognate with the other words in English or other Latin root languages (See Section2.8 for a detailed explanation of the "cognate effect"). Final version of the word stimuli is available in Appendix A.

Selection of Pseudowords: The pseudowords were selected using the Wuggy software after the Turkish plug-in extension was implemented (See Chapter 4 for more details). Each single word in the word stimuli set was used as a template on while generating the pseudowords. The software was set to generate 10 candidate pseudowords for each reference word. Each of these pseudowords differed from the template by one sub-syllabic element per each syllable. As the template was a bi-syllabic word, each candidate pseudoword differed from the template by two sub-syllabic out of six sub-syllabic components (onset, nucleus and coda per syllable). The pseudoword with minimum smallest deviation from the reference word was included to the stimuli set. Overall, the mean deviation of the pseudowords from their templates was 61.736 (sdv=49.85). Final version of the stimuli list is available in Appendix A.

5.2 Participants

There were 37 participants in the study (21 males and 16 females). The age range was between 21 to 35 years with mean age of 27.18 (sdv=3.32). The subjects were Bilkent University and Middle East Technical University graduate and undergraduate students, volunteered to participate in the study. All the participants signed an informed consent with content of procedures and protocols of the study reviewed by Middle East Technical University human subjects ethics review committee.

5.3 Experiment

Setup and Apparatus: The apparatus were a portable personal computer, and a keyboard as the response device. Software of the task was specially designed for the experiment on Java programming language (`http://java.sun.com`) using the PsychWithJava library (Boyaci, 2006). This library contains relevant functions for preparing psychophysics experiments.

Task: The experimental task was a lexical decision task: participants were expected to decide whether a presented verbal stimulus was a word or pseudoword on each trial. The experiment's software program started with a blank screen, then the participants pressed the ”Space” button whenever they were ready to proceed with the trials. The trial sequence was adapted from Keuleers et al. (2010b) (see Figure 5.1). Each trial started with a display of two parallel vertical fixation lines for 500 milliseconds which were located above and below the center. Then, either a word or a pseudoword stimuli appeared between the two vertical bars for maximally 2000 milliseconds. Subjects were expected to decide whether a presented stimulus was a word or pseudoword and then respond by pressing the ”M” or ”X” buttons. Each subject responded to a word stimulus with their dominant hands, and to a pseudoword stimulus with their non-dominant hands. For example, if a subject was right handed, he/she was expected to use the ”M” button to respond to ”word” stimulus and ”X” button for a ”pseudoword” stimulus. The inter-stimulus-interval (ISI) between the trials was 500 milliseconds. Overall, there were 500 trials in total and completing the entire experiment took approximately 20 minutes.

Procedure: Participants were seated to the front of the experiment computer in a secluded environment. All of the participants were expected to read and sign an informed consent form before starting to the experiment. Then, a questionnaire sheet was given to each participant, prepared for obtaining some personal information that might be related with their behavioral scores: their age, proficiency level in other languages (1-10 self-rating scale) and their reading habit (how many books do you read per month; See Appendix B). The procedure of the experimental task was explained in a detail to subjects before running the task. Then, the participants proceeded with the trials.

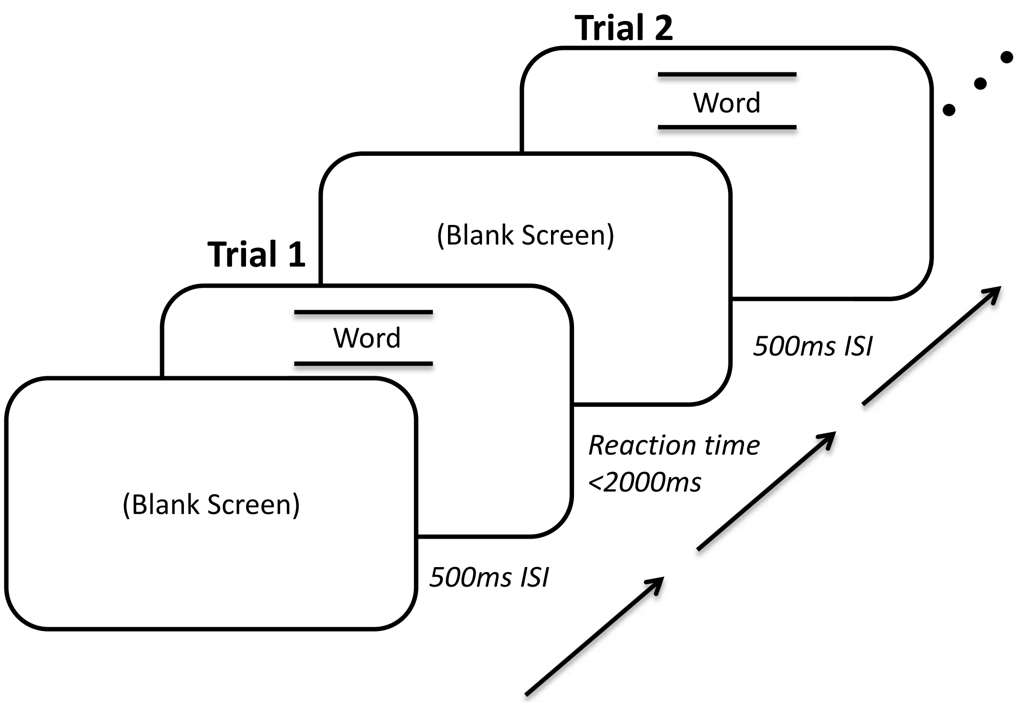

Figure 5.1: *The trial sequence of the experimental task (Adapted from (Keuleers et al., 2010b)).*

5.4 Summary

The task of the study was a lexical decision task with 500 trials. The stimuli set consisted of 250 words and 250 pseudowords. The words were selected randomly from a Turkish stem-list and varied in the dimensions of word-frequency (per-million), orthographic neighborhood (ON), OLD20 and imageability. The pseudowords were generated using the Wuggy pseudoword generator with the Turkish plug-in extension. Each of the words were used as a template to the pseudowords. The task was a lexical decision task and conducted on 37 subjects in the age range of 21 to 35.

CHAPTER 6

Results

6.1 Lexical Properties of Words in the Stimuli Set

Words per category: Figure 6.1 shows the distribution of words per-category in percentage. There are in total of 44 adjectives, 161 nouns, 38 verbs, 2 adverb, 1 post proposition and 4 reduplication. Consequently, more than half of the words in the stimuli set were nouns, and the number of adjectives were more than the nouns.

Frequency of ON: Figure 6.2 shows the distribution of ON size among words. Most of the words have ON sizes between the range of 0 to 5 (85%). In addition, there were only a few words having ON size above 10 (4%). Overall, the distribution of the scores were skewed towards left when compared with the normal curve.

Frequency of OLD20: Figure 6.3 shows the distribution of OLD20 scores among words. It can be seen that, most of the words have OLD20 scores between the range of 1.7 to 2.0 (73%). The scores were not skewed towards right or left when compared with the normal curve however there was a gap in distribution among the scores between the range of 2 and 2.5.

Word-frequency as a function of ON: Figure 6.4 shows the mean word-frequency scores of words among different ON ranges. Mean word-frequency of words in the *0 - 2* condition was 56.6 (sdv=158.5), in the *2 - 4* condition was 166.0 (sdv=534.4) and in the *5 and above* condition was 133.52 (sdv=338.9). Accordingly, the words in the ON range 2 to 4 have the highest word-frequency scores and the words in the ON range 0 to 2 have the lowest word-frequency scores. However, deviation of the scores for conditions *2-4* and *5 and above* were too high to make accurate inferences.

6.2 Behavioral Scores

Two measurements were derived from the lexical decision data: response time and error rates. The mean response time scores were obtained by averaging each subject's word detection duration in a condition. The error rates were the subjects' mean percentage of errors per-condition.

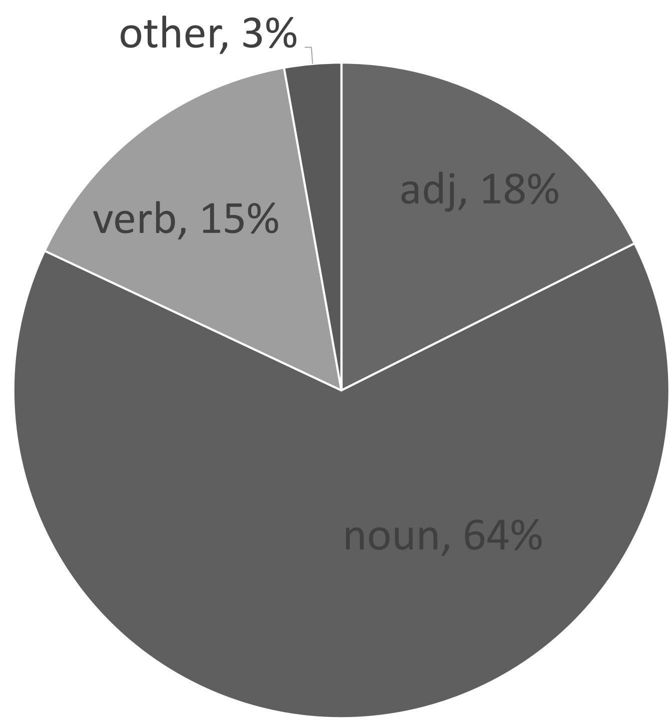

Figure 6.1: *Distribution of words per category.*

The error rates were calculated as:

$$100 * \left(1 - \frac{C}{N}\right)$$

where C was the total number of correct responses and N was the number of subjects.

Comparison of words and non-words: Two *paired samples t-test* were conducted to assess the magnitude of the difference between words and non-words for response times and error rates. As shown in Figure 6.5, the mean response time of words was 720.731 milliseconds (sdv=77.80) and it was 779.56 milliseconds (sdv=90.68) for the non-words. Overall, the words were detected to be 58.83 milliseconds faster than the non-words and this difference was significant at the *alpha* level of 0.05 [t(36)=-7.756, p=0.000].

