# EYE MOVEMENT CONTROL IN PERSIAN READING: A CORPUS-ANALYTIC APPROACH

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

FATEMEH SOLEYMANI

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

SEPTEMBER 2018

**EYE MOVEMENT CONTROL IN PERSIAN READING:**
**A CORPUS-ANALYTIC APPROACH**

Submitted by **FATEMEH SOLEYMANI** in partial fulfillment of the requirements for the degree of **Master of Science in Cognitive Science Department, Middle East Technical University** by,

Prof. Dr. Deniz Zeyrek Bozşahin
Dean, **Graduate School of Informatics** _____

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

Assoc. Prof. Dr. Cengiz Acartürk
Supervisor, **Cognitive Science Dept., METU** _____

**Examining Committee Members:**

Prof. Dr. Cem Bozşahin
Cognitive Science Dept., METU _____

Assoc. Prof. Dr. Cengiz Acartürk
Cognitive Science Dept., METU _____

Assoc. Prof. Dr. Bilal Kırkıcı
Foreign Language Education Dept., METU _____

Asst. Prof. Dr. Umut Özge
Cognitive Science Dept., METU _____

Asst. Prof. Dr. Mehdi Purmohammad
Psycholinguistics., Institute for Cognitive Science Studies _____

**Date:** ***September 6, 2018***

**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: Fatemeh Soleymani**

**Signature : _____**

## ABSTRACT

EYE MOVEMENT CONTROL IN PERSIAN:
A CORPUS-ANALYTIC APPROACH

Soleymani, Fatemeh
M.Ss. Cognitive Science
Supervisor: Assoc. Prof. Dr. Cengiz Acartürk

September 2018, 82 pages

In the Latin script languages, specifically the English language, there exist studies on the eye movements' pattern during reading. Thus, this study was conducted in order to investigate eye movement patterns in Persian. From these sentences, taken from the Bijankhan Persian Corpus, a number of eye movement measures were analyzed. The eye movement measures were first fixation landing position, first fixation duration, gaze duration, first run fixation count, regression in count, and regression out count. The word properties in this study were word length, word frequency, word predictability, word type (opacity and transparency), and phonemes. In order to control variances of random effects which were subject, sentence, and words, the linear mixed model were utilized in the analysis.

The results show that word length has an effect on first fixation landing position. Furthermore, the preferred viewing location (PVL) analysis showed that first fixations landed close to the end of short words, close to the center of medium-long words and close to the beginning of long words. These results are in line with former studies. In the first fixation duration analysis it was found that medium-long words were fixated shorter than short words. Moreover, it was found that longer gaze durations (GD) and higher first run fixation count (FRFC) corresponded with longer words. Also, a relationship between regression in count (RIC) and phonemes was obtained.



Key Words: Persian Reading, Eye movements, First Fixation Landing Position, First Fixation Duration, Linear Mixed Model

# ÖZ

FARSÇA OKUMADA GÖZ HAREKETLERİ KONTROLÜ:
DERLEM-ANALİTİK BİR YAKLAŞIM

Soleymani, Fatemeh
M.Sc. Bilişsel Bilimler
Danışman: Doç. Dr. Cengiz Acartürk

Eylül 2018, 82 pages

Latin dilleri ve özellikle İngilizce için okuma sırasında göz hareketi örüntülerini araştırma çalışmalar bulunmaktadır. Bu çalışmanın amacı Farsça okumada göz hareketlerinin araştırılmasıdır. Bijankhan Farsça derleminden seçilen cümleler üzerinde bir dizi göz hareketi metriği analiz edilmiştir. Bu metrikler arasında ilk sabitmele konumu, ilk sabitleme süresi, toplam bakış süresi, ilk geçiş sabitleme sayısı, içeri geridönüş sayısı, dışarı geridönüş sayısı bulunmaktadır. Bu çalışmadaki kelime özellikleri ise kelime uzunluğu, kelime sıklığı, kelime tahmin edilebilirliği, kelime türü (opasite ve saydamlık), ve fonemlerdir. Katılımcı, cümle, kelime gibi rastgele istatistiksel etkilerdeki varyansı kontrol edebilmek amacıyla analizlerde LMM (Linear Mixed Models) kullanılmıştır.

Sonuçları kelime uzunluğunun ilk sabitleme konumu üzerine etkisi olduğunu göstermektedir. Bunun yanında, tercih edilen gözleme konumu (PVL) analizleri ilk sabitlemelerin kısa kelimelerin sonuna, orta uzunluktaki kelimelerin ortasına ve uzun kelimelerin başına olduğunu ortaya koymaktadır. Bu sonuçlar literatürde raporlanan sonuçlar ile uyumludur. İlk odaklanma süresi analizinde orta uzunluktaki kelimelere kısa kelimelerden daha kısa süreyle sabitlendiği bulunmuştur. Bunun yanında, uzun gakış sürelerinin (GD) ve daha yüksek ilk geçiş sabitleme sayılarının (FRFC) uzun kelimelerle ilişkilendirilebileceği gözlenmiştir. İçeri geri dönüşler ile fonemler arasında da bir ilişki gözlenmiştir.

Anahtar Kelimeler: Farsça Okuma, Göz Hareketleri, İlk Sabitleme Konumu, İlk Sabitleme Süresi, Linear Mixed Model

*To my beloved family*

## ACKNOWLEDGMENTS

It would be impossible for me to fully thank and show my appreciation to everyone who helped me with my thesis. This was a huge and daunting undertaking that was made possible by the love and support of many individuals in my life. I cannot personally thank them all, but I would like to highlight the efforts of individuals whom I feel went above and beyond the call of duty to get me to where I am today.

I would first like to thank my supervisor Assoc. Prof. Dr. Cengiz Acartürk. I appreciate him taking me under his wing and providing me with support and guidance through this process. His critical eye and insight provided the cornerstone and foundation for my thesis. I am proud and honored to have worked with him on this project. I would also like to thank him for his patience as I found my way through my thesis.

Furthermore, I would also like to say thank you to my committee members, Prof. Dr. Cem Bozşahin, Assoc. Prof. Dr. Bilal Kırkıcı, Asst. Prof. Dr. Umut Özge, and Asst. Prof. Dr. Mehdi Purmohammad. I want to show them my gratitude for serving on my committee and providing me with additional feedback and insight on my thesis.

I want to express my deep gratitude and warmest thanks to Ayşegül Özkan. She made herself available for anything I needed, big or small, night or day. Her mind is like the Library of Alexandria with skills and knowledge that she was more than willing to share with me. With her help I was capable of understanding concepts that I initially feared I would never be able to understand. I am unsure how I could ever payback her dedicated support, but I hope to.

I would like to thank and show my love and appreciation to my parents Sima and Eynollah for their love and support, their love is a shining beacon of hope and serves as my guiding light and inspiration to continue to push myself harder and forward in life. It has inspired me and emboldened me, provided warmth when I was cold, brightened my day when there were clouds in the sky, and sat beside me when I was lonely. There were many nights where I missed them and my home dearly but knowing they are proud of me helped to keep moving me forward. I always keep a little piece of them, and my home, in my heart so that neither are never far from me, no matter where I go. I hope to continue making them proud of me as I take my next few steps in my passions and journey. With the love and support I receive from my parents; all things are possible.

I would also like to thank my siblings Leila, Reza, and Alireza for both their love and continued emotional support through this process. Words alone cannot describe my appreciation for standing by my side. You all know that I treasure you deeply and would do everything in my power to equally support you in your own endeavors through life. In particular, I would like to thank Reza and Sahar. I wouldn’t have made it this far without the full love and support of my dear dear family.

Additionally, I would like to thank all of my friends. I especially want to say thank you to Anders Patrik Segerfeldt, Sepehr Seyedian Choubi, and Tuğçe Nur Bozkurt for their love and dedication, and emotional support. I also want to thank Mehmet Yolcu, Mahyar Azizi, and Shima Ebrahimi for the merriment they brought to my life as I worked on my thesis. Their support have helped guide me through my thesis work. All my friends enrich my life. They are dear to me and I am thankful that they took this journey with me.

Finally, I am fortunate to have a friend such as TiMar Long in my life. Words alone do not suffice to thank him for the value he has brought into my life. He stood by me and never gave up on me. Whenever I was unsure about finishing he stepped in and helped to inspire and focus me so that I could see my thesis to the end. I couldn't begin to thank him for his unconditional support he was willing to provide me with. I treasure our friendship.

## TABLE OF CONTENTS

**LIST OF TABLES**

**LIST OF FIGURES**

## LIST OF ABBREVIATIONS

| Abbreviation | Meaning |
|---|---|
| FLP | First Fixation Landing Position |
| FFD | First Fixation Duration |
| GD | Gaze Duration |
| FRFC | First Run Fixation Count |
| RIC | Regression in Count |
| ROC | Regression out Count |
| WL | Word Length |
| WF | Word Frequency |
| WP | Word Predictability |
| WT | Word Type |
| PHON | Phoneme |
| LMM | Linear Mixed Model |
| PVL | Preferred Viewing Location |
| OVP | Optimal Viewing Position |
| IOVP | Inverted Optimal Viewing Position |
| RL | Right to Left |
| LR | Left to Right |
| IAUPS | Islamic Azad University's Pharmaceutical Sciences |
| POS | Part of Speech |
| EZ | Ezafeh Suffix |
| N | Noun |
| AP | Adjective Phrase |
| PP | Prepositional Phrase |
| NP | Noun Phrase |

# CHAPTER 1

# 1 Introduction

## 1.1 Reading

Post publication of Charles Darwin´s manuscript On The Origin of Species (1859), apes has in a humorous way been caricatured to be reading this very famous book (Kutschera, 2009). However, the human phenomenon of extracting information from printed symbols, aka. reading, is not merely old but a complex process as well. It involves basically three types of processes, which are called perceptual, attentional, and oculomotor (Kliegl et al., 2003) which guides readers in the extraction of information, either for pleasure or gaining knowledge (Wotschack, 2009).

It is generally viewed that there are five elements of knowledge and skills, which need to be mastered for a person to successfully extract information from a text. These are: 1) Phonemic awareness - which is the capacity of hearing, identifying and manipulating various types of sounds (phonemes) as building block of spoken words, 2) Phonics – the correlation between the spelling in written words and their spoken counter parts, aka. how graphemes correlate to phonemes, 3) Vocabulary - knowledge of the glossary (or words) which are used in language, hence both oral vocabulary (spoken or listened to words) as well as written vocabulary (words printed or recognized in text), 4) Fluency - the capability to quickly and correctly read through a text, and finally 5) Comprehension – to be able to comprehend or understand, not only individual words, but the greater content of text itself. Thus, to understand a written language, all these elements must be learned (Armbruster, H., Bikfalvi, A., Kinkel, S., & Lay, G., 2008).

Moreover, a noteworthy aspect of information extraction through reading is that despite a person reading has many correlative processes with a person listening, the reader unlike the listener has the capability to adjust the speed of information input to align it with the processes of his/her internal comprehension. In contrast to the listener the reader is able to pause, skip segments of text or re-read what he or she previously has read. These processes of internal comprehension may be studied by investigating the occurrence of these pauses which the reader makes while reading (Just, M. A. & Carpenter, P. A., 1980). Hence, studying eye movements in silent reading is a not only possible, however, but the preferred approach to investigating the processes of internal comprehension (Rayner, K., Pollatsek, A., Ashby, J., & Clifton Jr, C., 2012). Moreover, usual eye movement patterns can be gained from eye movement studies on reading where these patterns are connected to the comprehension of what is read (Hayhoe, 2004). In the following section I will outline the goals of this study. Afterwards, I will discuss my hypothesis.

## 1.1 Aim of the Study

By examining various types of eye movement measures (variables properties) with respect to their occurrence during silent reading of Persian sentences, and in relation to a number of word (fixed) properties, this thesis study aims to investigate eye movement patterns that occur during the course of silent reading Persian sentences. The eye movement measures and the word properties used in this study are listed below:

***Eye movement measures***

- First Fixation Duration (FFD): The duration of the first fixation in the current interest area.
- Gaze Duration (GD): The total duration of all fixations on the current interest area.
- First Run Fixation Count (FRFC): The total number of fixations in the first run of the current interest area in a trial.
- Regression-in Count (RIC): Number of times the interest area was entered from the left in Persian.
- Regressions-out Count (ROC): Number of times the interest area was exited to the right in Persian, before the upcoming fixated word in the trial.
- Fixation Landing Position (FLP): The character of the word on which the first fixation is made.

***Word properties***

- Word length: The length of a word with the following parameters:
    - The number characters in a word by including zero-width space as one independent character.
    - The number characters in a word by excluding zero-width space as an independent character.
    - The sum of pixels by using a mono-spaced font where each character is set to 14.03 pixels.
- Word Type: Phonologically transparency/opacity examines whether a word's spelling corresponds to its sound or not, respectively (Baluch, B. & Besner, D., 1991).
    - Opacity/transparency of the word by excluding Ezafeh affix as a single phoneme.
    - Opacity/transparency by including Ezafeh affix as a single phoneme.

- Word Frequency: The number of times whole-word occurs in the Bijankhan Corpus (Amiri, H., Hojjati, H., & Oroumchian, F., 2007)

- Phonemes: The number of phonemes including by Ezafeh affix as a single phoneme.

- Word predictability: Predictability of a word in a sentence by seeing the preceding part of the sentence (Taylor, 1953), and measured the fraction of the participants who accurately predicted the whole-word in question (Nilsson, 2012).

What the general aim of this study encompass is described above, in the following section the aspects of hypotheses, motivations and expected results will be described.

## 1.2 Hypotheses, Motivation and Expected Results

### *1.2.1 Hypotheses*

In this thesis, the central question around which the research was built is as follows: How do word length, word frequency, word type, word predictability, and number of phonemes affect eye movement measures such as first fixation duration, first run fixation count, gaze duration, first fixation landing position, regression-in, and regression-out counts during silent reading of Persian sentences?

On this follows a set of hypotheses which are assumed to be observed during reading of Persian sentences:

**i.** The initial hypothesis states that word length has an effect on first fixation landing positions, first fixation durations and gaze durations. It is also expected that these first fixations will have a landing position near the centers of short to average words, except for long words. It is hypothesized that the gaze duration will increase as the word length increases, due to that the total number of fixations would increase. Also, regression-ins will be expected to increase with longer words, as suggested by the previous literature on reading in other languages.

**ii.** The second hypothesis assumes that a word's opacity and transparency will affect both the first fixation duration as well as the gaze duration. It is further expected that since as a word is opaque, including the Persian Ezafeh affix (addition, adjunction), a vowel that can be found in the space after the word but is not represented (Samvelian, 2007), regression in count happens to the earlier word due to the Persian Ezafeh affix..

**iii.** The third hypothesis considers the effect of word frequency on first fixation landing position, first fixation duration, and the gaze duration. We assume that the fixation durations will increase on low frequent words in comparison to high frequent words. It is also expected that low frequency words will tend to have more regression-ins, as also suggested by the previous literature on reading in other languages.

**iv.** The final hypothesis considers the effect of word predictability. We expect to observe the effect of word predictability on first fixation durations, gaze duration, first run fixation count, as well as regressions-in count. Furthermore, highly predictable words will tend to have less first run fixation count as well as shorter first fixation duration.

In the subsequent section we will shortly state the general motivation for this study.

### *1.2.2 Motivation*

In the Latin script languages and specifically the English language there are studies on eye movement patterns in reading. Various word properties and eye movement measures have been investigated in different combinations, with different aims in understanding the underlying cognitive processes. However, there is a lack of studies on eye movements in the reading of Persian script-based languages such as Farsi, Dari and Baluchi. With this thesis we aim to fill this gap by investigating the previously mentioned eye movement measures in relation to the word properties in question based on the chosen research question. Moreover, we aim to investigate the effects of some word properties specific to the Persian language, such as zero-width space and Persian Ezafeh suffix in order to see how they affect various eye movement measures. While the motivation stated above describes the reason for conducting this study, the following general results are assumed to be obtained.

### *1.2.3 Expected Results*

Nevertheless, in a general sense it is reasonable to find certain intertwining relations between the eye movement measures and the word properties studied. Moreover, we may expect similar results as previous and comparable eye movement reading studies performed on various languages such as Uighur language, which uses Arabic alphabet as Persian language (Yan et al, 2014). However, since the Uighur language is a Turkic language and the Persian language is an Indo-Iranian language with certain morphological differences, this may cause significant differences in the results.

# CHAPTER 2

# 2 Background

## 2.1 History of reading

### *2.1.1 Early texts and writings*

One of the oldest writings, which to some extent could be viewed as writing, were the cave paintings found in caves of southern Europe, such as the Lascaux caves. These are estimated to be around 20,000 years old (Valladas et al., 2001; Rayner et al., 2012). However, the earliest types of writings, or what we would now view as writings, are much younger and are estimated to be up to 5,500 years old and are from Mesopotamia. Some of the oldest scripts of this time was the Sumerian proto cuneiform (arrow head), Sumerian cuneiform, and slightly later on Akkadian cuneiform (Jagersma, 2010; Rayner et al., 2012). Later in history, the Persian empires also used the cuneiform script (Rayner et al., 2012). A more recent one is the Persian script, which is used in modern Persia (Iran). In today's world, the Latin script and the Persian script are both commonly used scripts for reading and writing. Moreover, Persian script is composed of Arabic letters but also have alterations in some of the letters and addition of some more recent letters (Izadi, S., Sadri, J., Solimanpour, F., & Suen, C. Y., 2008).

A short description of the history texts and writings was given above, and in the subsequent sections a short introduction to early studies on the human eye is given.