Figure 6.6 shows the mean percentage of error rates of words which was 16.14% (sdv=6.18) and it was 11.94% (sdv=8.99) for the non-words. Contrary to the expectations, the error rates of words were 4.2% higher than the non-words and this difference was significant at the *alpha* level of 0.05 [t(36)=2.16, p=0.038]. Surprisingly, the error rates of words were higher than

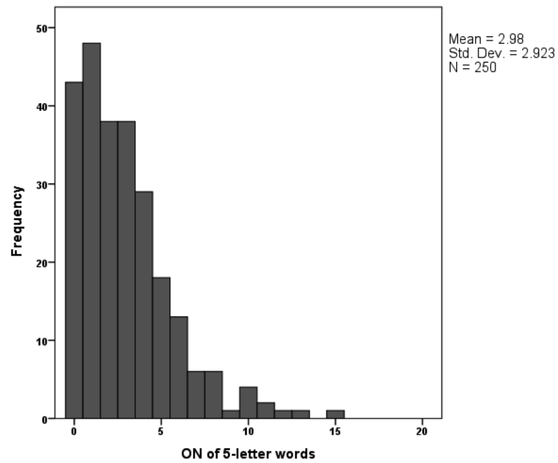

Figure 6.2: *Frequency of orthographic neighborhood of the words in the stimuli set.*

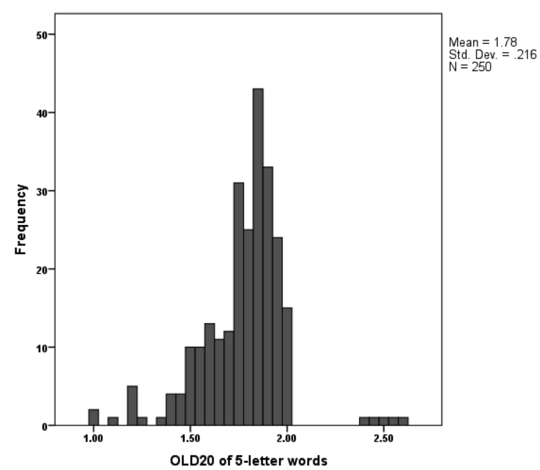

Figure 6.3: *Frequency of OLD20 scores of the words in the stimuli set.*

the non-words and such outcome was presumed to be due to the low-frequency words in the stimuli set. This assumption was tested by looking into the effect of frequency among word and non-word error rates.

The effect of *frequency* and *lexicality* on the error rates: Words and non-words were separated into two conditions according to their frequency level: *low-frequency* with mean the value of 3.65 (sdv=2.25) and *high-frequency* with the mean value of 201.49 (sdv=459.018). Non-words do not have a frequency value but they were divided according the frequency scores of

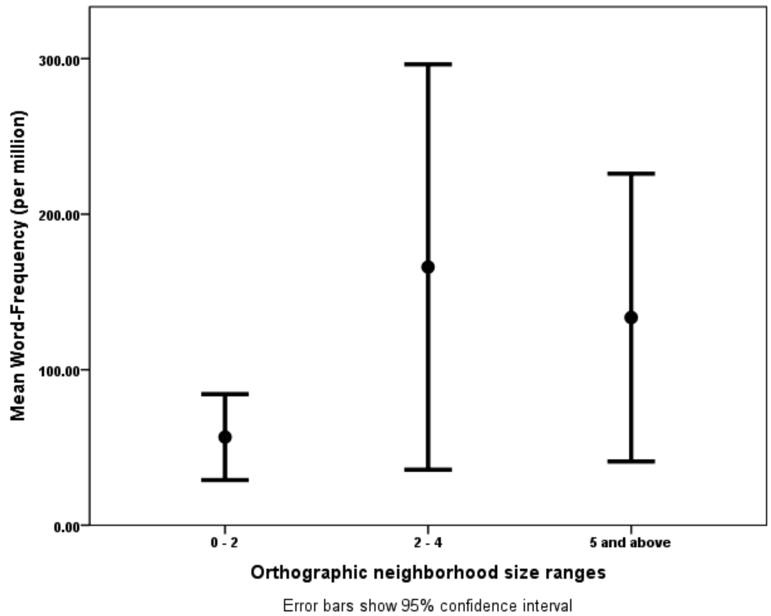

Figure 6.4: *Mean word frequency scores of words among different orthographic neighborhood (ON) size.*

their template words. Such division was applied in order to compare the error rates of words to their non-word counterparts in two different frequency conditions. Figure 6.7 shows the mean percentage of error rates of words and non-words for the two conditions of *low-frequency* and *high-frequency*. Overall, the words' error rates were 17% higher in the low-frequency condition and such difference did not exist for the non-words. A factorial repeated measures analysis of variance (ANOVA) was conducted with two within subject factors of *lexicality* (two levels of *words* and *non-words*) and *frequency* (two levels of *low-frequency* and *high-frequency*). The main effect of *lexicality* was significant at the alpha level of 0.05 [$F(1, 33)=28.67$, $p=0.000$]. The main effect of *frequency* was significant at the *alpha* level of 0.05 [$F(1, 33)=385.35$, $p=0.000$]. The *interaction of lexicality and frequency* was also significant at the *alpha* level of 0.05 [$F(1, 33)=2439.529$, $p=0.000$].

The repeated measures ANOVA analysis demonstrated a systematic interaction of the words' lexicality status and its frequency level on the error rates. The error rates were higher for low-frequency words and lower for high-frequency words when compared with their non-word counterparts. It should be reminded that, the pseudowords do not have frequency values and their template words' frequency values were used only for this case to compare the error rates among words and non-words in two different frequency conditions.

Which orthographic similarity metric (ON or OLD20) has more influence on the behavioral scores? As indicated in Chapter 2, ON is a classical orthographic similarity measurement,

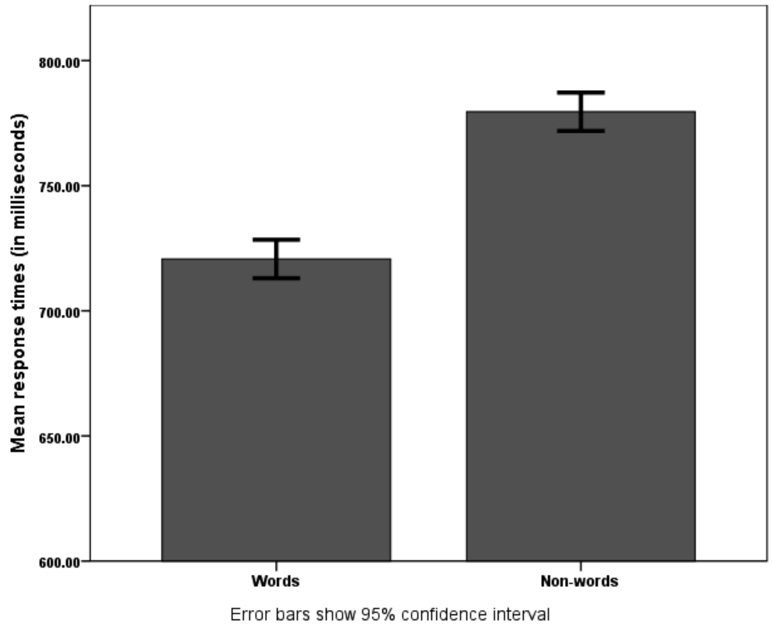

Figure 6.5: *The mean response times of words and non-words.*

whereas OLD20 is a recently developed metric alternative to ON (see Chapter 2). While the effect of ON on behavioral data is contradictory, the effect of OLD20 is consistently "facilitation" (Yap, 2008). The present study compares the response times among the low and high ON and OLD20 scores separately to observe whether an effect exists for Turkish words.

The words in the stimuli set (250 in total) was sorted according to their ON values for one comparison and their OLD20 values for another comparison. The low and high conditions was obtained by splitting the data into half and averaging the response times of each condition. The mean ON size in the *Low ON* condition was 0.96 (sdv= 0.79) and it was 5.13 (sdv= 2.82) in the *High ON* condition. The mean error rates in the *Low-OLD20* condition was 1.63 (sdv=0.17) and it was 1.92 (sdv=0.13) in the *High-OLD20* condition. Lower response times were an indication of a "facilitation effect" which facilitates the recognition of a word, and higher response times were an indication of an "inhibition effect" which is the opposite effect of facilitation, hinders the recognition of a word.

A paired samples t-test analysis was conducted to test the significance of the difference among low and high ON words. As shown in Figure 6.8, the mean response times of words in the *Low ON* condition (m=720.602; sdv= 75.02) was lower than the *High ON* condition (m=718.74; sdv=82.64). However, this difference was not significant at the *alpha* level of 0.05 [t(36)=0.948, p=0.350].

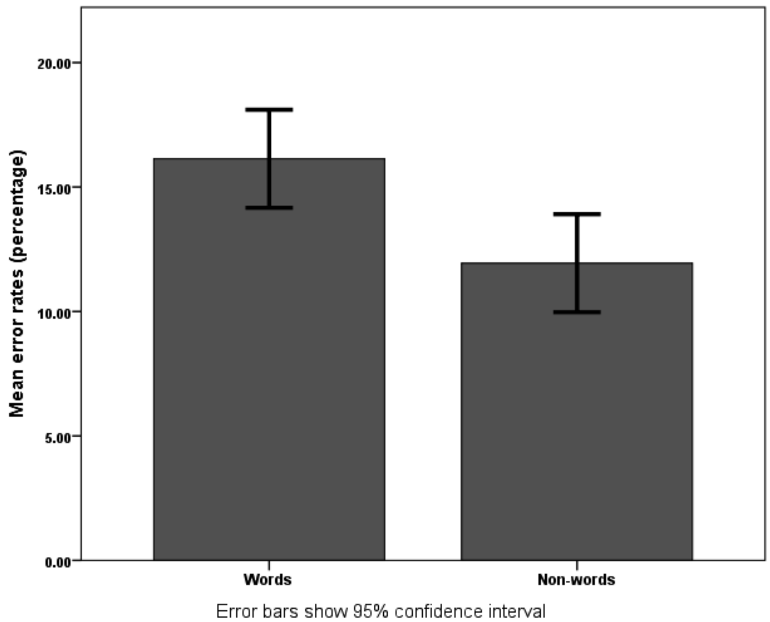

Figure 6.6: *The mean percentage of error rates of words and non-words.*

Figure 6.9 shows the mean response times of words among low and high OLD20 conditions. The mean response times in the *Low OLD20* condition was 724.39 (sdv= 81.11) and in the *High ON* condition was 716.79 (sdv=75.71). It can be seen from the figure that, the mean response times in the *Low OLD20* condition was higher when compared with the *High OLD20* condition and this difference was significant at the *alpha* level of 0.05 [t(36)=2.127, p=0.04]. Contrary to the expectations, low OLD20 scores induced a "facilitation" effect and this might be because of the uncontrolled frequency scores of the words. An interaction analysis was conducted to observe whether the frequency values of words were interacted with the OLD20 variable on response time scores.

Interaction of *frequency* and *OLD20*: The words' frequency scores were sorted from smallest to largest in both low and high OLD20 conditions and divided into half. Four conditions were obtained which were low-frequency (mean= 3.86; sdv=2.36) and low-OLD20 (mean= 1.63; sdv=0.18), low-frequency (mean=3.42 ; sdv=2.11) and high-OLD20 (mean= 1.91; sdv=0.09), high-frequency (mean=279.28 ; sdv= 594.92) and low-OLD20 (mean= 1.63; sdv= 0.17), high-frequency (mean= 117.21; sdv= 218.96) and high-OLD20 (mean= 1.94; sdv= 0.15).