### *2.1.2 Early studies on human eye and its functions*

The human eye is the major anatomical tool for visual information extraction from text. The anatomy of the eyes and its functions have, for a long time been, investigated in various levels by philosophers and scientists. Aristotle and Galen were some of the first philosophers and physicians to study and examine the eye. While the latter was more interested in the anatomy of the eyes, Aristotle focused more on gaining a basic knowledge of eye's functions, which was further elaborated upon by Ptolemy, Ibn al-Haytham, Bacon and many others (Wade, 2010).

In the previous section I introduced the early period of studies on eyes and their function. In the next section I introduce eye movement studies conducted post Middle Ages.

### *2.1.3 Eye movement studies in the post middle ages*

In the post middle ages and during the Renaissance, the number of scientists who studied how eyes work increased. Jan Purkinje, William Charles Wells and Johannes Müller were among those scientists who studied eye anatomy and its functions in early times (Wade, 2010).

In the 19th century, Wells studied voluntary versus involuntary eye movements by using afterimages. The research of Charles Bell drew the following conclusions (cited in Wade, 2010):

> When an object is seen, we enjoy two senses; there is an impression upon the retina; but we receive also the idea of position or relation which it is not the office of the retina to give. It is by the consciousness of the degree of effort put upon the voluntary muscles, that we know the relative position of an object to ourselves... (Bell, 1823)

In essence, Bell found that eyes do not only view objects but also the object's position relative to the eyes of the observer. An early 20th century psychologist, Leonard Carmichael, named Charles Bell the pioneer of eye movement research saying (cited in Wade, 2010):

> He made contributions to the study of the organs of sense, and did work upon the reciprocal innervation of antagonistic muscles. In the scientific study of the expressions of the emotions Bell's treatise is a classic which, according to Darwin, laid the foundations of the subject as a branch of science. (Carmichael, 1926)

Nevertheless, the topic of who pioneered eye movement research has been a controversial one with no consensus on who it was. Other researchers, such as Edwin G. Boring stated that the founder of eye movement studies was Johannes Müller, crediting Müller's book (monograph) *On Imaginary Apparitions*, published in 1826 (cited in Wade, 2010).

Recently, William Charlse Wells has been considered the founder of eye movement research (Wade, N. & Tatler, B. W., 2011). They base this claim on Wells' investigations on induced vertigo in healthy subjects while using the means of afterimages as an index. These afterimages served as a way for the retina to gain stability and to discern between the movement of the eyes versus the environment (Also see Tatler, B. W. & Wade, N. J., 2003; Wade, N. & Tatler, B. W., 2005).

Another eye movement researcher from the 19th century was Jan Evangelista Purkinje. He used the light reflected from the eye as a basic method for eye tracking.

For this, he utilized the natural ability of the cornea and lens to reflect visible light to accurately detect the position of the eye fixations. The image reflection on the eye was named the Purkinje Image in honour of him (Wade, 2010).

Figure 2.1 A spinning cylinder cage (or centrifuge)

used by J. E. Purkinje, and a Purkinje portrait. (Adjusted from Figure 10, Wade, 2010).

Purkinje also used a device which was invented by Erasmus Darwin (grandfather of Charles Darwin). E. Darwin developed a spinning cylinder cage in an attempt to treat mental illnesses (see Figure 2 .1). Purkinje used the spinning cylinder cage to induce vertigo in his subjects in order to study involuntary eye movements such as nystagmus (cited in Wade, 2010). The spinning cylinder was later utilized by the Austrian otologist Robert Bárány (cf. Bárány´s chair) for his study of the vestibular system and its correlation with post-rotational nystagmus. Nevertheless, it is generally accepted by scientists in this field that Jan Evangelista Purkinje is the founding researcher who studied the connection between the vestibular system and eye movements (Wade, 2010).

In the previous section I discussed the various aspects of the history of reading and introduced the field of eye movement research. In the following section I will describe the early period of eye movement studies with the focus on reading.

## 2.2 Initial Research on eye movements in reading

Late into the 19th century, a novel interest amongst many scientists formed. This interest focused not only on investigating the anatomy of the eyes, or the eye movements in general, but also studying how the eyes move while reading. For instance, the physician William Porterfield, the ophthalmologist Louis Émile Javal,

and the physiologist Karl Ewald Konstantin. Hering was among the researchers who investigated eye movements in reading during the 19th century (Wade, 2010). A notable researcher was L. E. Javal, who coined the phenomena of "saccades" in 1879 to describe an eye movement in which the eyes simultaneously moved between multiple phases of fixation while looking in the same direction (Wade, 2010).

Another scientist who studied eye movements that occurred while reading was Karl Ewald Konstantin Hering. Hering noticed that eye movements were not continuous but instead jump from one point to another point of fixation (Hering, 1942, a translation of Hering, 1879 into English by Carl A. Radde, cited in Wade, 2010).

This was supported by Edward B. Huey who, 30 years later, stated:

> As you watch the reading [of another person], you notice, too, that the eyes do not move continuously from left to right along the line, but proceed by a succession of quick, short movements to the end, then return in one quick, usually unbroken movement to the left. (Huey, 1908)

Hering later continued his research on eye movements by listening to sounds in the eyes by using a rubber tube put on the eye lid. He then drew the conclusion, after conducting similar experiments on curare-treated animals that the sounds that he heard were not made by blood flowing in veins but rather actions of muscles in the eye. He then, as a subsequent step preformed the same experiment for research on open eyes combined with the use of afterimages while reading. He found a relationship between short movements of the eyes and the sounds of the muscles he heard (Wade, 2009, 2010).

> One can observe the clapping sounds very clearly during reading. Although the eyes appear to glide steadily along the line, the clapping sounds disclose the jerky movement of the eyeball (Hering, 1879b, cited in Wade, 2010)

After Hering, Lamare reported that the eyes do a saccade once every 15$^{th}$ to 18$^{th}$ symbol in a row of text. Hence, Lamare and Hering may jointly be attributed to that the eyes do not move uninterrupted or smoothly across a sequence of words on a page, but rather makes leaps while reading (Wade, N. J., & Tatler, B. W., 2009; Wade, 2010).

Before proceeding with the development in research on eye movements in reading in the 20th Century, we first present basic characteristics of eye movements in the following section.

## 2.3 Characteristics of eye movements in reading

### *2.3.1 General characteristics - Eye movements and visual field*

Eye movements may be divided into four main categories based on their roles in the human visual sensory system:

- **Saccades** are the most commonly studied eye movement, a ballistic type that changes the fixation point rapidly.
- **Smooth pursuit movements** aim to slowly change the focus of the fovea in a gradual manner.
- **Vergence movements** are utilized when focus is needed on an object further away, by aligning the fovea accordingly.
- **Vestibulo-ocular movements** are needed to reimburse movements of the head, and there are stabilizing the eyes versus the external world (Purves et al., 2004).

Anatomically, there exist three major components of the eye: The fovea, the parafovea and the peripheral fields (Candy, T. R., Crowell, J. A., & Banks, M. S., 1998; Rayner et al., 2012; Moll, K., & Jones, M., 2013). Hence, they correspond with various spans of the visual field. First, the fovea covers a span of 2 degrees of the visual field relative to the point of fixation, while secondly the parafovea spans 10 degrees around the fixation point and from there the peripheral field starts (Rayner et al., 2012) (see Figure 2 .2).

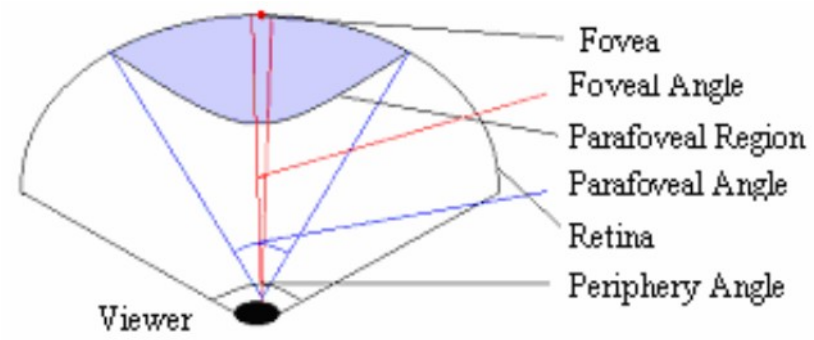


Figure 2.2 The boundaries of acuity in the visual field with respective angles

(from figure 2 in Biswas et al., 2008).

The retina consists of two specific types of receptors that are sensitive to light stimuli. These are cones and rods, where the former is responsible for discriminating between different colors (different wave length in the visual light spectra) as well as acuity or details, while the latter determines various shades of grey (white to black), detection of various levels of brightness as well as detection of movement. The fovea, which is the center of the visual span, is composed almost entirely of cones

that aids in the focus on words or parts of words, which demands detail. However, the concentration or density of cones decreases the further from the center of the fovea one looks, with a medium concentration in the parafoveal area and low concentration in the peripheral area. The nature of the concentration of rods is almost completely the opposite, as can be viewed in Figure 2 .3 (Rayner et al., 2012).

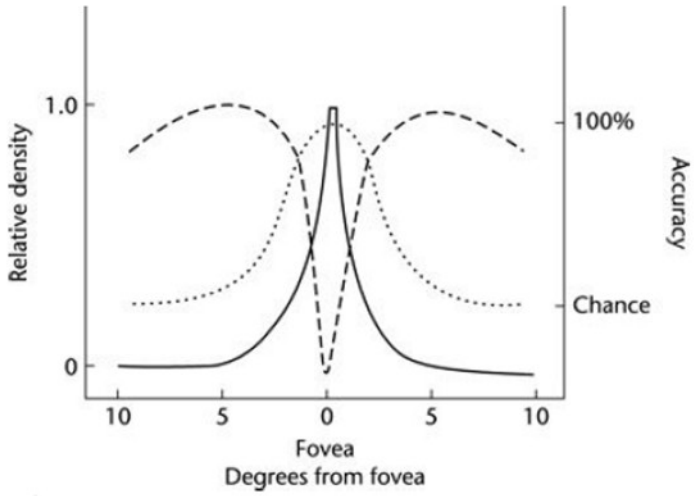


Figure 2.3. A displaying of the relative visual field density

as solid line (for cones) and as dashed line (for rods), and with the dotted line as the percent of accuracy in the visual field, with the highest at the fovea center (from figure 1.2 in Rayner et al., 2012).

We previously described some general and important characteristics of eye movement in reading research. Below we will describe the two most important eye movement measures in reading, fixations and saccades.

### *2.3.2 Fixations and saccades*

When measuring eye movements in reading, two main eye movement measures are of interest for systematic investigation. The first is when the eyes are resting or fixating at a point in the text, this is called fixations. Fixations may on average occur between 200-350 ms (milliseconds) but may range all from 100 ms to up to 500 ms or longer. These fixations are of importance since information is extracted from the text during fixations. They are then subsequently ended by a specific type of ballistic jumps, called saccades. Saccades are what move the eye focus from one point of fixation to the next point of fixation. Saccades are much faster than fixations, and they are generally as fast as 15 ms up to 40 ms (Reichle, E. D., Pollatsek, A., Fisher, D. L., & Rayner, K., 1998).

There are further eye movement measures that are important in reading research. For instance, a return sweep is a combination of a long saccade and a short correcting saccade, that is made when the focus changes from the end line of a text (right side of a paper in Latin scripts and vice versa in Persian scripts) to the beginning of the next or lower line of text (left side of a paper in Latin scripts and vice versa in Persian scripts). Another eye movement measure is regression. Regressions are

backward saccades to previously fixated words. Finally, skipping describes not conducting a fixation on a word that is usually adjacent to an already fixated word (Reichle et al., 1998; Izadi, 2008) (see Figure 2 .4).

In Figure 2 .4a series of measures are shown, where a number of adjacent words (n, n+1 & n+2) are viewed. How the reader saccades its point of fixation (a ring with a number) in various ways is described with bold arrows: A first saccade is made to fixation number one, which is left of the center of word n, followed by a refixation to the right of the center. This is followed by a skipping to word n +2 to fixation number 3 where the word n+1 is skipped. A subsequent regression to fixate (number 4) at the left of the center of the skipped word is made, followed by a longer saccade to fixate at the right of the center of word n+2.

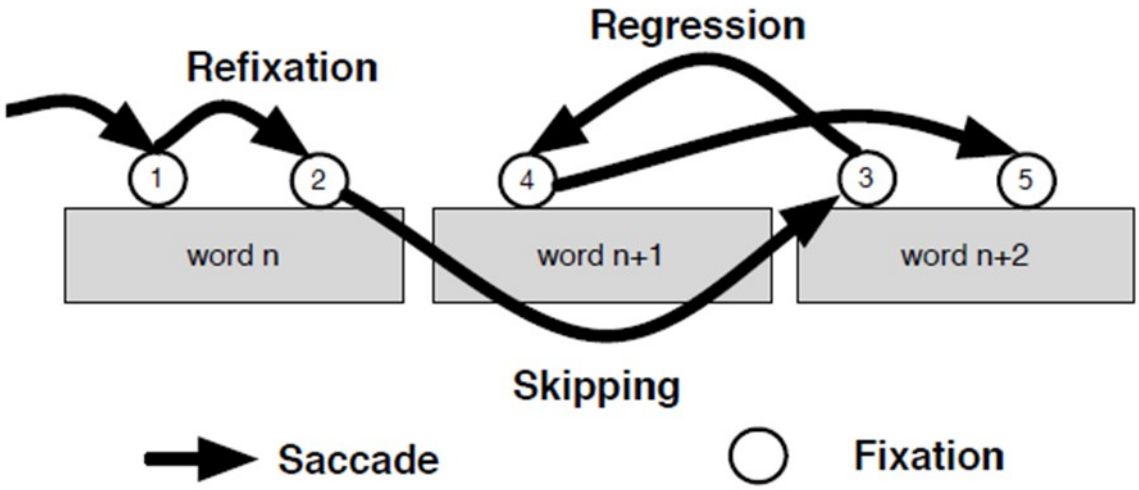


Figure 2.4 A picture showing some of the most common characteristics

measured in eye movement reading research (Adjusted from figure 2.1 in Wotschack, 2009).

In the previous section various eye movement measures were briefly described. In eye movement research related to these eye movement measures two basic research problems which are where the eye will move and when the eyes will move. These issues are explained in concisely below.

### *2.3.3 Where to move the eyes*

According to Clark and O´regan (1999; also see O´regan et al., 1984, 1987), there is a position in words that is ideal to fixate on for easier processing of the word. This point of fixation is called Optimal Viewing Position (OVP). According to Clark, O'regan and colleagues, the OVP is very close to the center of the word. They reported that if the initial fixation was at the Optimal Viewing Position, not only the probability of refixation on the word but also gaze duration and the total fixation time on the word decreased (Clarke et al, 1999). A contrasting finding to the Optimal Viewing Position, was reported by Rayner (1979). He found that during reading, the fixation which most often occurred was at a point left of the OVP. When this point was fixed it reduced the recognition time of the word in question. This position was

named the Preferred Viewing Location (PVL) (Reichle, E. D., Rayner, K., & Pollatsek, A., 2003).

In summary, according to previous research, there is a locational difference between where it is optimal to fixate on in a word and where it is preferred to. As part of the aim of this investigation, in the result section we will describe where the fixation occurs for readers of Persian text. In the next section the second research issue of when to move the eyes is described.

### *2.3.4 When to move the eyes*

Reichle et al. (1998) proposed that various factors influence eye movement patterns in reading. One of them is word predictability (i.e., predictability of a word in a sentence, aka. sentential predictability). Highly predictable words attract not only shorter gaze duration but also first fixation duration. It was also found that in general, words with higher frequencies are fixated at shorter times. Lexical, syntactic, and semantic characteristics also have various effects on the duration a word is fixated. However, these higher-level influences are usually conceived as beyond the scope at the current state of reading research.

As described there are various fixed effects affecting the question of when to move the eyes. How this occurs for Persian text reading will be discussed in the result section as part of the goal of this study. In the subsequent section I will give background on how eye movements in reading have been measured historically.

## 2.4 How to measure eye movements during reading

Huey (1908) adopted a crude mechanical technique of measuring eye movements, which was first developed by E. Raehlman for investigations of nystagmus and by A. Ahrens to investigate the eyes while reading. Huey used this technique himself, however he found that it had shortcomings despite getting the same results as the previous two researchers. Upon inspection he improved not only the method but also controlled the recording better with an individual/exchangeable biting-bar as well as with a clamp or fixation holder, in order to keep the head still. This, along with a calibration run improved the technique. However, the technique was still crude and put pressure on the eye lid which still limited its utilization as eye movement measuring in reading (Wade, 2010). Still, this was a genuine attempt of measuring eye movements in reading in a systematic manner (Kasprowski & Ober, 2004).

Huey drew important conclusions, such as the finding that word skipping was common in reading. He also found that regression of the eyes to parts of text which has already been fixated was common. Additionally, he also discovered that focus was often on a few letters of a words rather than on the entire word. Furthermore, he noted that the last word in a sentence is almost never fixated on, and that the first and sometimes the second word in a new line is skipped as well (Walczyk, J. J., Tcholakian, T., Igou, F., & Dixon, A. P., 2014).

According to Rashbass (1961), Raymond Dodge took a previous technique further for eye movement measurement, which was based on recording light reflected on a small mirror. He subsequently developed a new technique which was based on photography. This technique had several advantages, compared to the previous one in terms of being comfortable to use. It also had no physical connection to the eye lid or any part of the eye. Moreover, Dodge employed an assistant to watch his eye movements while reading (cited in Wade, 2010). Dodge's method set a starting point of techniques elaborating further on photographic-bases eye tracking (Richardson, D. C. & Spivey, M. J., 2004).

The eye tracking by photography invention is generally viewed as the spark which took eye tracking from its first wave into its second wave of research. The first era of eye tracking, which started by research in 1879, progressed until 1920s, when the ingenuity of photography-based eye tracking became common in many laboratories. The second wave, which focused more on application based aspects of eye tracking rather than on further developments of the techniques, is synonymous with Guy Buswelland, Alfred Yarbus. Yarbus did not only apply photography based eye tracking to textual reading, but also on picture viewing (Rayner, 1998; Duchowski, 2002; Wade, 2010).

Yarbus (1967) created improvements on previous eye tracking techniques and developed novel devices that has become famously known for over time (see Figure 2 .5) (DeAngelus, M. & Pelz, J. B., 2009).

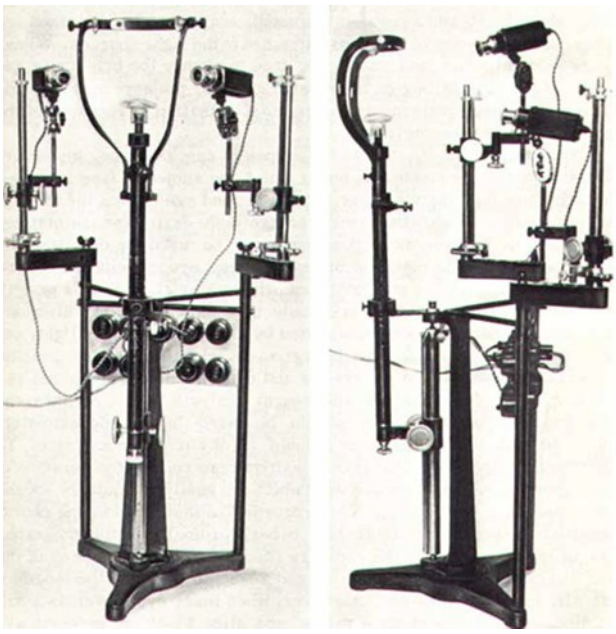

Figure 2.5 Front and side view of the Alfred Yarbus

supporting device/apparatus (adjusted from figure 3 of (DeAngelus & Pelz, 2009)).

Yarbus also developed suction "caps" to project images from the retina of the eye in order to be recorded (see Figure 2 .6) (DeAngelus & Pelz, 2009).

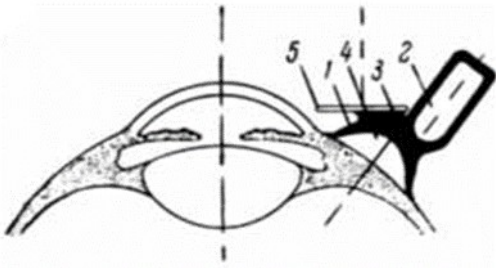

Figure 2.6 A diagram of one of the suction caps

used by Albert Yarbus (adjusted from figure 1 of (DeAngelus & Pelz, 2009)).

The third era of eye-tracking took place from the end of the 1960s (Yarbus publication in 1967) to the end of the millennia. According to Rayner (1998), the fourth (i.e., the recently present) wave of eye-tracking research got its start (Rayner, 1998; Duchowski, 2002; Bax, 2013) at the end of 1990s.

Today's eye-tracking techniques or systems may be categorized into two major groups, such as computer vision based eye tracking and sensory-based eye tracking. In computer vision based eye tracking, specialized cameras are employed to trace the movement of one or both eyes. The sensor-based eye tracking systems use electrodes with which electrical potential is measured, and are fixed at various points around the eyes. The latter group of eye-tracking systems is far less common than the first one (Al-Rahayfeh, A. & Faezipour, M., 2013).

The scientific driving force of eye movement investigations is to understand the underlying processes in reading. These processes can be categorized as orthographic, lexical, syntactic, and semantic and results in a vast field of studies. In order to conduct such an investigation I will describe some underlying aspects in eye-tracking research in the following section.

## 2.5 Underlying aspects of eye-tracking research

### *2.5.1 Eye movement and cognitive processes*

Cognitive processes may be conceived as inner processes that steer physical or other mental activities. In terms of eye movements in reading, Reichle (2006) states that one such major cognitive process is word identification. He goes on to say that this cognitive process steers when the eyes are to be moved, and that on a narrow "moment-to-moment" occurrence.

Eye movements in general (when a person is reading), as described by Wotschack (2009, via e.g. O'Regan, 1979), is influenced by various aspects of the text. Wotschack (2009) states that one of the most important reasons for studying eye movements in reading is gaining an understanding of the features that control the guiding process of eye movements, and the relevant cognitive processes like word identification. The link, as described by Reichle (2006), between these cognitive processes and eye movements in reading is a relevant model that enables a

comprehensive understanding of the reasons behind why we move our eyes the way we do during reading (Wotschack, 2009).

### *2.5.2 Eye movement control models in reading*

There exist two broad types of initial eye movement control models in reading research: cognitive-control models and oculomotor-control models. Some examples of each are the Morrison model and the READER model versus the O´Regan model, respectively (Thibadeau, R., Just, M. A., & Carpenter, P. A., 1982; Morrison 1984 via Reichle, 2006; Rayner, K. & Raney, G. E., 1996; Reichle, 2006). As evident by their names their focus on describing the eye movement controlling issues are different. The first assumes that some type of cognitive occurrence (or event) indicates the eyes to move, while the models of the latter group states that the oculumotor system is what guides the eye through a text (O´regan 1990, 1992 via Reichle, 2006; Rayner & Raney, 1996).

For example, in the Morrison's model (of the first group), the cognitive event is initially a combination of both eye gaze and visual attention aimed at the foveal word (wn). During the same fixation, when at some level of processing enough information is starting to be attained, the focus changes to the next foveal word (wn+1). Hence, this allows the parallel processing of a word along with the preparation for a "motoring program" of moving the eye to the next fixation point. However, O´regan (of the latter group) describes the cognitive occurrence as being determined largely due to the landing position (of the fixation) in a word. This landing position guides where the next fixation is made and for how long the present fixation is held (Morrison, 1984, and Rayner & Raney, 1996).

More recent models have been proposed such as the SWIFT model (Engbert, R. & Kliegl, R., 2005) and the E-Z reader model (Reichle et al, 1998). These two models contend that cognitive (of higher level) and oculumotor (of lower level) aspects are at play during eye movements when reading. The first model claims that issues with the processing of lexical material may inhibit the otherwise randomly started saccades, while the latter model says that during an early phase of lexical programming the oculumotor system in preparing for the subsequent saccade to occur (Reichle, 2006).

While SWIFT and E-Z reader share many aspects in their models, they also differ in numerous ways, such as their underlying assumptions on word identification. The former believes that many words are processed simultaneously, while the latter indicates that during each fixation two or more words are processed, not simultaneously, but rather on some one-by-one procedure (Rayner et al., 2012).

In this section, various theoretical control models that which are used in eye movement studies, has been described. In the subsequent section a data analysis approach, the linear mixed model, is presented.

### *2.5.3 Data analysis with the linear mixed model (LMM)*

The utilization of liner mixed models (aka. multi-level models), for statistical data analysis, in this study is based their relevance of being able to incorporate two basic types of parameters in the investigation. Their parameters are; fixed-effects parameters as well as random-effects parameters. While the nature of the former describes effects, which has an application relevance to parts of, or even encompass relevance of the effects of a complete population, but the latter types of parameters has effects which are related to the units of the experiment.

Hence, the term mixed, just as in the two related types of models; the nonlinear mixed models and the generalized linear mixed models they all encompass fixed-effects parameters along with random-effects parameters, respectively. This leads to the ability of studying the inter-dependency on a parameter-by-parameter approach of the previously named types of parameter groups.
Furthermore, due to the nature of these models, the occurrence of variation is found both in the per-observation noise (which is commonly found in less complex models e.g. linear regression models) as well as in the random-effects parameters (Bates, 2007). Thus, this model will be employed in this study. The effects of interest are found below (Table 2.1).

Table 2.1 The various types of effects

looked at in the linear mixed model (LMM).

| **Fixed effects** | **Random effects** |
|---|---|
| Word length | Subjects |
| Word type (Opacity/Transparency ) | Sentences |
| Word frequency | Words |
| Word predictablity | |
| Phoneme | |

The data analysis model (linear mixed model), that is utilized in this thesis study has briefly been described above. Moreover, it's common to apply some kind of general outline of how a study will be conducted. This will be addressed in the subsequent section, where I briefly scribe the cornerstones of the corpus analytic approach.

### *2.5.4 A corpus analytic approach*

In general the corpus-based analytic approach has four main attributes (Biber, D., Conrad, S., & Reppen, R., 1998; Wang, 2005);

- Studying the definite patterns of natural texts, in an empirical manner.
- The unitization of an encompassing collection of natural texts is used.
- Automatic as well as interactive computer techniques are broadly applied in the investigation.

- Qualitative along with quantitative analytical techniques are used and applied mainly on functional interpretations of the usage of language (Biber et al., 1998; Wang, 2005).

The corpus (dbrg.ut.ac.ir/Bijankhan/) used in the present investigation was created by the Database Research Group (at the faculty of Literature and Human Sciences) at the University of Tehran.

# CHAPTER 3

# 3 Methodology

## 3.1 Participants

This study was conducted at the Middle East Technical University, Turkey. Sixty students participated in this study, fifty participants spoke both Azari and Persian while the remaining ten only spoke Persian. There was an even gender split among the participants with thirty females and thirty males, whose ages ranged between twenty-two and forty-one (M=29.6, SD=4.03). Before initiating the experiment, the participants were asked to read and sign a form of informed consent that gave them details of the study as well as informing them that they could leave the experiment at any time and for any reason. Post experiment the participants were asked to fill out a demographic data form which gathered personal details as well as inquiries on their use of the language. It could be concluded from this information that none of the participants had previously been diagnosed with any language disorders, such as dyslexia. One of the participants did have difficulty pronouncing the letter “R” in some words. Furthermore, none of the participants wore any type of contact lenses. Nonetheless, sixteen participants (eight males and eight females) did wear glasses during the experiment and as such their test scores were removed from the study. One student was automatically aborted from the test and needed to retake the test. Due to this, we had to remove twenty-six sentences from this respondent’s test during the data analysis, which were sentences that the respondent had previously seen. In summary, this study was left with forty-four participants, twenty-two females and twenty-two males. This resulted in the new age range being from twenty-four to forty-one (M=29.81, SD=4.09). Each experimental session lasted thirty minutes.

An independent group of one hundred and twenty Persian speaking students from the Islamic Azad University’s Pharmaceutical Sciences Branch (IAUPS), located in Tehran, Iran, have participated in this study as well. Of these one hundred and twenty, sixty of these students were female and the remaining sixty were male, with an age range between eighteen and thirty-eight (M=23.66, SD=5.03). These students participated in a predictability test that help determine the predictability of words within the sentences. This type of experimental session took an average of fifteen minutes.

## 3.2 Materials

This study had a total of ninety-nine one-line sentences which were chosen from the Bijankhan Persian (Farsi) Corpus, which is a collection of 2.6 million manually tagged

words. There are forty Persian POS (part of speech) tags within the Bijankhan Persian Corpus. This collection was gathered from daily news articles and other various common texts. In this collection the documents are categorized into different subjects such as politics and culture. In this collection there was range of a total of four thousand three hundred various subjects (Amiri et al, 2007). The following criteria were utilized in order to choose which sentences would be used:

• The shortest sentences were merely nine words long while the longest was twelve words long (excluded zero-width space) (M=10.27, SD=1.01). The size of the characters was modified in order to fit into a line on the screen. Thus, the character sizes were no greater than 69 (M=55.76, SD=6.74). The font chosen for the experiment was Courier New with a font size of eighteen. This resulted in each character corresponding to a 0.42 degree of visual angle[1].

• None of the sentences used hyphenated words or numbers. In addition, no punctuation marks were utilized within the sentences themselves with the exception of a full stop at the end of a sentence.

• No abbreviations were used.

• Idioms were not used.

• Questions and interjective sentences were not used.

When counting the word length without a zero-width space as an independent character the length of the words ranged between one to twelve characters (M=4.78), SD=1.76). When counting the word length with counting zero-width space as an independent character ranged from one to thirteen (M=4.85, SD=1.9).

### 3.3 Apparatus

The subjects eye movements were tracked monocularly with an EyeLink 1000 Plus system (1000 Hz). A desktop mount with a chin and forehead rest was used in order to reduce any head movement. The sentences were printed out as a single line on the screen and were aligned on the right-hand side of the screen. They were displayed one at a time with a 1/3 vertical position from the top of the computer screen. The computers ran on Windows 7 professional with a 3.6 Ghz processor and a 24″ HDMI monitor with a display resolution of 1024 x 768. The participants were seated 73 cm from the screen with their head positioned on the forehead and chin rest. The camera was positioned 66 cm in front of them. The texts displayed with a black font on a white background with each character corresponding to 0.42 degrees of visual angle. There were 14.03 pixels (see Figure 3 .7).

[1] Multiply one character's size (i.e. 14.03pix*(390 mm (screen width)/1024 pix)) by 0.5 (i.e. 0.5A), and divide that by the viewing distance (730 mm) in the same units to find the tangent of the angle 0.5θ : 0.5A/D = tan (0.5θ ). Arctan (0.003659)= 0.2096 multiply by 2 (0.2096*2= 0.4192 rounded up to 0.42) to get θ in degrees or fractions of degrees.

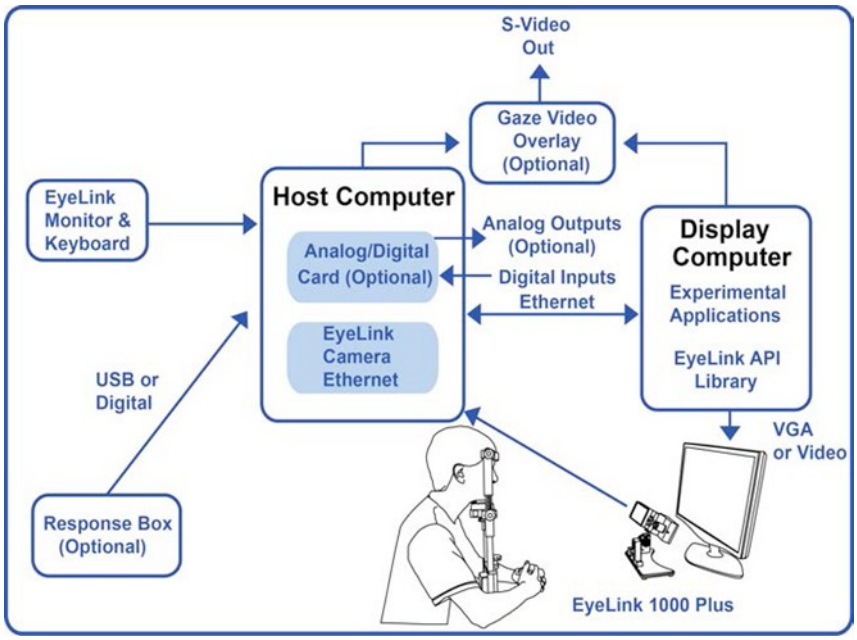

Figure 3.7 Eyelink 1000 Plus desktop mount

(Source: Eyelink 1000 Plus user manual 1.0.6, 2015)

## 3.4 Procedure

All participants signed an informed consent form that outlined the purpose of this study. They were also informed that this study may result in academic publications. The participants were also told that their personal identities would be kept confidential. Finally, this study was approved by the Middle East Technical University ethics committee.

### *3.4.1 Eye Tracking Experiment*

The participants were instructed to read the sentences silently at their normal reading pace in order to make sure they comprehended each sentence. After some trials they were presented with a series of yes or no questions. A Logitech Gamepad F310 was used for the responses. The right button was used to answer “Yes” while the left button was used to answer “No”. 89% (SD=0.3) accurately answered the questions. The participants were given breaks during the experiment and were allowed to rest for as long as they needed. In total there were 127 trials including 99 sentences, 4 practice trials, 4 practice questions and 20 questions.

General instructions were given before initiating the experiments. Pre-experiment a calibration was conducted for each participant using a standard nine-point grid. A bull’s eye was displayed on the right of the screen. Once the subject fixated on the bull’s eye during 500ms the sentence was displayed with the first letter of the sentence at the same point as the bull’s eye. The same process was repeated but with a bull’s eye displayed on the lower left side of the screen and post 500 ms fixation the screen would go blank and the sentence would be displayed at the same eye level as the first right bull’s eye.

Before starting the experiment each participant read four practice trial sentences with two follow up questions, after which there were nineteen trials made up of sentences and followed by random questions. At this point in time the participants were given their first break. This process was repeated an additional four times containing twenty trials. After each break participants were recalibrated and validated for accuracies. The questions and sentences used were randomly presented for each participant.

### *3.4.2 Corsi Block and Digit Span Experiments*

The participant's working memory was tested using the Corsi Block test and the Digit Span test. These tests were administered after the eye tracking experiment. Thirty participants took the Corsi Block test first and then the Digit Span test second. The other thirty took the Digit Span test first and the Corsi Block test second. The instructions for each test were given in Persian.

The Corsi Block test is administered by touching at max nine similar looking spatially separated blocks. The test starts out with two blocks and the difficulty continues to increase until the participant gets two sequences wrong. The Digit Span test is conducted by presenting a sequence of numerical digits that the participants are asked to recall in the order they were presented. The test starts with two digits and then becomes increasingly complex as the test progresses. After failing two consecutive sequences the test ends.

### *3.4.3 Predictability Experiments*

A Cloze task (Taylor, 1953) test was used in order to test the predictability of the words used in this experiment. The sentences were divided into conditions, with the shortest and longest having seven and ten words, respectively. A total of ten conditions. While each word in a sentence results in a different condition, the first and last words were removed and thus the predictability scores for these words were not calculated.