Figure 6.10 shows the mean response times of low and high-frequency words among low and high OLD20 conditions. The mean response time was 92 milliseconds lower for the high-frequency words when compared with the low-frequency words. Moreover, the response times were lowest in the *high-frequency and high OLD20* condition. A factorial repeated mea-

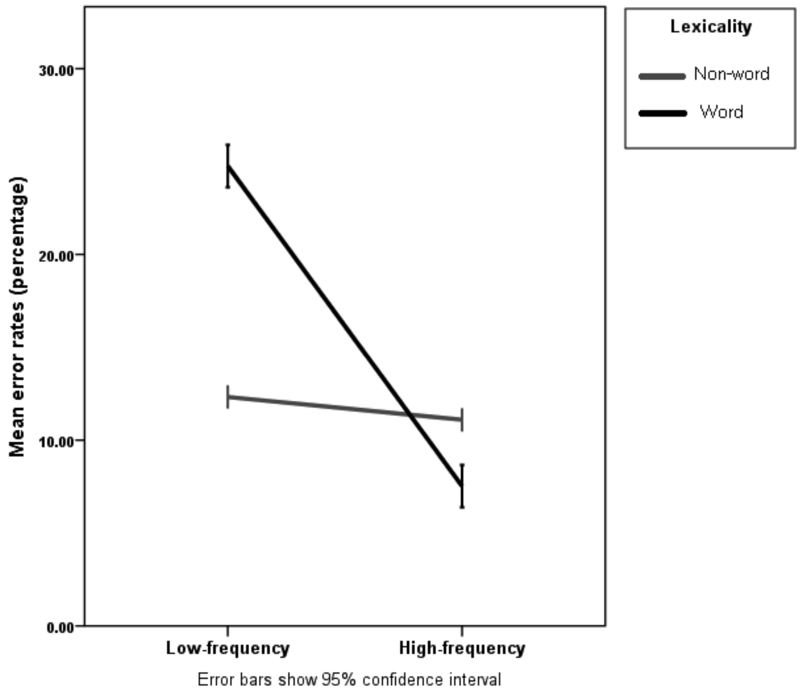

Figure 6.7: *The mean percentage of error rates of words and non-words among two conditions of low-frequency and high-frequency.*

sures ANOVA was conducted with two within subject factors of frequency (two levels of low-frequency and high-frequency) and OLD20 (two levels of low-OLD20 and high-OLD20). The main effect of frequency was significant at the *alpha* level of 0.05 [F(1, 36)=296.78, p=0.000]. The main effect of OLD20 was significant at the alpha level of 0.05 [F(1, 36)=4.59, p=0.039]. However, the interaction of frequency and OLD20 was not significant at the alpha level of 0.05 [F(1, 36)= 1.47, p=0.233]. Overall, conducting an ANOVA analysis demonstrate that OLD20 and frequency scores do not interact with each other.

Despite of the absence of any significant interaction between frequency and OLD20, Figure 6.10 demonstrates that the effect of OLD20 in low-frequency condition vanishes because the error bars of the two OLD20 condition overlaps. On the other hand, the effect of OLD20 is preserved in high-frequency condition because the error bars of the two OLD20 conditions do not overlap with each other. This pattern shows that the effect of OLD20 scores tends to be influenced by the words' frequency scores, especially the low-frequency ones. However, it is not possible to generalize such conclusion without a significant statistical analysis.

The effect of *imageability* on response time and error rates: The response times and error rate scores were sorted from smallest to largest and divided into half. The words which

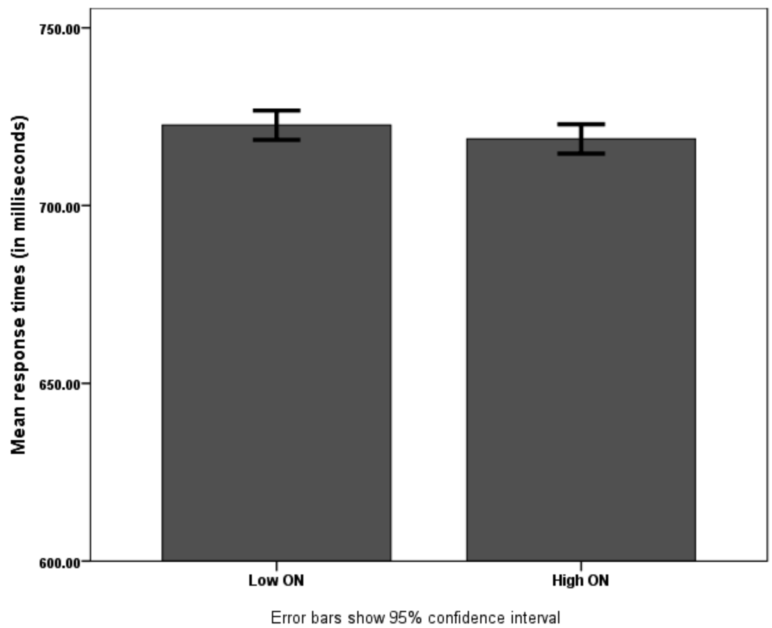

Figure 6.8: *Mean response times of words having low and high ON scores.*

have scores in the first half were labeled as ”low-imageable”, and it was labeled as ”high-imageable” for the second half. Mean *imageability* scores in *low-imageability* condition was 370 (sdv=58.03) and in *high-imageability conditio*n was 541 (sdv=49.46). Two pairwise t-test analyses were conducted to compare the behavioral scores of response time and error rates among low and high imageability condition.

Figure 6.11 shows the mean response times of the low and high imageable words. The mean response times of the low-imageable words was 729.817 (sdv=79.80) and it was 711.05 (sdv=77.49) for high-imageable words. Consequently, there was an 18.765 milliseconds response time difference between low and high imageable words and this difference was significant at the *alpha* level of 0.05 [t(36)=4.93, p=0.000].

Figure 6.12 shows the mean percentage of error rates for the low and high imageable words. The mean percentage of error rates in the *low-imageability* condition was 17.45 (sdv=6.86) and it was 14.74 (6.21) for the high-imageability condition. Overall, there was an 2.71% error rate difference between low and high imageable words and this difference was significant at the *alpha*level of 0.05 [t(36)=3.79, p=0.001].

As a result, the t-test comparisons pointed out a systematic effect of *imageability* on both response time and accuracy scores. An extra analysis was conducted to investigate the interaction of word-frequency and imageability. The response time scores were again divided into four conditions based on divding the imageability scores: low-frequency (mean= 3.69; sdv= 2.28)

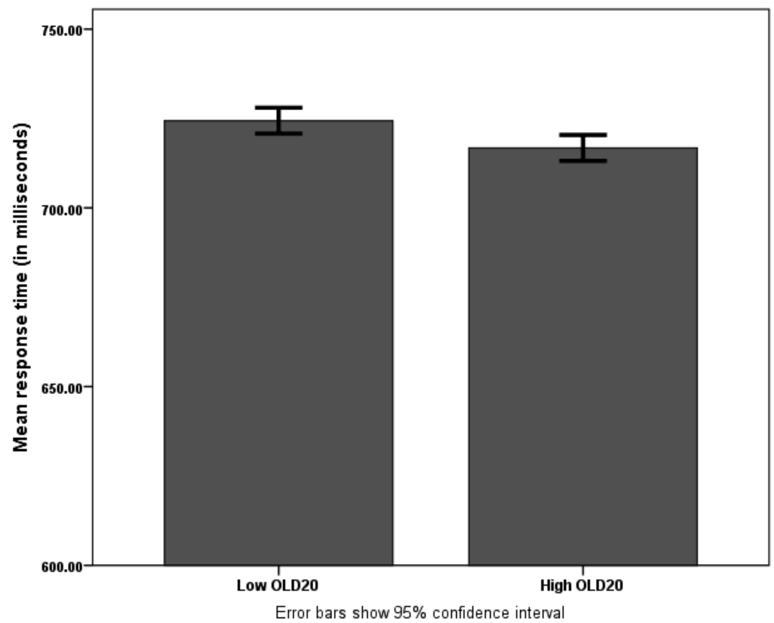

Figure 6.9: *Mean response times of words having low and high OLD20 scores.*

and low-imageability (mean= 371.24; sdv= 52.1), low-frequency (mean=3.61; sdv=2.23) and high-imageability (mean=526.49; sdv=43.84), high-frequency (mean=260.68; sdv=495.61) and low-imageability (mean=364.01; sdv=61.38), high-frequency (mean=141.33; sdv=417.36) and high-imageability (mean=551.01; sdv=54.95).

Figure 6.13 shows the mean response times of the low and high-frequency words among low and high imageability condition. The response times were lower for low frequency words when compared with the high frequency ones. Moreover, the response times were the lowest in the *high-frequency and high-imageability* condition. A factorial repeated measures analysis of variance (ANOVA) was conducted with two within subject factors of *frequency* (two levels of *low-frequency* and *high-frequency*) and *imageability* (two levels of *low-imageability* and *high-imageability*). The main effect of *frequency* was significant at the *alpha* level of 0.05 (F(1, 36)=296.807, p=0.000). The main effect of *imageability* was significant at the alpha level of 0.05 (F(1, 36)=33.91, p=0.00). However, the interaction of *frequency* and *imageability* was not significant at the *alpha* level of 0.05 (F(1, 36)=3.27, p=0.79). As a result, *imageability* and *frequency* influenced the response times individually but not interacted with each other.

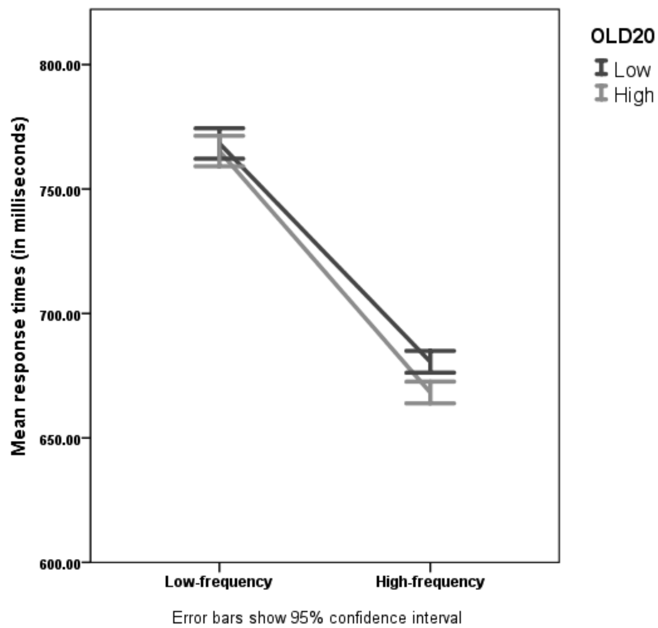

Figure 6.10: *Mean response times of low and high frequency words among low and high OLD20 condition.*

6.3 Control Tests

Relationship between the subjects' literacy level and their behavioral performances: A valid and reliable psycholinguistic data requires an unbiased sampling from the population. Consequently, it is necessary to control for the subjects' literacy level and it was tested by obtaining extra data from the subjects about their reading frequency and the knowledge of other languages (See Appendix B for details). The subjects filled out a questionnaire sheet and reported their proficiency level on non-native languages (on a 1 to 10 likert scale). Moreover, the reading behavior was also reported and quantified by reporting an approximate of number of books read per-month.