Consider the following sentence: "تلسکوپ جدیدی برای مشاهده دقیق سیاهچالهها طراحی شده است."
meaning "A new telescope has been designed for accurate observing black holes".
The sentence was truncated in this manner:

تلسکوپ…
Telescope…
تلسکوپ جدیدی…
A new telescope…
تلسکوپ جدیدی برای…
A new telescope for…
تلسکوپ جدیدی برای مشاهده…
A new telescope for observing…
تلسکوپ جدیدی برای مشاهده دقیق…
A new telescope for observing precise…
تلسکوپ جدیدی برای مشاهده دقیق سیاهچالهها…

A new telescope for accurate observing black holes…

تلسکوپ جدیدی برای مشاهده دقیق سیاهچالهها طراحی…

A new telescope for accurate observing black holes design…

The above is an example of how the sentences were truncated and how the sentences were put together in each condition. Therefore, this gives an idea of the manner in which the experiments were conducted. Hence, there were ten separate experiment sessions encompassing ninety-nine trials. Each participant was presented with a different set from the total of the available trials. The participants were asked to add the word that they thought fit the context of the sentence in the blank spaces. They were asked to use their intuition rather than thinking long on each sentence. By giving different conditions to the various participants, each word was tested by twelve participants. In case a participant predicted a word correctly they were given a score of 1. Any other response was scored as a 0. This resulted in a blind score in which only the actual word was scored as a 1, while any derivatives of that word were scored as 0. One such failure could be to choose the word "completely" instead of the word "complete". The predictability score of each word was presented as a proportion of the number of times the word was correctly predicted.

### 3.5 Eye Movement Data Analysis

The EyeLink Data Viewer was used for the purpose of data analysis, a software package utilized to view, process, and filter data collected using the EyeLink system. Furthermore, the program does also display auto segmented interest areas[2] (see Figure 3 .8).


Figure 3.8 A sample eye movement data in DataViewer.

The blue circles denotes fixations where the numbers are fixation durations, and the rectangles are interest areas.

*Fixation Report and Interest Area Report* from DataViewer are the output data from this investigation. First fixation durations (FFDs) along with gaze durations during first pass (GDs) from the Interest Area Report for interest areas were directly acquired.

IA_FIRST_FIXATION_DURATION and IA_FIRST_RUN_DWELL_TIME, was the names they were given in the data viewer, respectively.

[2] Date viewer works by truncating sentences in a word-wise manner which creates interest areas, which is conducted so that each word is associated with an interest area. By this follows that the Interest Area Report in describing the fixation properties within the related interest areas.

IA_FIRST_FIXATION_X was also rendered in the report resulting in the pixel position of the first fixation along the x-axis, inside of the current interest area.

To obtain the first fixation landing positions (FLPs) in terms of characters, then the following calculations were conducted:

Initially, the empty characters on the right side of the sentences along x-axis, composed of 42 pixels, were subtracted from the value of fixation position. Subsequently, this new obtained value was divided by 14.03 which correspond to one character in pixel. This results in the pixel value being converted into character unit. In the next step, the sum of word lengths as well as the number of spaces before the current word was subtracted for each word. The final value resulted in a fractional number, however it does not make sense to talk about fractional character indices; therefore, it was checked and observed carefully in order so that the rounding up of these values resulted that it landed in the exact character position; hence, they were rounded up in order to get the FLP value.

In the same manner, the total number of fixations in a trial in the first run of the current interest area, the number of times the interest area was entered from a higher IA_ID (from the left in Persian), and the number of times interest area was exited to a lower IA_ID (to the right in Persian) before a higher IA_ID was fixated in the trial were obtained directly from the Interest Area Report. They were named in Data Viewer as IA_FIRST_RUN_FIXATION_COUNT, IA_REGRESSION_IN_COUNT, and IA_REGRESSION_OUT_COUNT respectively.

The Interest Area Report also provided information on whether the current word had been skipped or not during first pass reading. The words which have blinks or not during first pass reading have been acquired via Fixation Report.

By following the approach of the analysis found in the literature (Kliegl, Nuthmann, &Engbert, 2006; as cited in Yan et al., 2014) log10 transformation was applied on word frequency which ranges from 1 to 120,558 per million while log2 transformation was applied to word length values. Furthermore, the natural logarithm transform was applied on the first fixation duration along with gaze duration, which deleted the skew in the distributions. Also, logit transformation was applied for predictability values (Kliegl, R., Grabner, E., Rolfs, M., & Engbert, R., 2004).

### 3.6 Data Selection

Before the data was analyzed, several exclusion criteria were applied. The first is that the first word and the last word were removed. As Yan et al. (2014) state this is a common practice for research conducted over reading. Second, all of the data was checked manually to make sure there was no loss of data and to correct for any issues. Trials that displayed any loss of data or had offset issues were excluded from the experiment. Lastly, words with shorter and longer first fixation duration (FFD) than 30 ms and 1000 ms, respectively, were removed. The same was done with gaze durations (GDs) longer than 1500 ms along with words which was blinked at during

fixation (Kliegl et al, 2004). Any fixation made prior to the words on the empty space was excluded from the analysis, as well.

### 3.7 Linear Mixed Model (LMM) Analysis

Random effects structure was controlled and the fixed effects were defined by the utilization of the linear mixed model (LMM). The effects of the following variables were investigated:

- Word length
- Word type
- Word frequency
- Word predictability
- Phonemes

The effects of previously described factors were investigated the eye movement parameters below:

- First fixation duration (FFD)
- First run fixation count (FRFC)
- Gaze duration (GD)
- First fixation landing position (FLP)
- Regression in count (RIC)
- Regression out count (ROC)

It should be noted that the results from the FLP was added to the FFD, FRFC and GD models since these parameters follows the FLP parameter (Yan et al., 2014).

The optimal models for each dependent variable in LMM rejected to keep the word length excluding zero-width space (as an independent character) and pixel variables due to collinearity error. Therefore, in the present study word length variable contains only characters including zero-width space. Regarding word type (transparency/opacity including Persian Ezafeh suffix (as a phoneme)) I decided to discard the variable due to lack of significance and to keep word type transparency/opacity excluding Persian Ezafeh suffix in the all models.

One of the important assumptions of linear models is regarding the independence assumption. In the case of behavioural studies, they generally violate the independence assumption due to the same subjects giving rise to multiple responses during the experiment. In such a setting, it is impossible to evaluate each response as independent (see Figure 3 .9).

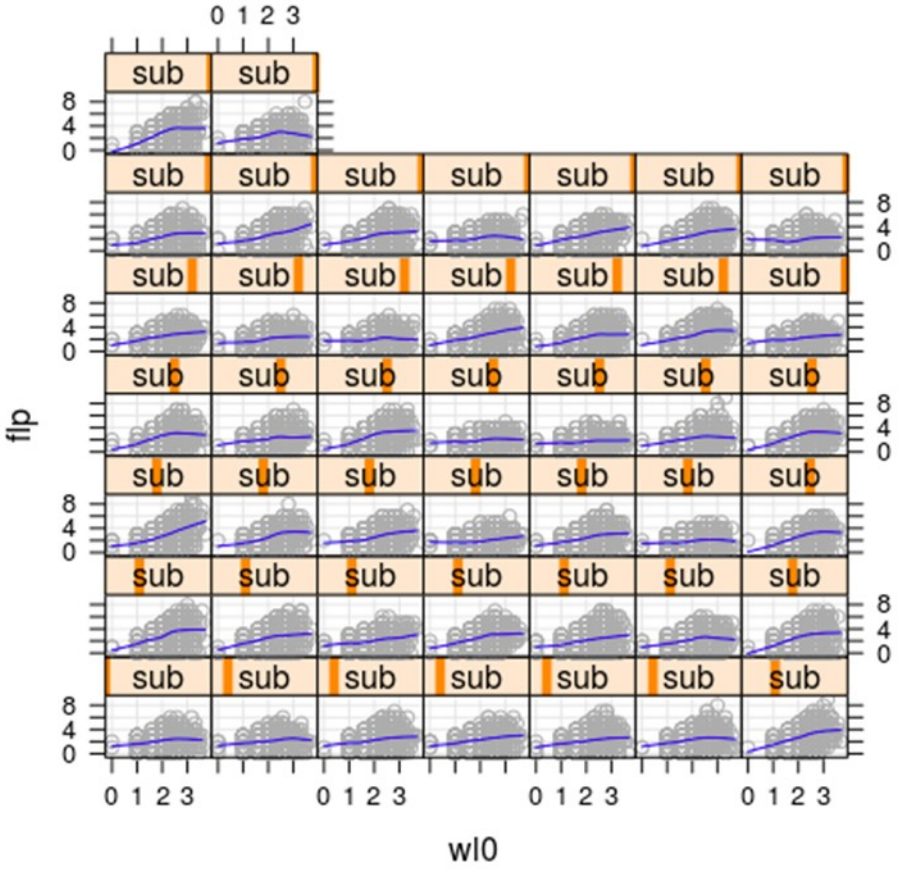


Figure 3.9 Quantile-quantile plot for FLP responses

among subjects. x axis is log2(wl0), y axis is log2(flp), each subgraph belongs to different subjects.

As Figure 3 .9 shows, it is possible to detect individual variations during the experiment. Therefore, to deal with the violation of the independence assumption in this study, I added a random effect for each subject which helped in creating different baselines for each of the subjects (Winter, 2014).

The addition of the random effect for each subject is a special feature of the linear mixed model, which we utilized. Unlike regular linear models, the linear mixed models not only deal with fixed effects but also with random effects as well, which captured individual differences. There is also the issue of by-sentences and by-words variations which violated independence assumptions. To overcome these issues, I added additional random variables to the models. Thus, by applying these aspects we have different intercepts for different subjects, sentences, and words which capture variations in dependent variables.

Thus far, we have only considered the random intercept model, however the fixed effects are not the same for all the subjects. Moreover, the slopes should also be adjusted for the random slope model. As an example, (1+wl|sub) infers that we have different baselines (or intercepts) as well as different responses relate to word length. Before I could construct the models, the outliers and missing values had to be examined. Due to no missing values in the data set occurred, no imputation methods

were applied. However, outliers were remove based on two main criteria (Kliegl et al., 2004):

1. Gaze duration (GD) records that are higher than 1500 ms are removed.

2. First fixation duration (FFD) records that are higher than 1000 ms or lower than 30 ms are removed, as well.

Furthermore, a log transformation is applied to normalize the dependent and independent variables in a manner as follows (Kliegl et al., 2004):

- The natural logarithm of skewed FFD and GD variables are used.
- The log2 of skewed FLP, PHON, and WL variables are used.
- The log10 of skewed WF and PIX values are used.
- A logit transformation is applied on the WP values.

As a result, my models are built following the standard modelling procedures as described below (Yan et al, 2014; Kliegl, R., Wei, P., Dambacher, M., Yan, M., & Zhou, X., 2011; Baayen, 2008a; Baayen, R. H., Davidson, D. J., & Bates, D. M., 2008b; Hohenstein, S. & Kliegl, R., 2014).
The application of random effects in the Linear Mixed Model (LMM):

- Initially, random intercepts were added for the subjects, sentences, and words.
- Secondly, random slope factors were added exclusively for the subjects.

The independent variables (word length, word frequency, word predictability, word type, and phonemes) features were the same within the sentences and words for all the subjects. Thus, it would not be sensible to add these variables into the random slopes of sentences and words.

In this study, I utilized the *lmer function* of the lme4 package according to the R. Modelling steps, as described below:

1. A dependent variable was chosen. This first dependent variable chosen was First Fixation Landing Position (FLP).

2. Secondly, the base model was constructed as simple as possible. Based on the methodology, I only utilized the word length for this.

3. An alternative model was constructed by adding an additional covariate to the base model.

4. These two models were subsequently compared by using the ANOVA test. In the case that a new term does give a significant result, i.e. additional power to explain a dependent variable, the new variable was kept. In case the alternative model failed to produce a significant result, the covariate was omitted, and a new covariate had to be added to the base model.

5. Step 3 and step 4 were repeated until all the independent variables are added to the models.

In addition, it should be stressed that in step 4 a log probability test using ANOVA function was conducted. By comparing the probability of the two models (reduced and full model) with each other I obtained a p-value from the ANOVA test. If the comparison proved to be significant (i.e. the p-value is less than 0.05 (p-value <0.05)), it would lead us to continue with the full model.

In summary, the present chapter has given a description of important information related the methodology utilized in this study. This methodology, consisting of materials, apparatuses, subjects, data analysis and data selection, as well as procedures with the latter being the application of the linear mixed models (LMM).

# CHAPTER 4

# 4 Results

In this part of the thesis I present the results from the eye movement measures: first fixation duration (FFD), first run fixation count (FRFC), gaze duration (GD), first fixation landing position (FLP), regression in count (RIC) and regression out count (ROC) from the runs of the liner mixed models (LMMs). For convenience follows a short recapitulation, describing the individual eye movement measures prior to reporting their results.

- First Fixation Duration (FFD): The duration, measured in units of milliseconds [ms], of the first fixation in the current interest area.
- Gaze Duration (GD): The total duration, measured in units of milliseconds [ms], of all fixations on the current interest area.
- First Run Fixation Count (FRFC): The total number of fixations, in the first run of the current interest area in a trial.
- Regression-in Count (RIC): Number of times, the interest area was entered from the left in Persian.
- Regressions-out Count (ROC): Number of times, the interest area was exited to the right in Persian, before the upcoming fixated word in the trial.
- Fixation Landing Position (FLP): The character of the word (i.e. character based measuring) on which the first fixation is made.

## 4.1 Fixation Landing Position (FLP)

In this section I present the results from the fixation landing position (FLP). The word length, word type, word frequency, and word predictability will be shown and how they relate to the fixation landing position.

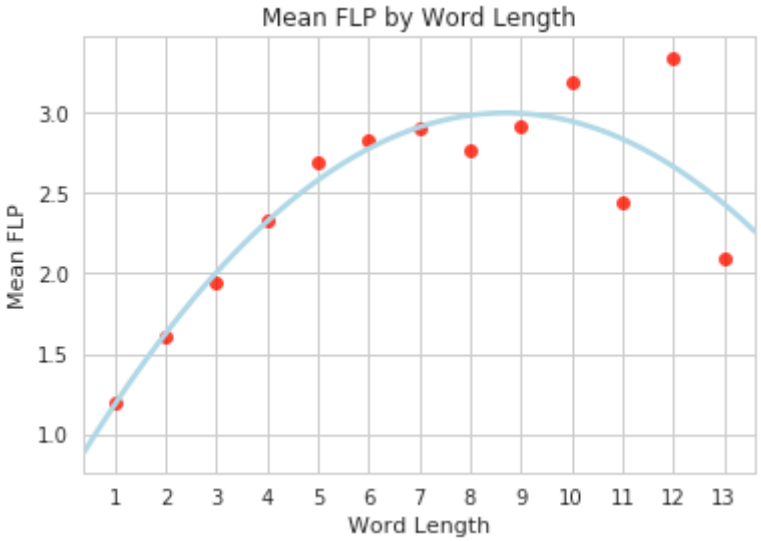


Figure 4.10 The word length (WL) plotted

against the mean first fixation landing position (FLP).

Figure 4 .10 (shown above) shows the lengths of the words against the mean of the first fixation landing position (FLP). This is shown as red dots and a blue trending line. From word length one until nine a logarithmic increase of the fixation landing position is seen. Post word length nine the advent of a fluctuation is observed. However, if looking at word length between one and five an almost liner increase in mean first fixation landing position is viewed in the graph. This linear increase of the mean FLP prior to word length five as well as the fluctuation in mean FLP post word length nine may both be explained by the higher and lower number of each word length, respectively (Appendix A;Error: Reference source not found).


Figure 4.11 Thirteen graphs, one for each word length,

1 to 13, respectively, display the first fixation landing position for each word length relative to the fixations count, which results in that the graphs show the preferred viewing location (PVL) for all the word lengths.

In the Figure 4 .11 above the preferred viewing location (PVL) in relation to the word length of the Persian words is shown. Hence, since Persian script is read from right to left, these results must be interpreted in an opposite manner versus Latin script words. This means in word length two and three PVL was found to be mainly at end or left side of the word. In words with length four and five PVL seems to be situated in the center or slightly right of the word. However, for words with length six to nine PVL seem to be found more towards the beginning of the words, hence to the right. For word length ten to thirteen the data points are too few to make any reliable interpretation. Hence, it seems that with increasing word length the PVL travels from the left to the right on those words.

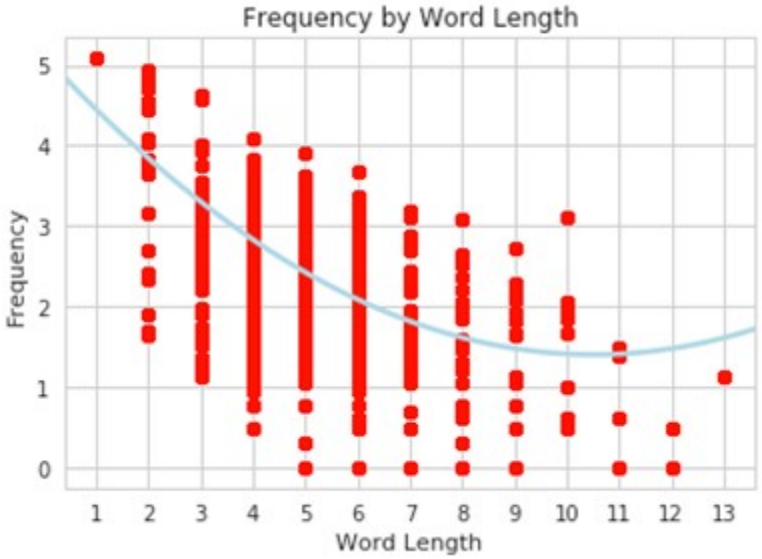


Figure 4.12 A graph displaying the relationship

between word length and word frequency.

The relation between word length and frequency can be displayed with a trend line (See Figure 4 .12). This figure shows that with increased word length the less frequent the word is.

A summary of the general effects of the independent variables on the eye movement measure first fixation landing position (FLP) from the linear mixed model analysis (LMM):

Random effects:

- It was found that the variance between the subjects is greater than the variance between sentences. However, it was also found that the highest variance was found to be between words.

Fixed effects:

- Regarding fixed effects, various effects were observed. A statistically positive effect of word length (WL) on the first fixation landing position (FLP) was found in that higher FLPs correspond with longer words (b = 0.63, SE = 0.06, t = 9.43) (see Figure 4 .10).

- In regards to word frequency (WF) a similar positive effect was found. Hence, higher FLPs corresponds with higher WF, i.e. with more frequent words (b = 0.103, SE = 0.02, t = 4.88) (see Figure 4 .13).

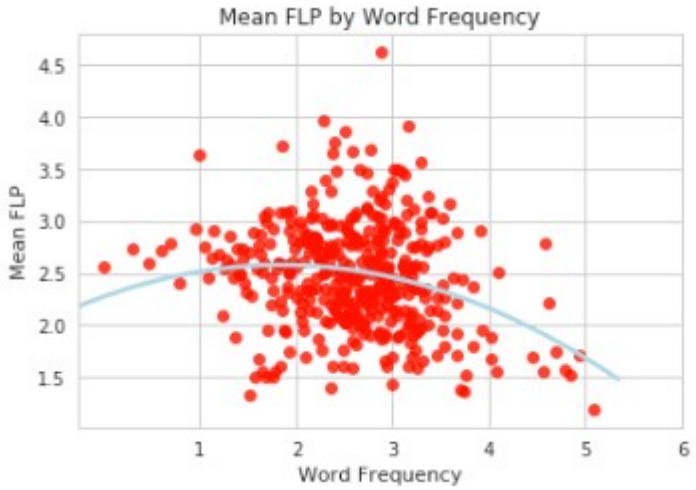


Figure 4.13 A graph showing the word frequency

(WF) plotted against the first fixation landing position (FLP).

- However, no significant effect was found between first fixation landing position and each independent variable: word predictability, word type (opacity, transparency) and phonemes.

- Lastly, an interaction displayed a positive effect on the first fixation landing position with regards to word length and word type (transparency). This means that with an increasing word length and if the word is transparent the results indicated an increase in FLP (b = 0.13, SE = 0.06, t = 2.21).

## 4.2 First fixation Duration

The results for how the independent variables affect the First Fixation Landing Position analyzed by Linear Mixed Model (LMM) was shown in the previous part. In the present part the result of the same variables' effect on the First Fixation Duration will be provided.

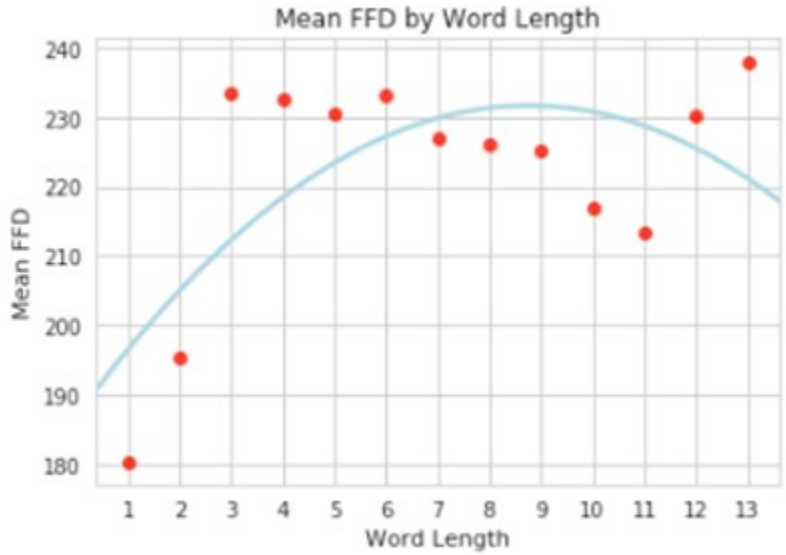


Figure 4.14 The word length (WL)

plotted against the mean first fixation duration (FFD).

The word length has been plotted against the mean for First Fixation Durations (FFDs), shown in the Figure 4 .14. Initially it can be observed that between the word lengths 1 and 4 the fixation duration increases almost proportionally, and post word length 4 the fixation duration starts to decrease. This decrease in the first fixation duration occurs until word length 11 where that the first fixation duration starts to increase until word length 13.

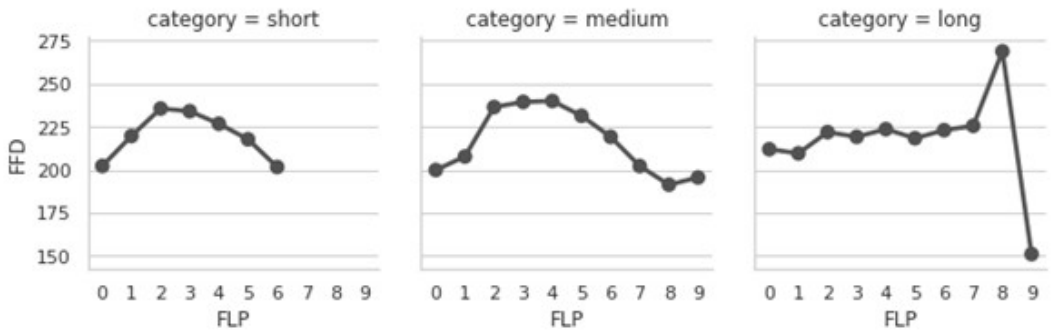


Figure 4.15 The fixation landing positions

(FLP) plotted against the first fixation durations (FFD) in three graphs based on word length category, which are 1 to 5 (short), 6 to 9 (medium) and 10 to 13 (long), respectively.

The general effects of the independent variables on the First Fixation Duration (FFD) from the running of the linear mixed model (LMM):

Random effects:

- The largest variance is found between subjects, which may be attributed to the personal traits of the subjects. Furthermore, it was found that the second

highest variance is found between words, while the lowest variance is found between the sentences.

Fixed effects:

- It was found that medium long words (in the span between 5 and 11 in character length) were fixated shorter compared to that those of short word length (1 to 4). This may be explained by high refixation rates for medium-long words (b = - 0.052, SE = 0.016, t = -3.15)

- With regards to word frequency it was found that a disproportionate relation exists between WF and the first fixation duration. This could be explained in that a word more commonly seen need less time of fixation to be processed, which even seems intuitive. (b = -0.048, SE = 0.005, t = -9.69) (see Figure 4 .16).

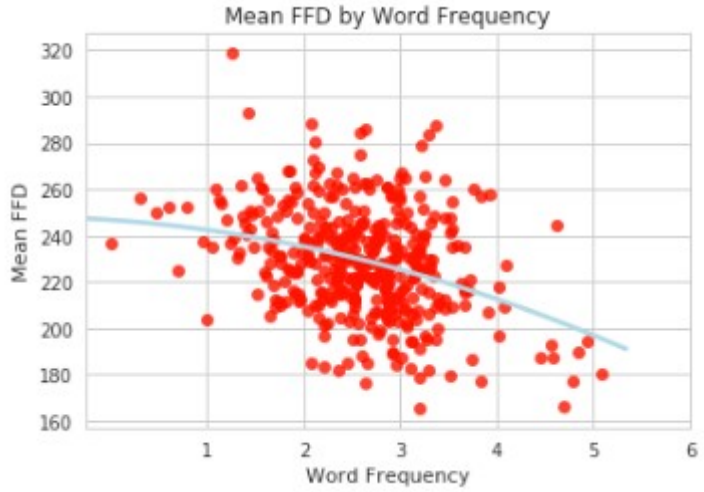


Figure 4.16 The word frequency (WF)

is plotted against the mean first fixation duration (FFD).

- Also, the fixation landing position and its relation to the first fixation duration was investigated. It was found that these eye movement measures are proportionately related to each other, hence if the landing position increases so does the fixation duration as well. (b = 0.014, SE = 0.001, t = 9.436)

- The interaction between two independent variables have a negative effect on the first fixation duration. This interaction when occurred between the word length and the phoneme has this effect. Hence, when the word length increases and the phoneme increases the first fixation duration decreased (b = -0.076, SE = 0.014, t = -5.26).

- Another interactive effect was between word length and word type, which resulted in a decrease in first fixation duration. The decrease occurs when the word length increases and the word is transparent, and this has a negative effect on the first fixation duration (b = -0.046, SE = 0.014, t = -3.13).

- Lastly, no significant effect was found on first fixation duration with respect to word predictability, word type, and phonemes.

## 4.3 Gaze Duration

In the previous section the results regarding the first fixation duration (FFD) were described, and in this section the results of the subsequent eye movement measure, Gaze Duration (GD), will be presented.

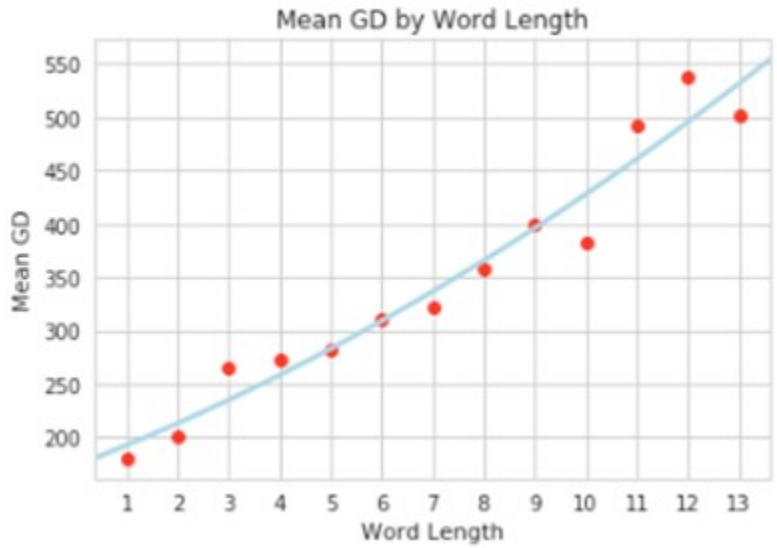


Figure 4.17 The word length (WL) is plotted against the mean gaze duration (GD).

In Figure 4 .17, word length is plotted against the mean gaze duration. Through most of the graph we see a strong proportionality between the word length and the mean gaze duration i.e. that an increasing word length is resulted in a long gaze duration time. It is plausible to assume that this is in part caused by a high fixation counts during the reading. However, a slight fluctuation can be seen from word length nine and onwards.

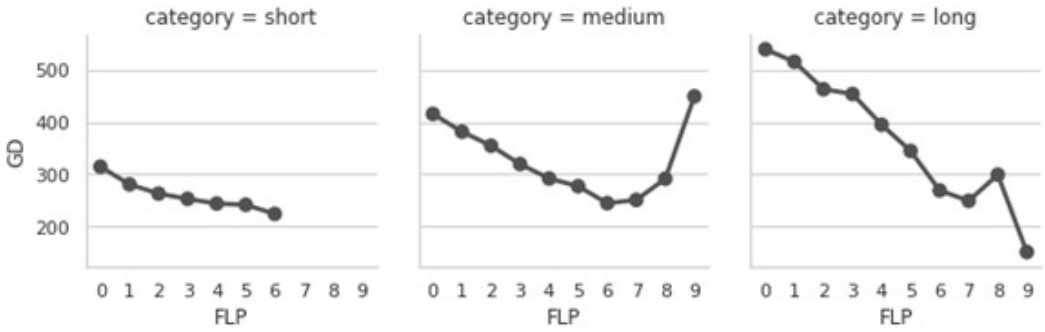


Figure 4.18 The fixation landing positions (FLP)

plotted against the gaze durations (GD) in three different graphs based on word length category, which are 1 to 5 (short), 6 to 9 (medium) and 10 to 13 (long), respectively.

A summary of the general effects of the independent variables on the eye movement measure Gaze Duration (GD) analyzed by the linear mixed model (LMM):

Random effects:

- Three random effects have been observed in the model: The variance between words displays the highest variance as a random effect followed by the variance between subjects, while the variance between sentences have shown the lowest variance of these three random effects

Fixed effects:

- A consistent and proportionate relation has been found in that increasing word length resulted in longer gaze durations. This could be hypothesized to be caused by a high fixation counts during the readings (b = -0.23, SE = 0.02, t = 10.44).

- It was also found that with increasing word frequency, shorter gaze durations were observed, hence, there is a negative effect of word frequency on gaze duration (b = -0.08, SE = 0.006, t = -13.085) (see Figure 4 .19).

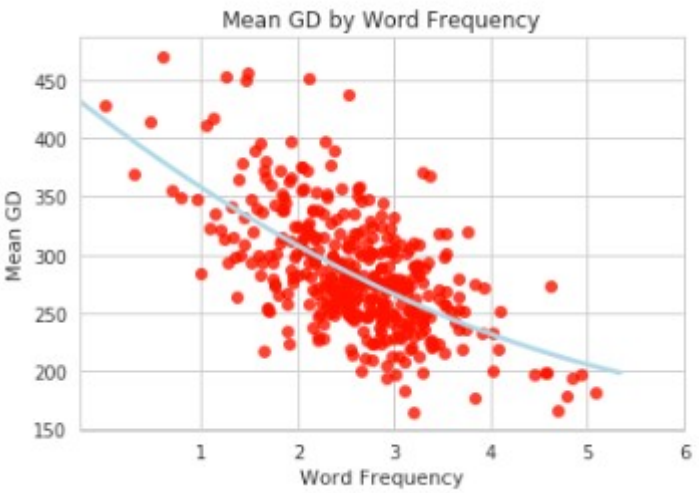


Figure 4.19 The word frequency (WF) is plotted

against the mean gaze duration (GD).

- In the manner as with word frequency, a negative effect of word predictability on gaze duration was observed i.e. when the word predictability is high the gaze duration is short (b = -0.027, SE = 0.009, t = -3.01).

- In relation between the eye movement measures fixation landing position and gaze duration it was found that when the fixation landing position increased, the gaze duration decreased. Hence these two measures show a disproportionate relation as well (b = -0.058, SE = 0.001, t = -30.814).

- It was found that the interaction between word length and word frequency resulted in a negative effect on the gaze duration, i.e. with elevated word length and frequency the value on gaze duration decreased (b = -0.033, SE = 0.01, t = -3.34).

- The same type of negative effect on the gaze duration was found in the interaction between world length and fixation landing position i.e. when the two latter variables was elevated the gaze duration decreased (b = -0.029, SE = 0.003, t = -9.02) (see Figure 4 .15).

- There was not any significant effect on GD respect to word type and phoneme.

## 4.4 First Run Fixation Count

In the previous section the results of the linear mixed model for the gaze duration was described. In this section the First Run Fixation Count (FRFC) results will be displayed as effected by the independent variables.

The general effects of the independent variables on the eye movement measure First Run Fixation Count (FRFC) analyzed by a linear mixed model (LMM):

Random effects:

- It was found that a higher variance existed between subjects than the variance between words in a sentence, but the variance between sentences was the lowest.

Fixed effects:

- Regarding fixed effects, the first run fixation count displayed a proportionate relation i.e. a positive effect of word length, i.e. with increasing word length corresponds an elevated first run fixation count (b = -0.48, SE = 0.025, t = 18.8) (see Figure 4 .20).

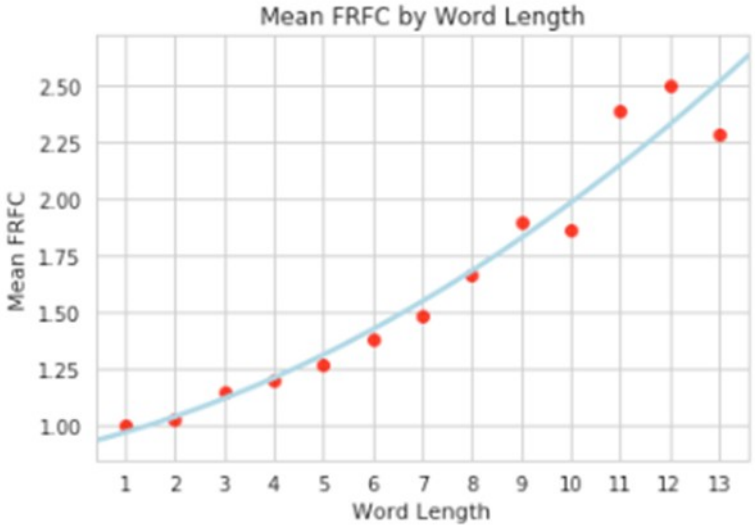


Figure 4.20 A graph showing the word length (WL)

plotted against the first run fixation count (FRFC).

- The effect of word frequency on first run fixation count was found to be negative, hence that with a higher word frequency yields a lower first run fixation count (b = -0.068, SE = 0.008, t = -8.34) (see Figure 4 .21).

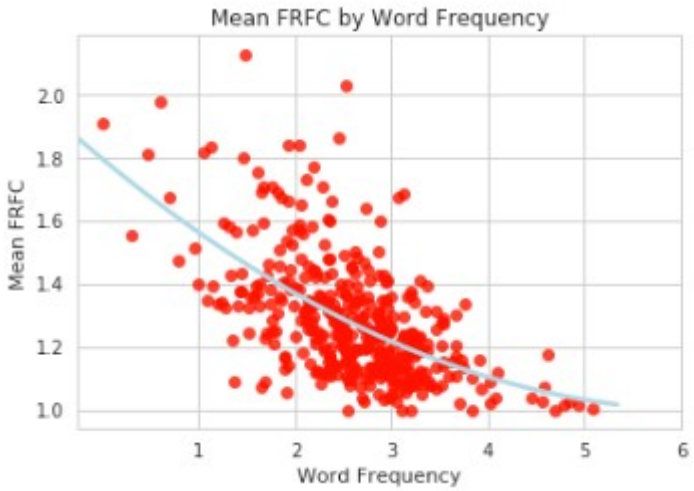


Figure 4.21 A graph showing the word frequency (WF)

plotted against the first run fixation count (FRFC).