Table 6.1 shows the descriptive statistics of the subjects' English proficiency level, number of books they read per-month and the behavioral performance scores. All of the subjects were able to use English proficiently since the median of English proficiency level was 8 on the 1 to 10 scale. Moreover, the median of reported number of book per-month was 1. Behavioral performance scores were obtained from the lexical decision data, the following section describes this process in detail.

Behavioral performance scores: Behavioral performance score is an indication of how accurate a subject responses in the experiment that the higher the score the better it is. The

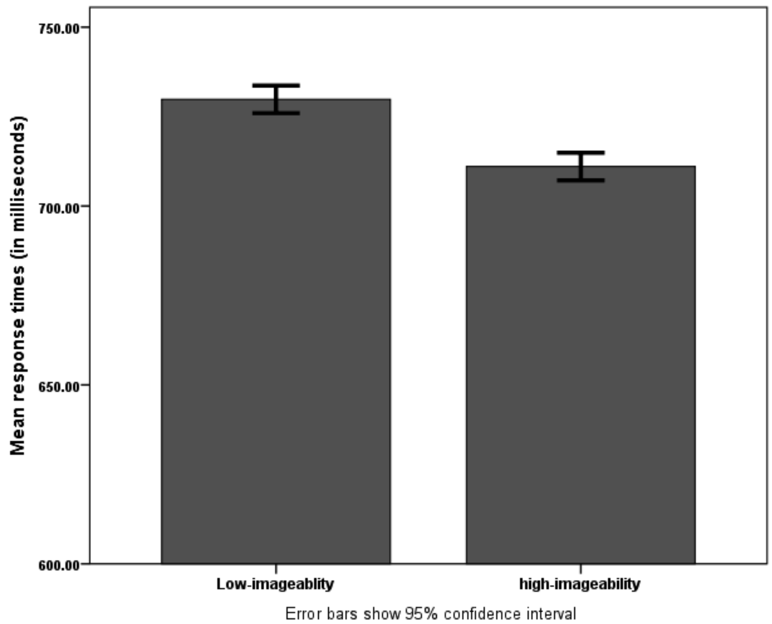

Figure 6.11: *Mean response times of words among low-imageability and high-imageability condition.*

Table 6.1: Descriptive statistics of subject's English proficiency level, Number of books they read per-month and behavioral performance.

Statistics

	Eng. Prof. Level	# of Books per Month	Behavioral Performance
N Valid	37	37	37
N Missing	0	0	0
Mean	8.16	1.568	174.0520
Median	8.00	1.000	177.2545
Mode	8	1.0	78.82
Std. Deviation	1.093	.9827	843.401
Variance	1.195	.974	843.401
Minimum	10	5.0	217.08
Maximum	10	5.0	217.08
Percentile 25	8.00	1.000	155.3284
Percentile 50	8.00	1.000	177.2545
Percentile 100	9.00	2.000	195.3604

calculated formula was designed to promote the correct answers and to penalize the wrong answers. Moreover, early correct responses were promoted over the late corrected responses

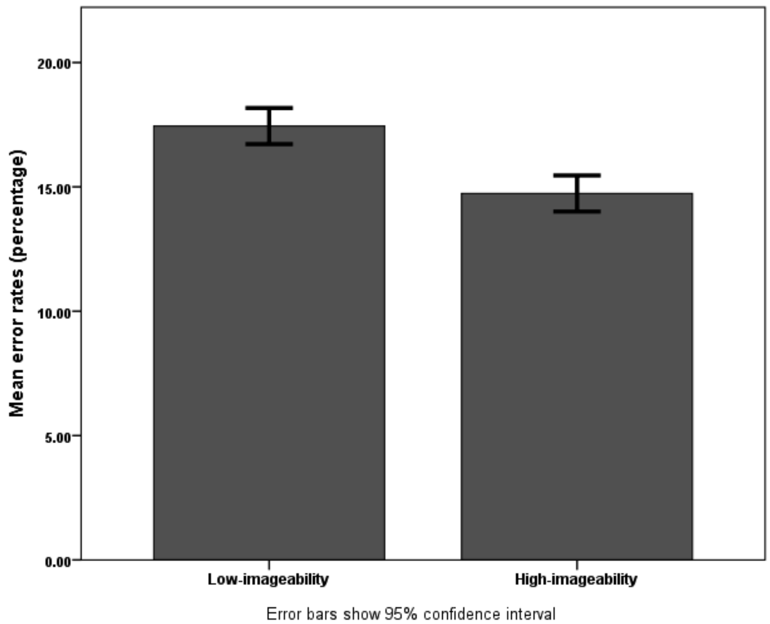

Figure 6.12: *Mean percentage of error rates of words between low-imageability and high-imageability condition.*

and the late wrong answers were penalized more than the early wrong answers. The first term in Formula 6.1 is for the correct answers. The R represents the response time of the user and the value 2000 was selected because it was the maximum time limit to respond for a trial in milliseconds (See section 5.3). The quicker a correct response is the higher the score will be (there is an extra promotion for correct responses). The second term is for the incorrect answers that the late wrong answers were demoted over the quick wrong answers. Lastly, the third term is for the missing answers, the ones that the subject couldn't answer within 2000 milliseconds. The value 2 was obtained by dividing 2000 by the average time to answer (1000). This was selected this way to promote the missing answers over the quick wrong answers. The formula of the behavioral performance score calculation is;

$$score = \log_{10}\left(\frac{2000}{R}\right) - \log_{10}\left(\frac{2000}{2000-R}\right) - \log_{10} 2 \tag{6.1}$$

where R is the response time of the subject.

Reported language proficiency: A Spearman's rho non-parametric correlation test was conducted to test the relationship between the reported English proficiency levels and the behavioral performances. As a result, there was not a significant relationship between the reported English proficiency level and the behavioral performances [r= 0.116, p=0.495, N=37].

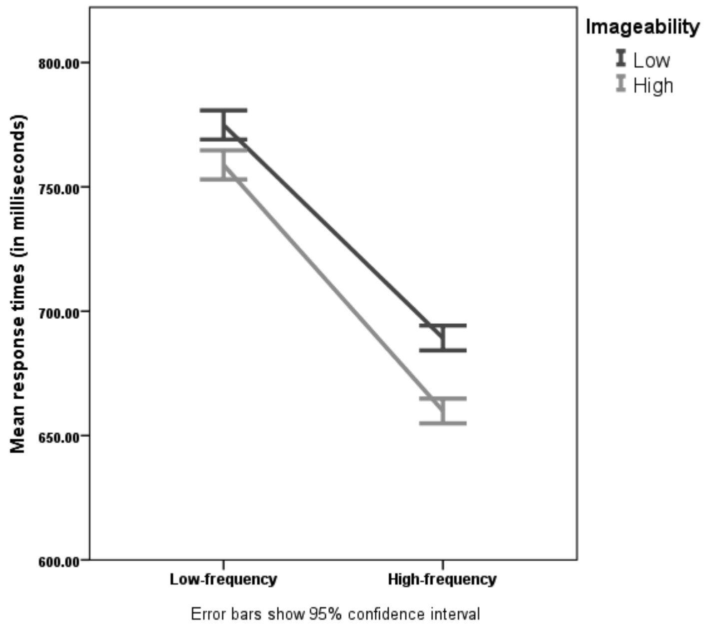

Figure 6.13: *Mean response times of low and high frequency words among low and high Imageability condition.*

Proficiency in more than one non-native languages may have influence on the subjects' lexical decision performances. Some of the subjects were able to speak two non-native languages proficiently. Only, the subjects' who rated a second language 5 and more (on the 1 to 10 scale) were assumed to be proficient on an additional language. A between subject's t-test was conducted to observe whether there was an influence of any second language on the behavioral scores.

Figure 6.14 shows the mean behavioral performance of one and two non-native language speakers. The mean behavioral performance scores were lower in *two-language condition* when compared with the *one-language condition*. Nevertheless, a between subject's t-test analysis showed that this difference was not significant at the *alpha* level of 0.05 [t(35)=1.535, p=0.134].

Reported reading frequency: A Spearman's rho non-parametric correlation test was conducted to assess whether there was any relationship between the subjects' reported reading frequency and their behavioral performances. As a result, there was not a significant relationship between the reported reading frequency and the behavioral performances [r = -0.94, p=0.580, N=37]. In short, the subjects' reading frequency was not anyway related with the behavioral performance.

An extra correlation analysis was conducted to test a possible relationship between the sub-

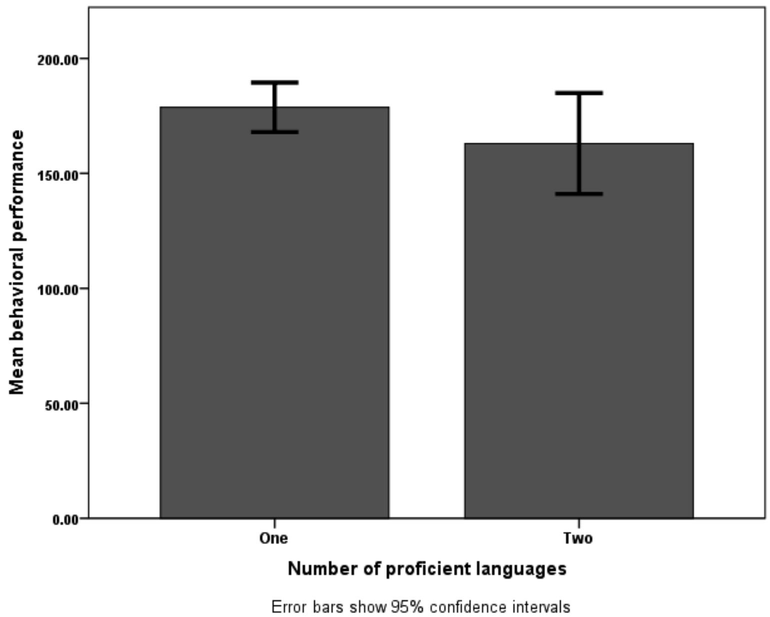

Figure 6.14: *Mean behavioral performance between one and two non-native language speakers.*

ject's reported English proficiency level and their reported reading frequency. Results of the Spearman's non-parametric correlation test highlighted a positive relationship between the two scores [r=0.472, p=0.003, N=37]. The explanatory power of this correlation test was 0.22. This means that 22% of the subjects' English proficiency levels are explained by their reported reading frequencies.

In this section, homogeneity of the subjects' literacy level was tested as a control for the validity of the lexical decision data. The literacy level measurements were obtained from a questionnaire which was quantified as the number of languages a subject knows, its English proficiency levels, and the number of books he/she reads per-month. As a result, there was not a significant relationship between the reported English proficiency levels and the behavioral performances. There was also not a relationship between the reported reading frequencies and the behavioral performances. Moreover, number of learned languages did not also had any effect on the behavioral performances. Whereas, there was a significant positive relationship between the reported English proficiency levels and the reading frequency scores. This significant relationship could be considered as a proof for a valid literacy level assessments.

Variations in the pseudowords' transition frequencies: Each non-word was derived from a word in the stimuli list (as its template word) using the *Wuggy software*. Each non-word was the best candidate to be in the stimuli set out of 10 candidates in terms of having the smallest transition frequency score. The transition frequency values are an indication of how

”wordlikely” a non-word is. The more a non-word has higher deviation score, the more it is dissimilar with its template word (See Chapter 4 for details).

A correlation analysis was conducted to observe the relationship between *deviation frequency* and *RT difference of non-words and words*. This measurement was obtained by subtracting each pseudoword's response time score from its template word. Results of Spearman's rho non-parametric correlation test showed that there was not a significant relationship between the two scores [r= 0.023, p=0.712, N=250]. Overall, the lexical decision data was not influenced by any variations in the ”word-likeliness” level of non-words. Moreover, this result was an indication of an homogeneity among the pseudowords.

6.4 Summary

The stimuli set was noun-dominant and most of the words had neighborhood size between 0 and 5, OLD20 scores between 1.7 and 2.0. The frequency scores among the words did not differ on different neighborhood size ranges.

Two behavioral measurements were obtained from the lexical decision task which was response times and error rates. The results showed that both of the scores were significantly different among words and pseudowords. For error rates, there was an interaction of word-frequency which resulted to be higher for words instead of pseudowords. The effect of ON and OLD20 on the behavioral data was also tested in this study. ON and OLD20 are both orthographic similarity measurements used in psycholinguistic studies. The results showed that OLD20 had an effect on the behavioral scores but not ON. The interaction of OLD20 and frequency was also tested and it was found to be not significant. The final linguistic variable to be tested was imageability, and the results pointed out a significant effect of the variable on both response times and accuracy scores. The interaction of frequency and imageability was also tested and again it was not significant.

There were two control tests in the study. The first control test was conducted to investigate whether variation on the subjects' literacy level had any influence on their behavioral performances. The literacy level was quantified as the reported reading frequency and the reported proficiency level on other languages. The results confirmed that there was not a significant relationship between both of the literacy level assessments and the behavioral performances. There was a significant relationship between proficiency level on English and reading frequency scores which was a proof of validity of the assessments. The validity of the pseudowords was also assessed with testing the relationship between the pseudowords' deviation scores and response time differences among words and pseudowords. As a result, there was not a significant relationship between the two scores.

CHAPTER 7

Discussion and Conclusion

Two psycholinguistic software programs were developed for Turkish visual word recognition studies: *KelimetriK* and *Wuggy with Turkish plug-in extension*. *KelimetriK* was developed to provide several orthographic statistics of a queried word. These orthographic statistics are word-frequency, bi-gram/tri-gram frequency, orthographic neighborhood (ON), transposed letter similarity, subset/superset similarity and OLD20 (See Chapter 2 for detailed explanation of the linguistic variables). The Wuggy pseudoword generator is now available to prepare stimuli sets for Turkish visual word recognition studies. The software programs were tested on a Turkish lexical decision study to provide sample hypothesis testing models to the users.

In the behavioral study, there were in total of 500 words (250 words and 250 pseudowords) all of which were bi-syllabic and composed of five-letters. The KelimetriK software was used for selecting the words with specified orthographic features, and Wuggy with Turkish plug-in extension was used for generating the pseudowords. All the words in the stimuli set were obtained from the filtered version of (Bozşahin et al., 2012) stem list which was filtered out to control for syllable and letter number. The stimuli words were randomly chosen from the filtered version of the list and varied in the dimensions of word-frequency (per-million), ON, OLD20 and imageability. The variance of each of these scores were selected to be similar to the variance of the scores in the filtered list. The imageability scores were included as a final step on the word stimuli preparation process. The selected pseudowords was the best candidate to be selected to the stimuli set because it had the smallest transition frequency from its template. The template words were taken from the stimuli set. It was proved with a control test that the variation of the transition frequency scores did not had any relationship with the detection duration difference of pseudowords and words.

Several descriptive statistics were obtained from the words' lexical properties. The results showed that, the distribution of ON size among words had a left-skewed curve because most of the words have an ON size between the ranges of 0-5. This distribution pattern of ON scores was also similar to five-letter English, Dutch, German and French words on standardized lexical databases(Duyck et al., 2004). On the other hand, the distribution of OLD20 words were not skewed on either sides and was more similar to the normal curve.

The first goal of this thesis work was to use the recently developed Turkish psycholinguistic tools on a visual word recognition study. This goal was accomplished by providing many different hypothesis-testing models to the users on a carefully controlled behavioral study. The second goal of the study was to investigate the effect of several linguistic variables on the behavioral scores for Turkish visual word recognition data and then to compare these results with other studies.

7.1 Lexical Processing of Turkish Words and Pseudowords

The difference between the words and the pseudowords were significantly different for both reaction time and accuracy scores. The lexical processing of the pseudowords took 59 milliseconds more duration for Turkish words. This duration was 15 milliseconds in British English, 20 milliseconds in Dutch and 126 milliseconds in American English (Balota et al., 2007; Keuleers et al., 2010b, 2012). However, direct comparison of the words among these studies is not so accurate because the present study had results only for bi-syllabic five letter words.

Contrary to the expectations, the error rates were higher for the words than the pseudowords. The frequency scores were detected to cause an artifact to the error rates which leaded to a huge difference among low and high frequency words. The difference was as large as 17% among the two-frequency conditions. Such *frequency effect* on the error rates was only present on this study and was not encountered on other visual word recognition studies. As a consequence, more research is necessary to find a plausible explanation for this kind of unpredictable effect.

7.2 The effect of Frequency on Turkish Lexical Processing

Word-frequency was the most effective variable on the Turkish lexical decision data. The consistency of this frequency effect on Turkish words were confirmed when its interaction with lexicality, OLD20 and imageability was tested on separate runs. The results showed that, word-frequency is a very powerful orthographic variable and influence the Turkish lexical decision data very strongly.

Overall, there were 92 milliseconds lexical detection difference among low and high frequency for Turkish words. For English words, this difference was 196 milliseconds on Forster and Chambers (1973) lexical decision study. However, the authors did not control for the effect of length on their stimuli set hence, the comparison may not be accurate.

7.3 Specific Influences of ON and OLD20 Variables on Turkish Lexical Processing

ON and OLD20 scores were developed as a measurement of orthographic similarity in visual word recognition studies (Coltheart et al., 1977; Yarkoni et al., 2008). As indicated in Chapter

2, both of the variables affected the lexical processing on some other languages (See section 2.4 and section 2.6). The present study investigated the specific influence of the ON and OLD20 variables on Turkish visual word recognition studies. Both of the variables were expected to influence the lexical processing and result in a facilitation effect, whereas a stronger effect is expected for the OLD20 variable. Unlike the predictions, the only systematic effect observed in this study was for the OLD20 variable whereas the effect was opposite with the previous studies (Yarkoni et al., 2008; Kresse et al., 2012).

Previous studies on English words highlighted an unambiguous effect of ON on the behavioral data, which could either facilitate or inhibit the lexical processing (Andrews, 1997). The present study could not obtain any systematic effect of the ON variable on Turkish lexical processing. The reason of this might be because of a hidden variable such as word-frequency or imageability. Thus, a possible future study should investigate the possible source of the confounding factor on the ON variable.

The studies which investigated the effect of OLD20 on English and German language reported a consistent effect of facilitation for the high OLD20 words (Yarkoni et al., 2008; Kresse et al., 2012). Contrary to the expectations, the present study obtained a facilitation effect for the words having low OLD20 scores. An interaction analysis was conducted to investigate the source of such an "opposite effect" which was presumed to be because of the uncontrolled frequency scores. As a result, the interaction of word-frequency and OLD20 was not significant. However, the effect of OLD20 for the high-frequency words was consistent with the expectations which was a facilitation effect for high OLD20 words. Overall, all these results demonstrate that the word-frequency scores should be controlled in visual word recognition studies before testing the effect of OLD20 on behavioral data.

7.4 The effect of *Imageability* on Turkish Lexical Processing

The effect of imageability was significant for both response times and accuracy scores. This result supported the hypothesis that the detection of the low-imageable words took 19 milliseconds more duration than the high-imageable words. Moreover, words that had higher imageability scores resulted in 3% more accurate results than the low imageable words. This result was consistent when compared with an English lexical decision study (McMullen and Bryden, 1987). However, the effect of imageability in this study was larger than the present study which was 90 milliseconds for response times and 10% for accuracy scores. This difference might be because of the variations of imageability assessments. The present study used the imageability scores prepared for English words (from the MRC database; see Section 2.8 for details) for a Turkish psycholinguistic study. As a consequence, it is necessary to replicate the study with imageability scores obtained for Turkish words.

This study also investigated a possible interaction among imageability and frequency on Turkish lexical detection durations. The results showed that both imageability and frequency had separate effect on word detection duration but not interacted with each other. The response times were higher in low frequency conditions and high imageable words had lower response times on both low and high frequency conditions. This pattern was also the same for English lexical decision latencies in McMullen and Bryden (1987) study which indicated a proof of validity of the assessments.

7.5 Limitations and Future Directions

All of the subjects in the present study were able to use English proficiently. Even though it was proved that knowledge of a second language did not had any effect on the subjects' behavioral performances, the results should be replicated with a group of subjects who can speak only Turkish as their native language to increase the reliability. The population sampling of this study was homogeneous in terms of the literacy levels that all of the subjects had at least a bachelor's degree. Additional control tests was also conducted to prove that the behavioral scores did not had any relationship with variations on the subjects' literacy levels. In any case, results of the study should be validated by replicating the results on subject populations from different literacy levels.

The present study investigated the effect of some of the lexical variables on Turkish lexical processing. However, there are still many other lexical variables left out for future studies. Some of these variables are word length, bi-gram and tri-gram frequency, subset and superset similarity, age-of-acquisition and concreteness.

Word stimuli in the present study was unbalanced on category division and mostly noun dominant. The words was randomly selected from a Turkish stem list, and the only factor considered was keeping the variance of the orthographic scores equal. A possible future study should replicate the same results with another stimuli list balanced on word category to increase its reliability.

The present study was only concerned with the orthographic representation of words because it was testing the recently developed Turkish psycholinguistic tools which only provided orthographic verbal scores. However, the users should be aware that phonology is also an important notion for any kind of verbal stimuli because visual lexical detection is not a core visual perception process that it also involves linguistic processing. The words are phonological units because their sub-components contain phonemic items in different categories like consonants and vowels which influence lexical processing (Frost, 1998). Therefore, a possible future study should investigate whether visual word processing of Turkish words is influenced by different type of consonants and vowels.