- In the same manner, with an elevated word predictability corresponds a decreased first run fixation count, as per negative effect (b = -0.035, SE = 0.008, t = -4.08).

- A negative effect of fixation landing position on first run fixation count was also found, hence an increase in the former resulted in a decrease in the latter (b = -0.108, SE = 0.002, t = -47.53).

- The interaction between word length and word frequency was found to display a negative effect on the first run fixation count (b = -0.051, SE = 0.01, t = -4.95).

- The same type of effect was found when word length interacted with word predictability on first run fixation count (b = -0.022, SE =0.011, t = -1.995).

- The interaction between word length and phonemes showed a positive effect on the first run fixation count (b = -0.14, SE = 0.019, t = 7.48).

- An interaction between the fixation landing position and word length showed a high negative effect on the first run fixation count (b = -0.06, SE = 0.003, t = -15.305).

- There is no significant effect on first run fixation count with respect to word type and phoneme

### 4.5 Regression in Count

In the previous section the results from the linear mixed model for the first run fixation count was described. In this section the findings of the Regression in Count (RIC) analyzed by linear mixed model (LMM) with be described.

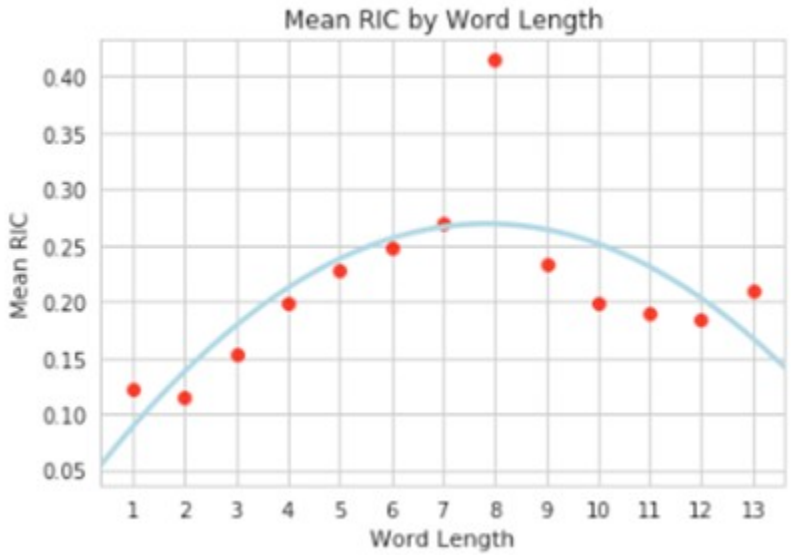


Figure 4.22 A graph showing the word length (WL)

plotted against the regression in count (RIC).

A summary of the general effects of the independent variables on the eye movement measure Regression In Count (RIC) analyzed by linear mixed model (LMM):

Random effects:

- The most notable random effect was that the highest variance between words, the second highest variance was that between subjects while the lowest variance was found between sentences. Noteworthy is that the latter variance was expected to be the lowest amongst the three types of random effects.

Fixed effects:

- Initially, it was found that a negative relation existed between the regression in count (RIC) and word frequency, i.e. that with frequent words in the text the RIC value was found to be lower (b = -0.022, SE = 0.008, t = -2.6) (see Figure 4 .23).

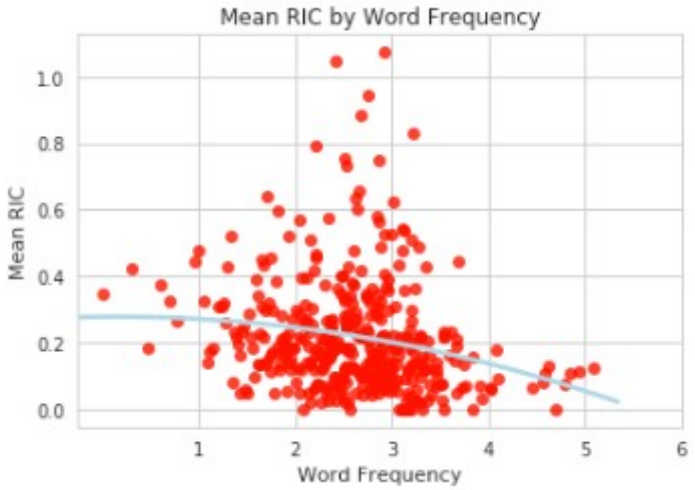


Figure 4.23 A graph showing the word frequency (Wf) plotted

against the regression in count (RIC).

- A negative effect of word predictability on regression in count was observed, hence with more predictable words the RIC values decreased (b = -0.076, SE = 0.011, t = -6.46).

- Thirdly, it was found that transparent words and regression in count expressed a positive relation with each other. Hence, with a transparent word the value of RIC increased (b = -0.053, SE = 0.021, t = 2.49).

- Another positive effect was found, this time effect of phoneme on regression in count. This means that when the number of phonemes increased the values of RIC increased as well (b = -0.08, SE = 0.032, t = 2.44).

- Lastly, it was found that the interaction between word length and word predictability had a negative effect on regression in count. Hence, when word length increased and word predictability increased the RIC decreased (b = -0.049, SE = 0.014, t = -3.37).

- There is no significant effect on RIC respect to word length and first fixation landing position.

## 4.6 Regression Out Count

In the last section the results of the linear mixed model on the eye movement measure regression in count (RIC) was discussed. In the present section the results of the LMM for Regression Out Count (ROC) will is displayed.

A summary of the general effects of the independent variables on the eye movement measure Regression Out Count (ROC) from the linear mixed model (LMM):

Random effects:

- Similar to the random effects in the latter section (i.e. the regression in count results) it was found that the highest variance was found between words, the second highest variance was found between subjects and the lowest variance was found, as was expected between sentences.

Fixed effects:

- A positive effect of word length on regression out count (ROC), thus with elevated word length an elevated value on the ROC was found (b = -0.05, SE = 0.023, t = 2.13) (see Figure 4 .24).

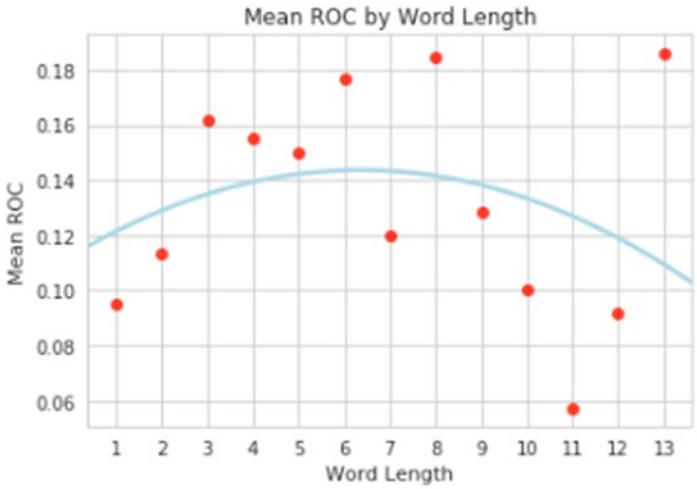


Figure 4.24 A graph showing the word length (WL) plotted

against the regression out count (ROC).

- In contrast to the previous fixed effect, a negative effect of word frequency was seen on the regression out count (ROC). This means that an increase in word frequency resulted in a decrease in ROC value (b = -0.014, SE = 0.007, t = -1.94) (see Figure 4 .25).

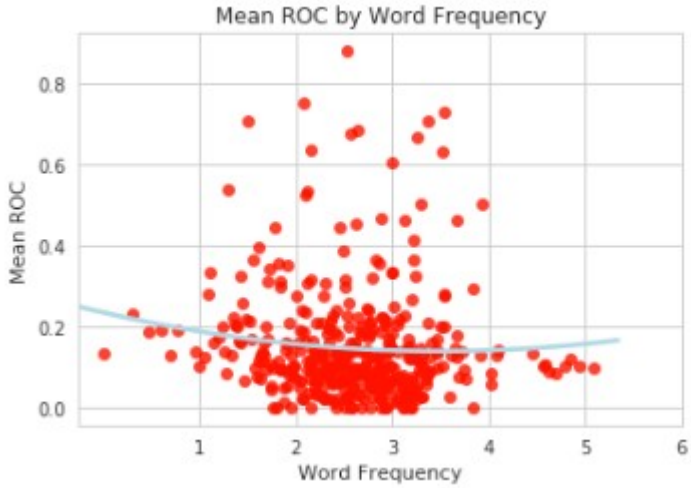


Figure 4.25 A graph showing the word frequency (WF) plotted

against the regression out count (ROC).

- A negative relation was observed between phonemes and regression out count (ROC). Consequently, with increasing numbers of phonemes the ROC value was found to be decreased (b = -0.07, SE = 0.026, t = -2.68).

- It was found that an interaction between the number of phonemes and the word length had a negative effect on the regression out count. Accordingly, with higher number of phonemes and with longer word length the ROC value decreased (b = -0.075, SE = 0.02, t = -3.58).

- Lastly, it was found no significant effect could be found by first fixation landing position, word predictability and word types on the regression out count value.

## 4.7 Result Summary

In this result section we have described the findings of all the eye movement measures,

- Fixation Landing Position (FLP)
- First Fixation Duration (FFD),
- Gaze Duration (GD),
- First Run Fixation Count (FRFC),
- Regression in Count (RIC),
- Regression out Count (ROC)

and how these eye movement measures have been affected by the independent variables,

- Word Length (WL)
- Word Type (WT)
- Word Frequency (WF)
- Word Predictability (WP)
- Phonemes (PHON)

and all of their relations or their lack that of. To describe their relations in an overview table, and their correlations they have been summarized and plotted in a correlation matrix, respectively. This correlation matrix, (see Figure 4 .26) incorporates colors' (blue and red) shades in its display, as can be seen below. The information of effects can be viewed in Table 4 .2.

Table 4.2 Displaying the relations
between the eye movement measures and the independent variables.

| *Independent variable* | *Eye movement measures* | | | | | |
|---|---|---|---|---|---|---|
| | (FLP) | (FFD) | (GD) | (FRFC) | (RIC) | (ROC) |
| (WL) | Positive | Negative | Positive | Positive | - | Positive |
| (WT) | - | - | - | - | Positive | - |
| (WF) | Positive | Negative | Negative | Negative | Negative | Negative |
| (WP) | - | - | Negative | Negative | Negative | - |
| Phonemes | - | - | - | - | Positive | Negative |
| (FLP) | - | Positive | Negative | Negative | - | - |


Figure 4.26: In this correlation matrix the linear relations

between the variables is described in terms of shades of colors. The darkest blue color (+1) indicates a strong positive correlation which decreases with decreasing shade of blue until it reaches white (0) which indicate a lack of correlation. The darkest brown color (-1) indicates a strong negative correlation which decreases with decreasing shade of brown until it reaches white (0) which indicates a lack of correlation.

# CHAPTER 5

# 5 DISCUSSION & CONCLUSION

In the present study I investigated how five independent variables relate to six eye movement measures during reading of Persian script, by native Persian speakers. The eye movement measures examined were first fixation landing position (FLP), first fixation duration (FFD), gaze duration (GD), first run fixation count (FRFC), regression in count (RIC), and regression out count (ROC). My independent variables were: word length (WL), word type (WT), word frequency (WF), word predictability (WP), and phonemes (PHON). These measures and variables were selected due to their appearance in literature related to similar studies as this one. This study was conducted based on the principles of corpus-based analytical approach, which incorporates four basic aspects which were outlined in the background in Chapter 2. The linear mixed model was chosen as the most suitable model with respect to its ability to control the random effects (subjects, words, and sentences) with respect to the eye movement measures and including individual differences in the analysis of experimental effects (Kliegl et al., 2010).

Initially, in regards to the first fixation landing positions (FLP), it was found that a logarithmic increase was seen between the mean FLP and the word length. However, this was specifically for word lengths from one to nine, but after that point a fluctuation starts which could be the result of much lower number of counts of longer words in the stimuli. One may also spot an almost linear relation between word length of one character and word length of five characters. These results are in line with what has previously been found in related studies (Yan et al., 2014; Nuthmann, A. & Kliegl, R., 2009; Deutsch, A. & Rayner, K., 1999; Rayner, K. & Raney, G. E., 1996; Rayner, 1979.) Regarding the preferred viewing location (PVL) it was found that the location seems to be different as a function of the word length of the word. The results showed that shorter words were focused near the end of the word (left - as per Persian script), while longer words close to the beginning (right - as per Persian script), but those of medium length were focused closest to the center. Thus, the PVL seemed to shift from left to right as the word length increases. This is in good agreement with the hypothesized effect that medium (or average) long words have their FLP near the center of the word. A positive relation was also found between the mean FLP and word frequency for low to medium frequency words. In the present study values of whole-word frequency were utilized, and it was observed that longer words have less frequency values. Therefore, the frequency effects are likely influenced by word length due to the close relationship between word length and frequency.

Regarding the results of first fixation duration (FFD), we see a mainly positive effect of word length on FFD. However, after word length six or seven until thirteen a slight fluctuation is seen. This means that short words result in shorter durations, medium-long words result in longer durations and the longest words result in long durations as similar to medium-long words. Furthermore, the inverted optimal viewing position effect was observed in this study: The longest durations were observed neither at the beginning nor at the end of the words but close to the center of the words, but with a slight shift to the beginning (right – as per Persian script). These findings are in line with what has been reported in previous studies (Vitu, F., McConkie, G. W., Kerr, P. & O'Regan, J. K, 2001; Nuthmann, A., Engbert, R., & Kliegl, R., 2005; Hyöna, J. & Bertram, R., 2011; as cited in Yan et al., 2014). Lastly, it was found that a negative (or disproportionate) effect of word frequency on FFD was found, which has been hypothesized. We found that lower frequency words have a higher FFD while more common words had a lower FFD. Hence, this hypothesis has been confirmed by the results, as well.

Gaze durations (GD), which is the sum of all the fixation durations when they are read for the first time, i.e. during the first-pass reading are found to be positively related with word length in what seems to evolve into an almost exponential increase with increasing word length. That is in line with the hypothesis on the relation between GD and word length. It was also found that fixations were longer when fixation landing position was in the beginning (right - as per Persian script) of the words in all the word lengths categories. Furthermore, it was observed that with increasing word predictability, as well as with increasing word frequency decreasing GD followed.

Regarding the eye movement measure first run fixation count (FRFC), it was found that the same relations was found towards word length, word frequency, and word predictability as between these independent variables and the previous eye movement measure gaze duration (GD), i.e. positive, negative, and negative, respectively. Hence, the hypothesis regarding that the effect of the word predictability on first run fixation count will be negative was found to be true.

As with the previously discussed eye movement measures a negative effect of word frequency and word predictability was found on regression in count (RIC). Which was hypothesized, as well. However, a relation was found between RIC and word length, which displayed a positive relation from length one to seven and a negative from nine until thirteen. It could be speculated that the latter relation between RIC and word length could be explained in part due to the negative effect found in the RIC when word length and word predictability interacted. Moreover, a significant effect of phonemes was found for the first time on an eye movement measure (regression in count), which was a positive effect. The Persian Ezafeh affix might be the explaining factor here, which is an enclitic vowel linked to the head noun as related modifiers and possessor noun phrase (Samvelian, 2007) and has very two types, i.e. adjectival and nominal Ezafeh (Perry, J. R. & Kaye, A. S., 2007), Below a schematic structure of an example of Persian noun phrase is shown:

| N-EZ | AP-EZ | PP-EZ | PP-EZ | PP-EZ | NP(Poss) |
|---|---|---|---|---|---|
| sibha-ye | ghermez-e | abdar-e | bagh-e | pedarbozorg-e | Arash |
| apples-EZ | red-EZ | juicy | ezgarden-EZ | grandpa-EZ | Arash |

‘red juicy apples from Arash’s grandpa’s garden’

Therefore, we could extract that participants showed regression ins for text integration (annexation, suppletion) and comprehension as the very affix links the head noun and the possessor NP (e.g. Arash) together along with the all elements.

With regard to regression out count (ROC), a positive effect of word length on ROC was found. In addition, phonemes on their own proved to affect ROC negatively, as did word frequency.

Finally, during reading of Persian sentences by native Persian readers, various properties of words have been shown to correspond, and in some cases in an interactive manner with each other, with different aspects of eye movements during the reading process. Moreover, most of the findings in this study are in line with the results in previous studies on eye movement measures in reading when word related independent variables have been utilized. Thus, this seems, as per the existing scientific literature on eye movements during reading, to be the first study focusing on eye movement during silent reading of Persian sentences.

There are, however, a number of limitations that warrant mention and consideration. An aspect which may influence the cognitive ability of information processing is the phenomena of bilingualism, or the ability of communicating via more than one language. It has been shown; however not conclusively, that bilinguals may have an advantage over monolinguals in that their cognitive abilities is elevated. The latter case has been true when bilinguals learn a novel language. In a study by Mokhtari Khiyavi et al. (2016) on EFL (English as a Foreign Language) learners, it was shown that Turkish-Persian bilinguals outperformed their Persian monolinguals peers in English reading comprehension. However, Jalalipour et al. (2017) observed that the opposite was true when Arabic-Persian bilingual and Persian monolingual fifth grade peers were reading Persian. In this study, which was focused on reading accuracy and reading comprehension, it was found that the monolinguals in general excelled over their bilingual peers in both areas of reading. Thus, the number of languages which the reader knows and masters influences not only the reading process of texts in one of those languages, but also the reading process while learning a novel language. The first limitation to the present study is related to the number of bilingual subjects, 32 subjects out of 44, in total, were bilingual, which includes native speakers of Persian and Azari. Azeris (an ethnic group of Iranians who live in northern Iran), in Iran, are generally fluent in Persian. However, it is good to consider that Azari is a Western Oghuz Turkic language while Persian is and Indo-European language (Rashidvash, 2012). This could have unintended effects which should be considered for future research. Such a study could focus on the possible effects of bilingualism between two languages from two different language families

might have on the reading process. In the present study any effect due to bilingualism is outside its scope, and hence has not been investigated in this master thesis. However, in future investigations this phenomenon ought to be addressed, and perhaps incorporating various bilingual variations in various studies, such as Azeri-Persian, Kurdish-Persian, Arabic-Persian, or Armenian-Persian speakers along with Persian monolinguals.

The second limitation to the present study is related to generalization of the results; how and to what extent the results are applicable to the Persian language. The stimuli were chosen from a corpus but to what degree it represents the Persian language is unclear.

Thirdly, the study of morphological complexity (affixation) such as suffixes, prefixes, and infixes were beyond the scope of this research but do open up an additional avenue of study of this topic. By considering a target-word approach it opens the possibility to observe morphological aspects. By this follows that word frequencies and predictability tests may be added, in future investigations, in the affixation level. Moreover, it should be noted that this study did not encompass the Corsi Block and Digit Span tasks into the eye movement measures analysis. Hence, future investigations would be better suited to address these tasks.

One of the interests which could be considered in Persian reading studies is comparing its results with a different language which uses the same alphabet. One such language could be Arabic. The Arabic morphology is a non-concatenative morphology as opposed to the Persian morphology which is based on an affixal system. While the Persian language is utilizing the Perso-Arabic alphabet, the Arabic language is using the Arabic alphabet. Both these alphabets have their origin in the Aramaic alphabet, but have evolved in slightly different directions (Healey, J. F. & Smith, G. R., 2012). Hence, it would be constructive to compare the findings of eye movement control during reading from these two different languages, which encompasses different morphologies, while similar orthographical characteristics. Various investigations have been conducted in Arabic. In one of these studies Hermena, Liversedge, and Drieghe (2017) found that native Arabic readers tend to fixate words close to the respective word center, or more precisely prior to it in all conditions including 5 and 7 character long words. Moreover, FLP were more likely to land closer to the beginning of longer words in comparison with shorter words. Consequently, these findings, as well as the results of the present study, are in line with the classic findings.

Moreover, the results of the present study is in line with the findings of Hermena et al. (2017) with regards to first fixation duration and gaze duration, i.e. longer fixations on longer words. The authors also found longer first fixation duration on shorter words, spatially, rather than longer words which are in line with the findings in the present study.

According to what has been described above, it could be stated that findings of two different languages of the same non-Latin alphabet with the same orthography, are both in line with the literature. However, there is a big vacuum, specifically, for

comparative future works which investigates both languages within the same study. Moreover, as a plausible future aims, studies should consider suffixation effects on OVP and IOVP measures by using further paradigms such as the boundary paradigm, as described by Hyönä, Yan, and Vainio (2017) in their study within the Finish language. Such future investigations could illuminate information regarding morphological, typographical and lexical aspects and their effects on eye movement measures in reading.

Nonetheless, it should also be stressed that it would be of high interest for future investigations in similar eye movement experiments as the present, to incorporate and consider words with similar orthography but with differing pronunciations.

This thesis aimed at contributing to reading research by reporting findings of a relatively less-studied language, namely Persian. It is necessary to conduct further studies with broader stimuli, with alternative paradigms (e.g., the target-word approach), as well as with a larger population.

## REFERENCES


Al-Rahayfeh, A., & Faezipour, M. (2013). Eye tracking and head movement detection: A state-of-art survey. *IEEE journal of translational engineering in health and medicine*, *1*.

Amiri, H., Hojjat, H., & Oroumchian, F. (2007). Investigation on a feasible corpus for Persian POS tagging. In *Proceedings of the 12th International CSI Computer Conference (CSICC)*.

Armbruster, H., Bikfalvi, A., Kinkel, S., & Lay, G. (2008). Organizational innovation: The challenge of measuring non-technical innovation in large-scale surveys. *Technovation*, *28*(10), 644-657.

Baayen, H. (2008a). *Analyzing linguistic data: A practical introduction to statistics using R*. Cambridge University Press.

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008b). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of memory and language*, *59*(4), 390-412.

Baluch, B., & Besner, D. (1991). Visual word recognition: Evidence for strategic control of lexical and nonlexical routines in oral reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *17*(4), 644.

Bates, D. (2007). Linear mixed model implementation in lme4. *Manuscript, University of Wisconsin*, *15*.

Bates, D. (2011). Computational methods for mixed models. *Vignette for lme4*.

Bates, D., Maechler, M., Bolker, B., Walker, S., Christensen, R. H. B., Singmann, H., & Rcpp, L. (2015). Package 'lme4'

Bax, S. (2013). The cognitive processing of candidates during reading tests: Evidence from eye-tracking. *Language Testing*, *30*(4), 441-465.

Bell, C. (1823). XV. On the motions of the eye, in illustration of the uses of the muscles and nerves of the orbit. *Philosophical Transactions of the Royal Society of London*, *113*, 166-186.

Biber, D., Conrad, S., & Reppen, R. (1998). *Corpus linguistics: Investigating language structure and use*. Cambridge University Press.

Candy, T. R., Crowell, J. A., & Banks, M. S. (1998). Optical, receptoral, and retinal constraints on foveal and peripheral vision in the human neonate. *Vision research*, *38*(24), 3857-3870.

Carmichael, L. (1926). Sir Charles Bell: A contribution to the history of physiological psychology. *Psychological review*, *33*(3), 188.

Clark, J. J., & O'regan, J. K. (1999). Word ambiguity and the optimal viewing position in reading. *Vision Research*, *39*(4), 843-857.

DeAngelus, M., & Pelz, J. B. (2009). Top-down control of eye movements: Yarbus revisited. *Visual Cognition*, *17*(6-7), 790-811.

Deutsch, A., & Rayner, K. (1999). Initial fixation location effects in reading Hebrew words. *Language and Cognitive Processes*, *14*(4), 393-421.

Duchowski, A. T. (2002). A Breadth-First Survey of Eye Tracking Applications. *Behavior Research Methods, Instruments & Computers (BRMIC)*, *34*(4), 455-470.

Engbert, R., & Kliegl. R. (2003). Chapter 6 - *Binocular Coordination in Microsaccades* (124-140). Cognitive and Applied Aspects of Eye Movement Research

Hayhoe, M. M. (2004). Advances in relating eye movements and cognition. *Infancy*, *6*(2), 267-274.

Healey, J. F., & Smith, G. R. (2012). *A Brief Introduction to the Arabic Alphabet* (Vol. 3). Saqi.

Hermena, E. W., Liversedge, S. P., & Drieghe, D. (2017). The influence of a word's number of letters, spatial extent, and initial bigram characteristics on eye movement control during reading: Evidence from Arabic. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *43*(3), 451.

Hohenstein, S., & Kliegl, R. (2014). Semantic preview benefit during reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *40*(1), 166.

Huey, E. (1908). *The psychology and pedagogy of reading*. New York: Macmillan.

Hyönä, J., & Bertram, R. (2011). Optimal viewing position effects in reading Finnish. *Vision Research*, *51*(11), 1279-1287.

Hyönä, J., Yan, M., & Vainio, S. (2017). Morphological structure influences the initial landing position in words during reading Finnish. *The Quarterly Journal of Experimental Psychology*, *1*-10.

Izadi, S., Sadri, J., Solimanpour, F., & Suen, C. Y. (2008). A review on Persian script and recognition techniques. In *Arabic and Chinese Handwriting Recognition* (pp. 22-35). Springer, Berlin, Heidelberg.

Jagersma, A. H. (2010). *A descriptive grammar of Sumerian* (Doctoral dissertation, Faculty of the Humanities, Leiden University).

Jalalipour, M., Majdinasab, F., Khedri, H., & Chadorian, B. (2018). A Comparison of Reading Accuracy and Reading Comprehension in Bilingual Arabic/Farsi-speaking and Monolingual Farsi-speaking Elementary Students in the Fifth Grade. *Journal of Rehabilitation Sciences and Research.*

Just, M. A., & Carpenter, P. A. (1980). A theory of reading: From eye fixations to comprehension. *Psychological review*, *87*(4), 329.

Khiyavi, B. M., Rostami, G., & Gholami, H. (2016). Effects of Bilingualism and Monoligualism on Iranian EFL Learner s' performance on Reading Comprehension Test. *Journal of Applied Linguistics and Language Research, 3*(4), 225-233.

Kliegl, R., Grabner, E., Rolfs, M., & Engbert, R. (2004). Length, frequency, and predictability effects of words on eye movements in reading. *European Journal of Cognitive Psychology*, *16*(1-2), 262-284.

Kliegl, R., Masson, M. E., & Richter, E. M. (2010). A linear mixed model analysis of masked repetition priming. *Visual Cognition*, *18*(5), 655-681.

Kliegl, R., Wei, P., Dambacher, M., Yan, M., & Zhou, X. (2011). Experimental effects and individual differences in linear mixed models: Estimating the relationship between spatial, object, and attraction effects in visual attention. *Frontiers in Psychology*, *1*, 238.

Kutschera, U. (2009). Charles Darwin's Origin of Species, directional selection, and the evolutionary sciences today. *Naturwissenschaften*, *96*(11), 1247.

Moll, K., & Jones, M. (2013). Naming fluency in dyslexic and nondyslexic readers: Differential effects of visual crowding in foveal, parafoveal, and peripheral vision. *The Quarterly Journal of Experimental Psychology*, *66*(11), 2085-2091.

Morrison, R. E. (1984). Manipulation of stimulus onset delay in reading: evidence for parallel programming of saccades. *Journal of Experimental psychology: Human Perception and performance*, *10*(5), 667.

Nilsson, M. (2012). *Computational Models of Eye Movements in Reading: A Data-driven Approach to the Eye-mind Link*(Doctoral dissertation, Acta Universitatis Upsaliensis).

Nuthmann, A., Engbert, R., & Kliegl, R. (2005). Mislocated fixations during reading and the inverted optimal viewing position effect. *Vision research*, *45*(17), 2201-2217.

Nuthmann, A., & Kliegl, R. (2009). Preferred viewing locations: A validation and an extension. *Perception*, *38*(6), 901-902.

O'Regan, J.K. (1979). Eye guidance in reading: Evidence for the linguistic control hypothesis. *Perception & Psychophysics, 25*, 501-509.

Perry, J. R., & Kaye, A. S. (2007). Persian morphology. *Morphologies of Asia and Africa*, *2*, 975-1019.

Purves, D., Augustine, G. J., Fitzpatrick, D., Hall, W. C., Lamantia, A. S., Mcnamara, J. O., ... & Dackow, R. J. (2005). A Classic Guide to the Neurophysiological Foundations of Emotions and Behaviors. *Psyccritiques*, *50*(7).

Rashbass, C. (1961). The relationship between saccadic and smooth tracking eye movements. *The Journal of Physiology*, *159*(2), 326-338.

Rashidvash, V. (2012). The Iranian and Azari languages. *Research on Humanities and Social Sciences*, *2*(5), 28-36.

Rayner, K. (1979). Eye guidance in reading: Fixation locations within words. *Perception*, *8*(1), 21-30.

Rayner, K., & Raney, G. E. (1996). Eye movement control in reading and visual search: Effects of word frequency. *Psychonomic Bulletin & Review*, *3*(2), 245-248.

Rayner, K. (1998). Eye movements in reading and information processing.

Rayner, K., Pollatsek, A., Ashby, J., & Clifton Jr, C. (2012). *Psychology of reading*. Psychology Press.

Reichle, E. D., Pollatsek, A., Fisher, D. L., & Rayner, K. (1998). Toward a model of eye movement control in reading. *Psychological review*, *105*(1), 125.

Reichle, E. D., Rayner, K., & Pollatsek, A. (2003). The EZ Reader model of eye-movement control in reading: Comparisons to other models. *Behavioral and brain sciences*, *26*(4), 445-476.

Reichle, E. D. (2006). Computational models of eye-movement control during reading: Theories of the eye-mind link. *Cognitive Systems Research*, *7*(1), 2-3.

Richardson, D. C., & Spivey, M. J. (2004). Eye tracking: Research areas and applications. *Encyclopedia of biomaterials and biomedical engineering*, 573-582.

Samvelian, P. (2007). A (phrasal) affix analysis of the Persian Ezafe. *Journal of Linguistics*, *43*(3), 605-645.

Tatler, B. W., & Wade, N. J. (2003). On nystagmus, saccades, and fixations. *Perception*, *32*(2), 167-184.

Taylor, W. L. (1953). "Cloze procedure": A new tool for measuring readability. *Journalism Bulletin*, *30*(4), 415-433.

Team, R. C. (2015). R: A language and environment for statistical computing. Vienna, Austria; 2014. URL http://www. R-project. org.

Thibadeau, R., Just, M. A., & Carpenter, P. A. (1982). A model of the time course and content of reading. *Cognitive Science*, *6*(2), 157-203.

Valladas, H., Clottes, J., Geneste, J. M., Garcia, M. A., Arnold, M., Cachier, H., & Tisnérat-Laborde, N. (2001). Palaeolithic paintings: evolution of prehistoric cave art. *Nature*, *413*(6855), 479.

Vitu, F., McConkie, G. W., Kerr, P., & O'Regan, J. K. (2001). Fixation location effects on fixation durations during reading: An inverted optimal viewing position effect. *Vision Research*, *41*(25-26), 3513-3533.

Wade, N., & Tatler, B. W. (2005). *The moving tablet of the eye: The origins of modern eye movement research*. Oxford University Press, USA.

Wade, N. J., & Tatler, B. W. (2009). Did Javal measure eye movements during reading?. *Journal of Eye Movement Research*, *2*(5).

Wade, N. J. (2010). Pioneers of eye movement research. *i-Perception*, *1*(2), 33-68.

Wade, N. J., & Tatler, B. W. (2011). Origins and applications of eye movement research. *Oxford handbook on eye movements*, 17-46.

Walczyk, J. J., Tcholakian, T., Igou, F., & Dixon, A. P. (2014). One hundred years of reading research: Successes and missteps of Edmund Burke Huey and other pioneers. *Reading psychology*, *35*(7), 601-621.

Wang, S. P. (2005). Corpus-based approaches and discourse analysis in relation to reduplication and repetition. *Journal of Pragmatics*, *37*(4), 505-540.

Winter, B. (2013). Linear models and linear mixed effects models in R with linguistic applications. *arXiv preprint arXiv:1308.5499*.

Wotschack, C. (2009). *Eye movements in reading strategies: How reading strategies modulate effects of distributed processing and oculomotor control* (Vol. 1). Universitätsverlag Potsdam.

Yan, M., Zhou, W., Shu, H., Yusupu, R., Miao, D., Krügel, A., & Kliegl, R. (2014). Eye movements guided by morphological structure: Evidence from the Uighur language. *Cognition*, *132*(2), 181-215.

YARBUS, A. L. (1967). *Eye Movements and Vision*, pp. 23-25, Plenum Press, New York.

# APPENDICES

## Appendix A

Table A.3 Means (standard deviations) for eye movement measures according to word length

| WL | N | FFD | FRFC | GD | FLP | RIC | ROC |
|---|---|---|---|---|---|---|---|
| 1 | 21 | 182(59) | 1(0.07) | 182(59.7) | 1(0) | 0.12(0.34) | 0.09(0.28) |
| 2 | 154 | 196(63) | 1(0.17) | 201(75.1) | 1.5(1) | 0.11(0.36) | 0.1(0.31) |
| 3 | 99 | 229(79) | 1.1(0.38) | 255(107) | 1.8(1.1) | 0.15(0.41) | 0.15(0.42) |
| 4 | 182 | 229(79) | 1.1(0.43) | 264(113) | 2.2(1.2) | 0.19(0.48) | 0.14(0.4) |
| 5 | 166 | 229(77) | 1.2(0.48) | 274(119) | 2.6(1.4) | 0.22(0.53) | 0.14(0.41) |
| 6 | 89 | 230(76) | 1.3(0.55) | 295(131) | 2.7(1.5) | 0.24(0.56) | 0.17(0.45) |
| 7 | 48 | 225(72) | 1.4(0.59) | 311(136) | 2.8(1.5) | 0.27(0.57) | 0.11(0.35) |
| 8 | 25 | 225(74) | 1.5(0.66) | 336(142) | 2.7(1.6) | 0.4(0.72) | 0.17(0.42) |
| 9 | 18 | 225(69) | 1.8(0.71) | 379(149) | 2.9(1.6) | 0.23(0.52) | 0.12(0.35) |
| 10 | 10 | 215(61) | 1.8(0.72) | 368(156) | 3.1(1.7) | 0.19(0.47) | 0.1(0.31) |
| 11 | 4 | 210(58) | 2.3(0.77) | 466(148) | 2.4(1.6) | 0.2(0.44) | 0.04(0.21) |
| 12 | 2 | 222(51) | 2.3(0.76) | 499(152) | 3.4(1.5) | 0.18(0.42) | 0.1(0.3) |
| 13 | 1 | 232(51) | 2(0.75) | 440(170) | 2.1(1.4) | 0.21(0.47) | 0.21(0.41) |

Note: WL: word length (characters including zero-width space as an independent character); N: number of words; FFD: first run fixation duration; FRFC: first run fixation count; GD: gaze duration; FLP: first fixation landing position; RIC: regression in count; ROC: regression out count.