7.6 Summary and Conclusion

KelimetriK and Wuggy with Turkish plug-in was the two software programs developed as an helpful tools for Turkish visual word recognition studies. The KelimetriK software calculates several orthographic such as the OLD20 and the ON size of a queried word. Wuggy is a pseudoword generator adaptable to multiple languages. The present thesis tested the two recently developed software programs on a lexical decision task. All of the 250 words in the stimuli set was prepared using the KelimetriK software and they were varied on the word-frequency (per-million), ON, OLD20 and imageability. The 250 pseudowords in the stimuli set was obtained using the Wuggy software with Turkish plug-in. The task of the behavioral study was a lexical decision task and the response time and error rate scores were obtained for hypothesis testing.

The results showed that the processing of words and pseudowords are significantly different for both the response times and the error rates. Moreover, there was a significant effect of frequency, OLD20, and imageability scores on the Turkish words. The only significant interaction obtained was between the lexicality and frequency of words in the error rate scores which caused the scores to be higher for words.

Overall, the present study tested the behavioral effect for some of the linguistic variables on Turkish lexical decision data whereas, there are still many other linguistic variables are waiting to be investigated as a possible future study. To conclude, the present study confirmed that the effect of linguistic variables for Turkish lexical decision data was consistent with the results obtained for other languages.

CHAPTER 8

References

Adelman, J. S., Brown, G. D., and Quesada, J. F. (2006). Contextual diversity, not word frequency, determines word-naming and lexical decision times. *Psychological Science*, 17(9):814–823.

Andrews, S. (1997). The effect of orthographic similarity on lexical retrieval: Resolving neighborhood conflicts. *Psychonomic Bulletin & Review*, 4(4):439–461.

Baayen, R. H., Piepenbrock, R., and van H, R. (1993). The {CELEX} lexical data base on {CD-ROM}.

Balota, D. A., Yap, M. J., Hutchison, K. A., Cortese, M. J., Kessler, B., Loftis, B., Neely, J. H., Nelson, D. L., Simpson, G. B., and Treiman, R. (2007). The english lexicon project. *Behavior Research Methods*, 39(3):445–459.

Bird, H., Franklin, S., and Howard, D. (2001). Age of acquisition and imageability ratings for a large set of words, including verbs and function words. *Behavior Research Methods, Instruments, & Computers*, 33(1):73–79.

Bowers, J. S., Davis, C. J., and Hanley, D. A. (2005). Automatic semantic activation of embedded words: Is there a ”hat” in ”that”? *Journal of Memory and Language*, 52(1):131–143.

Bozşahin, C., Köprü, Ş., and Şirin, U. (2012). Morphology by itself and morphology for itself: Morpho-supertagging without analysis.

Brysbaert, M. and New, B. (2009). Moving beyond kučera and francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for american english. *Behavior Research Methods*, 41(4):977–990.

Burgess, C., Livesay, K., and Lund, K. (1998). Explorations in context space: Words, sentences, discourse. *Discourse Processes*, 25(2-3):211–257.

Burnard, L. (1995). Users reference guide british national corpus version 1.0.

Carroll, J. B. and White, M. N. (1973). Word frequency and age of acquisition as determiners of picture-naming latency. *The Quarterly Journal of Experimental Psychology*, 25(1):85–95.

Chumbley, J. I. and Balota, D. A. (1984). A word's meaning affects the decision in lexical decision. *Memory & Cognition*, 12(6):590–606.

Cibelli, E. (2012). Shared early pathways of word and pseudoword processing: Evidence from high-density electrocorticography.

Coltheart, M. (1981). The mrc psycholinguistic database. *The Quarterly Journal of Experimental Psychology*, 33(4):497–505.

Coltheart, M., Davelaar, E., Jonasson, T., and Besner, D. (1977). Access to the internal lexicon.

Davis, C. J. (2005). N-watch: A program for deriving neighborhood size and other psycholinguistic statistics. *Behavior Research Methods*, 37(1):65–70.

Davis, C. J. and Perea, M. (2005). Buscapalabras: A program for deriving orthographic and phonological neighborhood statistics and other psycholinguistic indices in spanish. *Behavior Research Methods*, 37(4):665–671.

Dijkstra, T., Miwa, K., Brummelhuis, B., Sappelli, M., and Baayen, H. (2010). How cross-language similarity and task demands affect cognate recognition. *Journal of Memory and language*, 62(3):284–301.

Dutilh, G., Vandekerckhove, J., Forstmann, B. U., Keuleers, E., Brysbaert, M., and Wagenmakers, E.-J. (2012). Testing theories of post-error slowing. *Attention, Perception, & Psychophysics*, 74(2):454–465.

Duyck, W., Desmet, T., Verbeke, L. P., and Brysbaert, M. (2004). Wordgen: A tool for word selection and nonword generation in dutch, english, german, and french. *Behavior Research Methods*, 36(3):488–499.

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

Forster, K. I. and Chambers, S. M. (1973). Lexical access and naming time. *Journal of verbal learning and verbal behavior*, 12(6):627–635.

Frost, R. (1998). Toward a strong phonological theory of visual word recognition: true issues and false trails. *Psychological bulletin*, 123(1):71.

Gibbs, P. and Van Orden, G. C. (1998). Pathway selection's utility for control of word recognition. *Journal of Experimental Psychology: Human Perception and Performance*, 24(4):1162.

Good, R. and Kaminski, R. (2002). Dynamic indicators of basic early literacy skills . eugene, or: Institute for the development of educational achievement. retrieved feb. 1, 2004. *dibels. uoregon. edu*.

Grainger, J., Bouttevin, S., Truc, C., Bastien, M., and Ziegler, J. (2003). Word superiority, pseudoword superiority, and learning to read: A comparison of dyslexic and normal readers. *Brain and Language*, 87(3):432–440.

Hauk, O., Patterson, K., Woollams, A., Watling, L., Pulvermüller, F., and Rogers, T. (2006). [q:] when would you prefer a sossage to a sausage?[a:] at about 100 msec. erp correlates of orthographic typicality and lexicality in written word recognition. *Journal of cognitive neuroscience*, 18(5):818–832.

Henry, R. (2012). Complex word processing in teenage poor readers-does morphological knowledge help or hinder?

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

Keuleers, E. and Brysbaert, M. (2011). Detecting inherent bias in lexical decision experiments with the ld1nn algorithm. *The Mental Lexicon*, 6(1):34–52.

Keuleers, E., Brysbaert, M., and New, B. (2010a). Subtlex-nl: A new measure for dutch word frequency based on film subtitles. *Behavior Research Methods*, 42(3):643–650.

Keuleers, E., Diependaele, K., and Brysbaert, M. (2010b). Practice effects in large-scale visual word recognition studies: A lexical decision study on 14,000 dutch mono-and disyllabic words and nonwords. *Frontiers in Psychology*, 1.

Keuleers, E., Lacey, P., Rastle, K., and Brysbaert, M. (2012). The british lexicon project: Lexical decision data for 28,730 monosyllabic and disyllabic english words. *Behavior Research Methods*, 44(1):287–304.

Kresse, L., Kirschner, S., Dipper, S., and Belke, E. (2012). Towards exploring the specific influences of wordform frequency, lemma frequency and old20 on visual word recognition and reading aloud. *Lexical Resources in Psycholinguistic Research*, page 9.

Kučera, H. and Francis, W. N. (1967). *Computational analysis of present-day American English*. Dartmouth Publishing Group.

Lass, R. (1984). *Phonology: an introduction to basic concepts*. Cambridge University Press.

Levenshtein, V. I. (1966). Binary codes capable of correcting deletions, insertions and reversals. In *Soviet physics doklady*, volume 10, page 707.

Lupker, S. J., Perea, M., and Davis, C. J. (2008). Transposed-letter effects: Consonants, vowels and letter frequency. *Language and Cognitive Processes*, 23(1):93–116.

MacKay, P. A. (1988). Turkish hyphenations for tex. *TUGboat*, 9(1):12–14.

Mainy, N., Jung, J., Baciu, M., Kahane, P., Schoendorff, B., Minotti, L., Hoffmann, D., Bertrand, O., and Lachaux, J.-P. (2007). Cortical dynamics of word recognition. *Human brain mapping*, 29(11):1215–1230.

Manning, C. D., Schütze, H., and MITCogNet (1999). *Foundations of statistical natural language processing*, volume 999. MIT Press.

Marcel, A. J. and Patterson, K. E. (1978). Word recognition and production: Reciprocity in clinical and normal studies. *Attention and performance VII. Hillsdale, NJ: Erlbaum*, pages 209–226.

Massaro, D. W., Venezky, R. L., and Taylor, G. A. (1979). Orthographic regularity, positional frequency, and visual processing of letter strings. *Journal of Experimental Psychology: General*, 108(1):107.

McMullen, P. A. and Bryden, M. P. (1987). The effects of word imageability and frequency on hemispheric asymmetry in lexical decisions. *Brain and Language*, 31(1):11–25.

Medler, D. and Binder, J. (2005). Mcword: An on-line orthographic database of the english language. *Retrieved May*, 14:2008.

New, B., Ferrand, L., Pallier, C., and Brysbaert, M. (2006). Reexamining the word length effect in visual word recognition: New evidence from the english lexicon project. *Psychonomic Bulletin & Review*, 13(1):45–52.

Paivio, A., Yuille, J. C., and Madigan, S. A. (1968). Concreteness, imagery, and meaningfulness values for 925 nouns. *Journal of experimental psychology*, 76(1p2):1.

Peereman, R., Bonin, P., et al. (1998). Is perception a two-way street? the case of feedback consistency in visual word recognition. *Journal of Memory and Language*, 39(2):151–174.

Peeters, D., Dijkstra, T., and Grainger, J. (2013). The representation and processing of identical cognates by late bilinguals: Rt and erp effects. *Journal of Memory and Language*, 68(4):315–332.

Perea, M., Duñabeitia, J. A., and Carreiras, M. (2008). Transposed-letter priming effects for close versus distant transpositions. *Experimental Psychology*, 55(6):384.

Perea, M. and Lupker, S. J. (2004). Can¡ i¿ caniso¡/i¿ activate¡ i¿ casino¡/i¿? transposed-letter similarity effects with nonadjacent letter positions. *Journal of Memory and Language*, 51(2):231–246.

Perea, M. and Pollatsek, A. (1998). The effects of neighborhood frequency in reading and lexical decision. *Journal of experimental psychology: human perception and performance*, 24(3):767.

Price, C. J., Wise, R., and Frackowiak, R. (1996). Demonstrating the implicit processing of visually presented words and pseudowords. *Cerebral Cortex*, 6(1):62–70.

Radanović, J. and Milin, P. (2011). Morpho-semantic properties of serbian nouns: Animacy and gender pairs. *Psihologija*, 44(4):343–366.

Raettig, T. and Kotz, S. A. (2008). Auditory processing of different types of pseudo-words: an event-related fmri study. *Neuroimage*, 39(3):1420–1428.

Rastle, K., Harrington, J., and Coltheart, M. (2002). 358,534 nonwords: The arc nonword database. *The Quarterly Journal of Experimental Psychology: Section A*, 55(4):1339–1362.

Rugg, M. D. (1990). Event-related brain potentials dissociate repetition effects of high-and low-frequency words. *Memory & Cognition*, 18(4):367–379.

Sak, H., Güngör, T., and Saraçlar, M. (2011). Resources for turkish morphological processing. *Language resources and evaluation*, 45(2):249–261.

Schmalz, X., Marinus, E., and Castles, A. (2013). Phonological decoding or direct access? regularity effects in lexical decisions of grade 3 and 4 children. *The Quarterly Journal of Experimental Psychology*, 66(2):338–346.

Seidenberg, M. S. and McClelland, J. L. (1989). A distributed, developmental model of word recognition and naming. *Psychological Review; Psychological Review*, 96(4):523.

Simos, P. G., Breier, J. I., Fletcher, J. M., Foorman, B. R., Bergman, E., Fishbeck, K., and Papanicolaou, A. C. (2000). Brain activation profiles in dyslexic children during non-word reading: a magnetic source imaging study. *Neuroscience letters*, 290(1):61–65.

Stubbe, R. (2012). Do pseudoword false alarm rates and overestimation rates in yes/no vocabulary tests change with japanese university students' english ability levels? *Language Testing*.

Van Rossum, G. and Drake Jr, F. L. (1995). *Python reference manual*. Centrum voor Wiskunde en Informatica Amsterdam.

Vega, F. C., Nosti, M. G., Gutiérrez, A. B., and Brysbaert, M. (2011). Subtlex-esp: Spanish word frequencies based on film subtitles. *Psicológica: Revista de metodología y psicología experimental*, 32(2):133–143.

Westbury, C. F., Hollis, G., and Shaoul, C. (2007). Lingua: the language-independent neighbourhood generator of the university of alberta. *The Mental Lexicon*, 2(2):273–286.

Yarkoni, T., Balota, D., and Yap, M. (2008). Moving beyond coltheart's n: A new measure of orthographic similarity. *Psychonomic Bulletin & Review*, 15(5):971–979.

APPENDIX A

Word List

Words	Type	English	Img	Word-Freq	OLD20	ON	Non-Word	Sum-Dev
afyon	noun	opium	487	1.473	1.9	2	afcın	14
ağlaş	verb	weep	523	1.344	1.8	0	aplaz	90
akşam	noun	evening	559	341.045	1.9	2	akyem	6
aktar	verb	transfer	313	245.131	1.85	2	akrer	52
anten	noun	aerial	567	14.602	1.95	1	alken	56
atfet	verb	attribute	295	8.762	1.75	2	ağcet	10
aymaz	adj	heedless	361	5.636	1.75	4	ayriz	45
ayyaş	adj	drunkard	527	1.101	1.9	2	ahkaş	45
bakır	noun	copper	548	14.74	1.5	6	makir	70
barut	noun	gunpowder	606	8.12	1.85	2	marıt	92
başla	verb	start	359	2561.53	1.4	4	mapla	80
bayıl	verb	faint	466	43.567	1.75	3	bamul	40
beyaz	adj	white	566	299.74	1.95	1	besiz	38
bıçak	noun	knife	633	55.992	1.9	1	bıfik	79
bitki	noun	plant	605	91.256	1.8	2	bifti	21
bıyık	noun	moustache	630	16.135	1.85	3	bıvek	73
bohça	noun	package	529	2.661	2	0	biyça	8
bucak	noun	edge	495	8.696	1.75	5	buhik	54
buzul	noun	glacier	580	10.799	1.95	0	bumıl	23
çağır	verb	call	424	126.04	1.55	7	çacur	18
çalgı	noun	instrument	521	9.287	1.55	4	çansı	36
çalış	verb	work	458	3149.437	1.6	4	paluş	77
canlı	noun	alive	426	170.64	1.65	7	cerlı	52
çarık	noun	shoe	601	2.013	1.55	5	paruk	92
çatal	noun	fork	598	9.431	1.9	2	pamal	66
cebir	noun	algebra	510	5.645	1.75	4	reşir	20
çekin	verb	hesitate	354	49.552	1.4	6	penin	128
celil	adj	supreme	378	5.129	1.7	5	cerel	68
çeliş	verb	contradict	285	15.943	1.75	3	peliz	60

cenap	noun	respect	343	1.998	1.85	3	renaç	27
çeşit	noun	variety	372	463.516	1.9	1	pebit	30
çifte	adj	double	426	35.66	1.9	0	çılte	19
cihaz	noun	device	391	144.812	1.85	3	ciciz	7
cılız	adj	skinny	502	5.843	2	0	cıluş	79
çorap	noun	sock	553	19.615	1.95	1	poraç	56
coşku	noun	enthusiasm	464	38.42	2.4	0	rışku	11
cubur	dup	destitute	447	1.164	1.95	1	cuşır	19
darıl	verb	reprove	301	2.553	1.65	3	dalül	81
denet	noun	glance	395	31.272	1.5	6	betet	132
deney	noun	experiment	527	48.352	1.85	2	denah	46
denge	noun	balance	429	248.325	1.85	2	delye	71
deniz	noun	sea	606	468.25	1.75	3	dezaz	54
denli	adj	temperate	422	42.07	1.65	4	disli	59
desen	noun	pattern	453	26.379	1.8	3	beyen	152
devin	verb	move	413	2.13	1.55	5	bebin	139
devre	noun	period	492	200.238	1.55	3	dakre	85
diret	verb	insistence	324	5.471	1.6	4	boret	245
dizin	noun	index	386	24.654	1.75	1	bimin	176
dökük	adj	dilapidated	513	2.466	1.75	4	dönuk	111
donan	verb	equip	410	7.946	1.75	2	bunan	179
dönem	noun	period	429	1370.22	1.6	5	dötim	54
doruk	noun	peak	546	18.718	1.6	6	dolük	77
dövüş	verb	fight	543	5.096	1.85	0	döduş	42
düğme	noun	button	580	3.927	1.8	4	dohme	3
dümen	noun	rudder	520	7.304	1.7	6	büden	199
durum	noun	circumstance	210	1817.553	1.65	5	dulüm	83
düşük	adj	low	378	185.221	1.75	3	dübuk	75
düşür	verb	reduce	251	243.514	1.7	4	dübor	56
duvak	noun	veil	516	1.374	1.8	4	dudik	69
duygu	noun	feeling	370	259.283	1.8	2	dunyu	11
efrat	noun	individual	440	1.689	1.85	2	izrat	58
eğlen	verb	entertain	435	46.219	1.75	1	iplen	47
eklem	noun	knuckle	520	15.649	1.65	4	irlem	58
ekmek	noun	bread	619	102.658	1.9	1	eldek	45
eksiz	adj	sketch	510	0.021	1.7	4	elkiz	37
emsal	noun	similar	324	12.848	1.95	1	eynal	8
erbap	noun	expert	495	5.747	1.9	2	ersip	34
erinç	noun	ease	327	2.076	1.85	2	erest	21
erzak	noun	provisions	343	2.409	2	0	erhik	69
eşleş	verb	pair	480	6.653	1.75	2	etlez	97
esmer	noun	swarthy	608	31.881	1.9	1	ekder	47

etmen	noun	factor	269	7.613	1.8	1	ezden	37
evlat	noun	child	619	47.299	1.9	1	inlat	89
evrim	noun	evolution	402	26.943	1.75	1	enlim	57
fırın	noun	oven	599	30.48	1.85	2	cırun	87
fütur	noun	languor	359	2.037	2	0	füzır	33
gamze	noun	dimple	518	21.667	2	0	gatbe	2
gazap	noun	wrath	377	4.242	1.85	1	gamip	47
girdi	noun	input	356	15.451	1.95	0	gurmi	32
güçlü	adj	strong	463	1.599	1.8	2	gablü	13
haham	noun	rabbi	557	2.367	1.9	2	hacem	33
hakça	adv	fair	439	2.619	1.9	0	hahfa	29
halen	adv	now	276	169.545	1.9	1	haran	213
halim	adj	mild	359	4.479	1.4	10	hariş	209
hanım	noun	lady	571	158.575	1.75	5	hazum	54
harap	adj	desolate	435	5.852	1.7	4	yaraç	93
harıl	dup	glutton	548	5.462	1.5	8	halül	81
haset	noun	jealously	361	1.968	1.5	6	yayet	123
hasım	noun	enemy	497	5.3	1.45	10	hayem	38
hazan	noun	sonbahar	622	1.32	1.6	6	yadan	180
hazne	noun	store	506	2.337	1.9	0	havte	13
hedef	noun	aim	383	621.725	1.9	2	hesaf	53
helak	noun	perish	404	1.47	1.85	2	yelar	240
heves	noun	zeal	347	29.253	2	0	hesus	43
hudut	noun	border	453	8.804	1.95	1	huvıt	50
hülya	noun	delusion	396	31.932	2	0	hünka	9
hüner	noun	talent	399	3.201	2	0	hütir	52
ibret	noun	lesson	446	9.521	1.85	1	edret	58
içkin	adj	immanent	301	2.142	1.85	1	ihtin	14
iffet	noun	namuslu	366	1.455	1.95	0	ihşet	6
ihtar	noun	warn	359	11.45	1.85	3	içmar	25
ikram	noun	treating	360	23.272	1.75	5	etram	50
ikrar	noun	avowal	208	1.662	1.8	4	etrar	50
ilham	noun	inspiration	275	13.043	1.6	6	ilvem	17
incin	verb	hurt	465	4.152	1.75	3	ilgin	29
ıslah	noun	recovery	438	12.662	1.9	1	ünlah	25
ıslak	adj	wet	509	12.488	1.8	3	ünlak	25
işlev	noun	function	294	72.409	1.85	2	imlej	22
ıslık	noun	whistle	574	5.483	1.85	2	ünlık	25
itlaf	noun	destruction	505	7.16	1.9	1	imlah	35
izlem	noun	watch	525	1.965	1.75	4	etlem	78
kaban	noun	coat	572	1.635	1.2	11	kasen	190
kabus	noun	nightmare	485	24.27	2.5	0	kaşes	11

kadeh	noun	wineglass	585	14.671	1.9	2	kasih	67
kadük	adj	outmoded	478	1.521	2	0	kayuk	77
kafes	noun	cage	585	14.914	1.95	0	kahus	42
kahin	noun	prophet	467	3.303	1.9	1	kaben	113
kamış	noun	reed	520	2.367	1.6	2	kazuş	44
kapla	verb	cover	443	42.244	1.2	2	kağla	25
kasti	adj	deliberate	341	1.05	1.9	0	kaksi	55
katıl	verb	join	340	2.841	1.5	7	kamul	62
katır	noun	mule	608	36.338	1.65	5	kanir	45
kayna	verb	boil	533	32.235	1.55	2	kalba	30
kayra	noun	grace	441	1.095	1.7	0	kalta	110
kazık	noun	stick	517	8.894	1.2	12	kayek	64
kenar	noun	edge	495	113.34	2	0	keter	69
kibir	noun	conceit	402	5.588	1.8	4	kişır	44
kilim	noun	rug	591	7.793	1.75	4	kirem	79
kilit	noun	lock	532	46.135	1.9	2	kirat	121
kıpır	dup	slow	377	4.781	1.45	10	kıcur	30
kızak	noun	sledge	490	5.249	1.7	5	kımik	70
kodla	verb	code	460	7.82	1.7	1	kuçla	75
köfte	noun	meatball	618	18.415	1.95	0	köğde	22
köpük	noun	foam	600	43.567	1.75	3	köğuk	64
körel	verb	rust	547	3.351	1.9	1	kölil	68
korku	noun	fear	394	177.773	1.8	0	kontu	46
koşut	adj	parallel	324	3.549	1.5	7	kobıt	36
kubat	adj	rough	491	1.683	1.8	1	kudit	102
küçük	adj	little	502	557.439	1.9	1	küfuk	66
kurgu	noun	speculation	315	37.544	1.8	3	kırtu	47
kurul	noun	commission	481	891.033	1.5	5	kulül	81
küvet	noun	bathtub	601	3.603	1.95	1	kuyet	187
larva	noun	nymph	546	1.638	1.95	0	rarza	47
latif	adj	graceful	483	4.212	1.85	2	lanaf	36
lazım	adj	necessary	268	354.115	1.7	5	latum	31
lütuf	noun	favor	438	5.036	2.55	0	lüzıf	30
mayın	noun	mine	522	24.627	1.65	5	madon	48
medet	noun	aid	413	6.608	1.9	1	teset	142
melal	noun	boredom	406	1.812	1.5	9	telat	195
menfi	adj	negative	400	3.996	1.85	1	mevşi	20
menşe	noun	root	565	6.047	1.95	0	mevfe	23
meşru	adj	legitimate	321	40.298	1.85	2	başru	84
mesul	adj	responsible	348	3	1.75	5	meyıl	28
meyil	noun	predispozition	261	3.033	1.8	4	mevel	78
mezar	noun	grave	619	78.829	1.8	4	meser	82

milli	adj	national	400	31.683	1.9	1	mindi	35
mısra	noun	verse	489	3.654	1.95	1	momra	26
muska	noun	charm	455	2.004	2	0	munsa	27
mutat	adj	habitual	334	1.605	1.85	3	bumat	86
nakil	adj	transmission	430	8.003	1.95	1	nanel	80
nakış	noun	embroidery	478	6.371	1.6	3	şakiş	17
nalan	noun	moaner	475	2.475	1.5	8	ralak	222
necat	noun	salvation	408	1.05	1.95	1	cehat	15
nefer	noun	soldier	578	1.719	1.85	3	recer	41
nesne	noun	object	408	33.635	1.85	0	nekde	19
nezih	noun	upright	453	5.438	1.85	3	rezif	47
nikah	noun	wedding	594	37.727	2.6	0	şikoh	34
niyet	noun	intention	286	187.624	1.85	2	civet	76
nizam	noun	order	352	7.391	1.85	0	şizem	20
olgun	adj	mature	363	54.924	1.8	1	oktun	34
önlük	noun	apron	565	2.511	1.85	1	uslük	5
ördek	noun	duck	632	8.969	1.8	2	öntek	21
örtük	adj	tacit	354	1.893	1.8	0	ustük	26
örtüş	verb	overlap	399	19.498	1.75	1	örguş	16
paten	noun	skate	511	7.016	1.75	4	çanen	48
pisle	verb	dirty	485	1.143	1.95	0	ponle	14
pulla	verb	spangle	356	1.173	1.85	1	çölla	33
radde	noun	degree	521	1.02	1.85	3	cedde	11
rakam	noun	digit	489	294.013	1.85	3	cakım	55
ramak	noun	almost	305	1.923	1.85	3	catak	69
refet	verb	mercy	373	1.026	1.95	0	cecet	30
rezil	adj	vile	444	15.97	1.95	1	cezel	46
sabık	adj	previous	276	1.332	1.55	8	sağek	42
safha	noun	phase	319	25.77	1.75	3	şaşpa	30
saksı	noun	vase	563	5.387	1.9	1	sahrı	28
salak	adj	stupid	381	5.984	1	22	salen	248
saman	noun	straw	568	10.073	1.4	8	tatan	122
sarın	verb	wrap	482	1.794	1.2	10	salun	112
sataş	verb	tease	390	6.92	1.75	4	samaz	112
sayın	adj	reverend	343	412.924	1.35	8	sason	39
sazak	noun	boreas	535	1.14	1.45	11	samik	70
sebat	noun	perseverance	344	6.332	1.9	1	sedit	102
seçik	dup	sharp	495	4.997	1.9	1	sefak	70
sefer	noun	journey	520	153.185	1.6	7	sepir	61
sefil	adj	miserable	429	6.833	1.75	4	secel	34
şehir	noun	city	605	340.14	1.75	5	necir	25
semer	noun	saddle	578	2.214	1.55	6	seyar	75

serin	adj	chilly	460	17.227	1.45	3	selın	137
seyir	noun	course	391	62.495	1.55	6	sesur	49
sıçra	verb	jump	506	31.392	1.95	0	sigra	3
sıkış	verb	representative	380	84.153	1.5	2	sıkiz	54
simge	noun	symbol	447	42.991	1.95	0	sişye	6
sirke	noun	vinegar	562	10.586	1.85	1	sürte	17
şişir	verb	blow	458	8.93	1.8	0	nibir	18
sıtma	noun	malaria	484	4.994	1.85	3	süşma	3
sonra	postp	after	217	1780.309	1.75	4	sirra	30
sonuç	noun	result	324	1441.003	2	0	sonıp	16
soyla	verb	trace	384	1.722	1.6	2	simla	19
soyun	verb	strip	562	46.585	1.65	4	soson	53
sümük	noun	mucus	570	1.569	1.75	3	süzuk	52
süper	adj	foundation	429	216.349	1.9	1	süğir	57
suret	noun	appearence	233	120.794	1.85	1	toret	169
tabir	noun	remark	321	34.889	1.6	7	taşer	45
takat	noun	strength	470	1.857	1.55	6	makit	92
takva	noun	piety	330	3.789	1.7	3	tenva	7
tavuk	noun	chicken	619	60.581	1.8	4	tadük	51
tekin	adj	deserted	395	1.626	1.1	5	mekın	155
tente	noun	sunshade	592	3.129	1.95	1	takte	29
tonoz	noun	vault	550	1.377	2	0	totüz	60
tüfek	noun	gun	613	28.494	1.9	2	tücık	30
tüken	verb	exhaust	520	40.289	1.85	2	münen	131
türkü	noun	ballad	578	59.208	1.9	1	türlo	17
tutar	noun	amount	316	181.598	1.8	4	mumar	94
tüzel	adj	judicial	310	36.338	1.65	5	tümil	30
tüzük	noun	regulations	345	47.572	1.85	3	mümük	76
ümmet	noun	community	416	7.049	2	0	üzdet	38
ürkek	adj	timid	404	6.617	1.95	1	üntek	39
vapur	noun	steamboat	631	26.589	1.85	3	vağır	28
vasat	noun	mediocre	322	10.007	1.7	4	zayat	56
velet	noun	kid	525	1.155	1.85	2	zeret	198
vergi	noun	tax	446	729.234	1.65	5	zargi	29
yahni	noun	stew	587	5.213	1.9	1	yakdi	19
yakış	verb	suit	536	97.388	1.2	8	yakiz	54
yalan	noun	lie	385	125.414	1	15	halak	247
yamuk	noun	askew	428	1.656	1.8	3	yatük	81
yanay	noun	profile	572	5.375	1.6	6	yatey	42
yapay	adj	artificial	386	26.973	1.85	3	yacey	45
yarat	verb	invention	408	656.69	1.65	3	yalit	121
yarık	adj	fissure	381	1.443	1.25	13	yaluk	139

yaşıt	noun	peer	376	7.853	1.6	6	yabut	34
yazgı	noun	fate	343	4.449	1.8	3	yavrı	33
yengi	noun	victory	461	3.537	1.75	4	genci	73
yolcu	noun	passenger	529	146.978	1.85	3	yoygu	14
yükün	noun	ion	348	1.908	1.75	2	hükun	111
yürek	noun	heart	617	123.005	1.85	3	yülık	92
zabit	noun	officer	593	1.011	1.85	3	vaşit	38
zalim	adj	tyrant	494	11.306	1.8	3	zarem	79
zinde	adj	energetic	422	4.167	2	0	ziste	34
zorla	verb	force	437	242.245	1.7	3	zekla	48
zulüm	noun	cruelty	422	19.105	2.45	0	zurum	83

APPENDIX B

Pre-Experiment Participant Form

(TR: Deneye başlamadan önce, katılımcıdan deneyin sonuçları için gerekli olabilecek bilgileri öğrenmek için hazırlanmıştır.)

(ENG: This document was prepared to obtain relevant information from the subjects which might be necessary for the results).

Tarih (Date): ____________

Katılımcının (Participant's);

İsim Soyisim (Name Surname): ________________________

Yaş (Age): ________________________

Kaç tane dil konuşabiliyorsunuz? ________________________

(How many languages do you speak?)

Bu dile ne kadar hakimsiniz? (Lütfen derece skalasında değerlendiriniz.)

What is your proficiency level on these languages? (Please rate according to the scale.)

1. ________________________ Seviye (Level): 1-2-3-4-5-6-7-8-9-10
2. ________________________ Seviye (Level): 1-2-3-4-5-6-7-8-9-10
3. ________________________ Seviye (Level): 1-2-3-4-5-6-7-8-9-10

Ayda ortalama kaç tane kitap okuyorsunuz? ________________________

(How many books do you read per-month?)

Yazı yazarken hangi elinizi kullanıyorsunuz? Sağ / Sol

(Which hand do you use while writing? Right / Left)

***English translations were not included in the original document.

TEZ FOTOKOPİSİ İZİN FORMU

<u>ENSTİTÜ</u>

Fen Bilimleri Enstitüsü	☐
Sosyal Bilimler Enstitüsü	☐
Uygulamalı Matematik Enstitüsü	☐
Enformatik Enstitüsü	X
Deniz Bilimleri Enstitüsü	☐

<u>YAZARIN</u>

Soyadı : Erten
Adı : Begüm
Bolümü: Bilişsel Bilimler

<u>TEZİN ADI</u>: Adapting and testing psycholinguistic toolboxes for Turkish visual word recognition studies

<u>TEZİN TÜRÜ</u>: Yüksek Lisans X Doktora ☐

1. Tezimin tamamından kaynak gösterilmek şartiyla fotokopi alınabilir. ☐
2. Tezimin içindekiler sayfası, özet, indeks sayfalarından ve/veya bir bölümünden kaynak gösterilmek şartiıyla fotokopi alınabilir. ☐
3. Tezimden bir (1) yıl süreyle fotokopi alınamaz. X

<u>TEZİN KÜTÜPHANEYE TESLİM TARİHİ</u>: 2 Ekim 2013