## Appendix B

Table B.4 Linear mixed model estimates for first fixation landing position (FLP)

| Fixed effects | Estimate | SE | t-Value |
|---|---|---|---|
| Grand Mean(GM) | 2.488 | 0.05 | 47.90 |
| Word Length (WL) | 0.636 | 0.06 | 9.43 |
| Frequency (WF) | 0.103 | 0.02 | 4.88 |
| Predictability (WP) | 0.02 | 0.02 | 0.89 |
| Word Type (WT) | 0.06 | 0.05 | 1.24 |
| Phoneme (PHON) | 0.08 | 0.07 | 1.11 |
| WL × WF | 0.03 | 0.03 | 1.21 |
| WL × WP | 0.04 | 0.03 | 1.19 |
| WL × WT | 0.13 | 0.06 | 2.21 |
| WL × PHON | -0.03 | 0.05 | -0.60 |

| Random Effects | Variance | SD | Correlation | | | |
|---|---|---|---|---|---|---|
| Words (GM) | 0.128 | 0.35 | | | | |
| Sentences (GM) | 0.004 | 0.069 | | | | |
| Subjects (GM) | 0.098 | 0.313 | | | | |
| Subjects- WL | 0.031 | 0.177 | 0.94 | | | |
| Subjects- WF | 0.002 | 0.053 | -0.15 | 0.20 | | |
| Subjects- WT | 0.011 | 0.108 | -0.02 | -0.27 | -0.74 | |
| Subjects- WP | 0.0001 | 0.012 | -0.91 | -0.87 | 0.08 | -0.12 |
| Subjects- PHON | 0.008 | 0.091 | 0.75 | 0.64 | -0.29 | 0.47 |
| Residual | 1.555 | 1.247 | | | | |

Table B.5 Linear mixed model for first fixation duration (FFD)

| Fixed Effects | Estimate | SE | t-Value | | | |
|---|---|---|---|---|---|---|
| Grand Mean (GM) | 5.344 | 0.018 | 285 | | | |
| Word Length (WL) | -0.052 | 0.016 | -3.15 | | | |
| Frequency (WF) | -0.048 | 0.005 | -9.69 | | | |
| Predictability (WP) | -0.001 | 0.007 | -0.26 | | | |
| Word Type (WT) | -0.015 | 0.012 | -1.25 | | | |
| Phoneme (PHON) | 0.015 | 0.017 | 0.88 | | | |
| First Fixation Landing Position (FLP) | 0.014 | 0.001 | 9.43 | | | |
| WL × WF | -0.007 | 0.007 | -0.97 | | | |
| WL × WP | 0.007 | 0.008 | 0.87 | | | |
| WL × WT | -0.046 | 0.014 | -3.13 | | | |
| WL × PHON | -0.076 | 0.014 | -5.26 | | | |
| WL × FLP | 0.006 | 0.002 | 2.38 | | | |
| **Random Effects** | **Variance** | **SD** | **Correlation** | | | |
| Words (GM) | 7.465 | 0.086 | | | | |
| Sentences (GM) | 4.341 | 0.006 | | | | |
| Subjects (GM) | 1.367 | 0.116 | | | | |
| Subjects- WL | 6.689 | 0.025 | -0.30 | | | |
| Subjects- WF | 1.266 | 0.011 | -0.25 | -0.85 | | |
| Subjects- WF | 4.812 | 0.021 | -0.10 | 0.54 | -0.50 | |
| Subjects- WT | 2.968 | 0.017 | 0.30 | -0.42 | 0.26 | -0.54 |
| Subjects- PHON | 6.620 | 0.025 | -0.14 | -0.66 | 0.74 | -0.92 |
| Residual | 1.015 | 0.318 | | | | |

Note: WL: log2 values of word length; WF: log10 values of word frequency; WP: logit transformed values of word predictability; WT: word type including opacity and transparency; PHONE: log2 values of phonemes; FLP: first fixation landing position. Log Likelihood: -8631.8; REML Deviance: 17264; Number of observations: 28998, N of sentences: 99, N of subjects: 44.

Table B.6 Linear mixed model for First Run Fixation Count (FRFC)

| Fixed Effects | Estimate | SE | t-Value | | | |
|---|---|---|---|---|---|---|
| Grand Mean (GM) | 5.673 | 0.021 | 260 | | | |
| Word Length (WL) | 0.237 | 0.022 | 10.44 | | | |
| Frequency (WF) | -0.087 | 0.006 | -13.08 | | | |
| Predictability (WP) | -0.027 | 0.009 | -3.01 | | | |
| Word Type (OTZT) | -0.020 | 0.018 | -1.14 | | | |
| Phoneme (PHON) | 0.025 | 0.024 | 1.04 | | | |
| First Fixation Landing Position (FLP) | -0.058 | 0.001 | -30.81 | | | |
| WL × WF | -0.033 | 0.010 | -3.34 | | | |
| WL × WP | -0.009 | 0.010 | -0.89 | | | |
| WL × OTZT | -0.023 | 0.020 | -1.14 | | | |
| WL × PHON | 0.005 | 0.018 | 0.29 | | | |
| WL × FLP | -0.029 | 0.003 | -9.02 | | | |
| Random Effects | Variance | SD | Correlation | | | |
| Words (GM) | 0.0127 | 0.112 | | | | |
| Sentences (GM) | 0.0003 | 0.019 | | | | |
| Subjects (GM) | 0.0180 | 0.134 | | | | |
| Subjects- WL | 0.0032 | 0.056 | 0.25 | | | |
| Subjects- WF | 0.0002 | 0.017 | -0.32 | -0.09 | | |
| Subjects- WP | 0.0007 | 0.028 | -0.23 | 0.37 | 0.28 | |
| Subjects- OTZT | 0.0004 | 0.02 | -0.90 | - | 0.38 | - |

| | | | | | | |
|---|---|---|---|---|---|---|
| | | 1 | | 0.51 | | 0.14 |
| Subjects- PHON | 0.0021 | 0.04 | -0.24 | - | - | - |
| | | 5 | | 0.65 | 0.10 | 0.85 |
| Residual | 0.1494 | 0.38 | | | | |
| | | 6 | | | | |

Note: WL: log2 values of word length; WF: log10 values of word frequency; WP: logit transformed values of word predictability; OTZT: word type including opacity and transparency including Persian Ezafeh Suffix; PHONE: log2 values of phonemes; FLP: first fixation landing position. Log Likelihood: -14313; REML Deviance: 28625; Number of observations: 28998, N of sentences: 99, N of subjects: 44.

Table B.7 Linear mixed model for Regression in Count (RIC)

| Fixed Effects | Estimate | SE | t-Value | | | |
|---|---|---|---|---|---|---|
| Grand Mean (GM) | 0.199 | 0.018 | 10.64 | | | |
| Word Length (WL) | -0.034 | 0.026 | -1.31 | | | |
| Frequency (WF) | -0.022 | 0.008 | -2.60 | | | |
| Predictability (WP) | -0.076 | 0.011 | -6.46 | | | |
| Word Type (WT) | 0.053 | 0.021 | 2.49 | | | |
| Phoneme (PHON) | 0.080 | 0.032 | 2.44 | | | |
| WL × WF | -0.004 | 0.013 | -0.34 | | | |
| WL × WP | -0.049 | 0.014 | -3.37 | | | |
| WL × WT | 0.019 | 0.025 | 0.75 | | | |
| WL × PHON | -0.008 | 0.024 | -0.32 | | | |
| Random Effects | Variance | SD | Correlation | | | |
| Words (GM) | 0.0235 | 0.153 | | | | |
| Sentences (GM) | 0.0039 | 0.062 | | | | |
| Subjects (GM) | 0.0105 | 0.102 | | | | |
| Subjects- WL | 0.0007 | 0.026 | -0.60 | | | |
| Subjects- WF | 0.0001 | 0.012 | -0.22 | -0.29 | | |
| Subjects- WP | 0.0006 | 0.026 | -0.95 | 0.66 | 0.38 | |
| Subjects- WT | 0.0008 | 0.029 | 0.99 | -0.59 | -0.16 | -0.93 |
| Subjects- PHON | 0.0068 | 0.082 | 0.88 | -0.77 | 0.26 | -0.76 |

| | | |
|---|---|---|
| Residual | 0.2166 | 0.465 |

Note: WL: log2 values of word length; WF: log10 values of word frequency; WP: logit transformed values of word predictability; WT: word type including opacity and transparency; PHONE: log2 values of phonemes. Log Likelihood: -19729; REML Deviance: 39458; Number of observations: 28998, N of sentences: 99, N of subjects: 44.

Table B.8 Linear mixed model for Regression out Count (ROC)

| Fixed Effects | Estimate | SE | t-Value |
|---|---|---|---|
| Grand Mean (GM) | 0.230 | 0.012 | 18.18 |
| Word Length (WL) | 0.050 | 0.023 | 2.13 |
| Frequency (WF) | -0.014 | 0.007 | -1.94 |
| Predictability (WP) | 0.018 | 0.010 | 1.71 |
| Word Type (WT) | -0.025 | 0.018 | -1.40 |
| Phoneme (PHON) | -0.071 | 0.026 | -2.68 |
| First Fixation Landing Position (FLP) | -0.026 | 0.001 | -14.48 |
| WL × WF | -0.025 | 0.011 | -2.26 |
| WL × WP | 0.011 | 0.012 | 0.97 |
| WL × WT | -0.007 | 0.021 | -0.33 |
| WL × PHON | -0.075 | 0.020 | -3.58 |
| WL × FLP | 0.004 | 0.003 | 1.43 |
| Random Effects | Variance | SD | Correlation |
| Words (GM) | 0.018 | 0.135 | |
| Sentences (GM) | 0.000 | 0.000 | |
| Subjects (GM) | 0.003 | 0.06 | |

| | | | | | | |
|---|---|---|---|---|---|---|
| Subjects- WL | 0.001 | 2<br>0.041 | 0.83 | | | |
| Subjects- WF | 0.000 | 0.015 | -0.39 | -0.20 | | |
| Subjects- WP | 0.001 | 0.036 | 0.37 | 0.75 | -0.12 | |
| Subjects- WT | 0.001 | 0.032 | -0.66 | -0.67 | 0.79 | -0.38 |
| Subjects- PHON | 0.002 | 0.053 | -0.72 | -0.90 | 0.49 | -0.65 |
| Residual | 0.140 | 0.375 | | | | |

## Appendix C

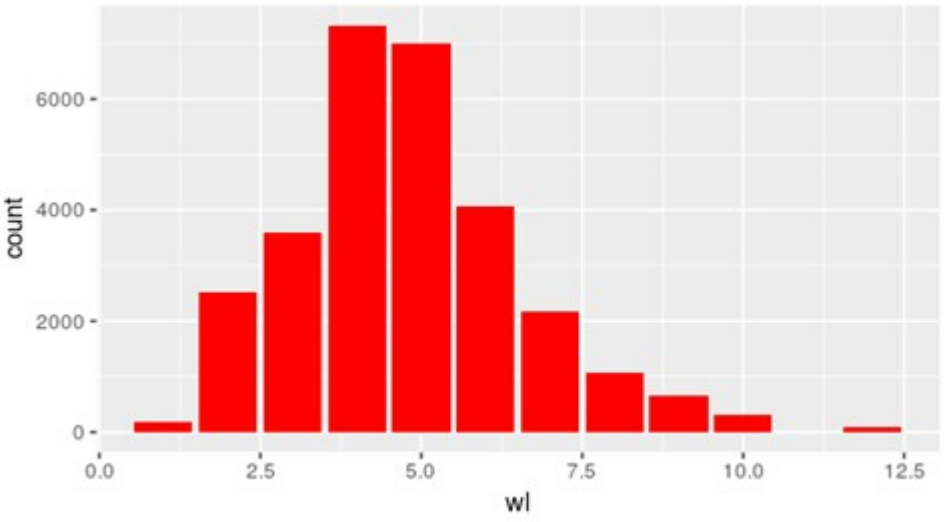


Figure C.27 X axis shows word length,

y axis shows how many times each word length occurs in the dataset. The distribution is right-skewed which means the mean is on the right side of the peak or median. We should apply log2 transformation on WL values to resemble it a normal distribution.

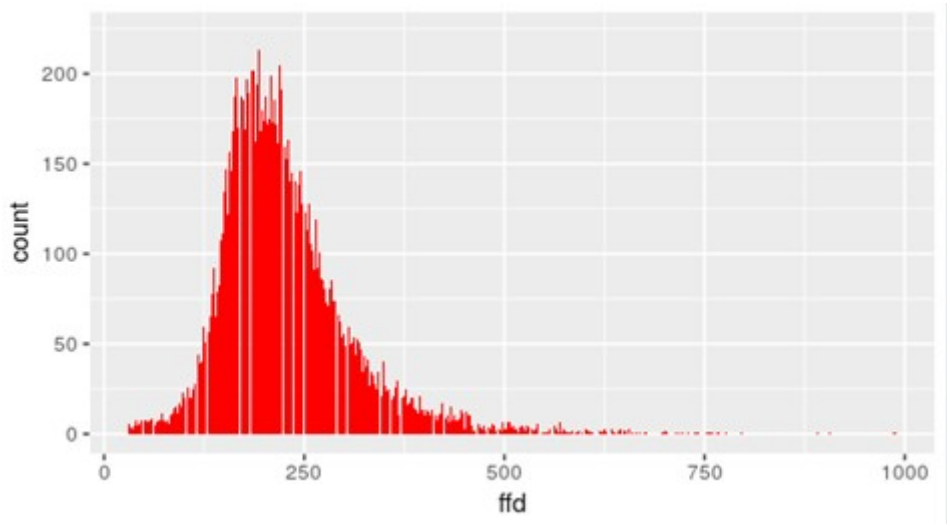


Figure C.28 X axis shows each word's FFD value,

y axis shows how many times each first fixation occurs in the dataset. The distribution is right-skewed which means the mean is on the right side of the peak or median. We should apply natural logarithm transform for FFD values to resemble it a normal distribution.

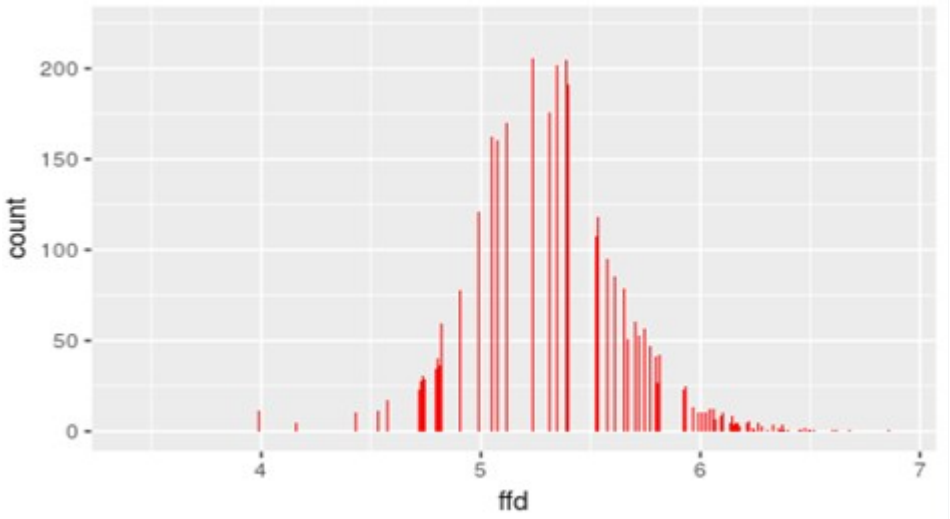


Figure C.29 Histogram of FFD after applying natural logarithm transformation

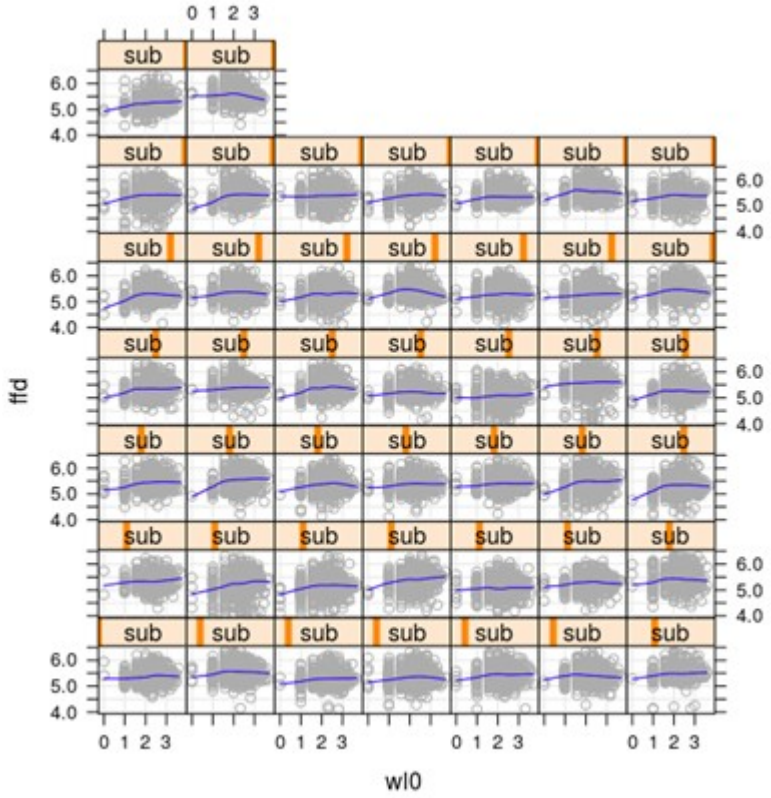


Figure C.30 Distributions of FFD

(after natural log transformations) among subjects. It shows how each subject responds to the same task. We could say that there is a high variance between subjects. Therefore, we prefer Linear Mixed Model to Linear Regression since the former helps us to assign unique intercept & slope to each subject.

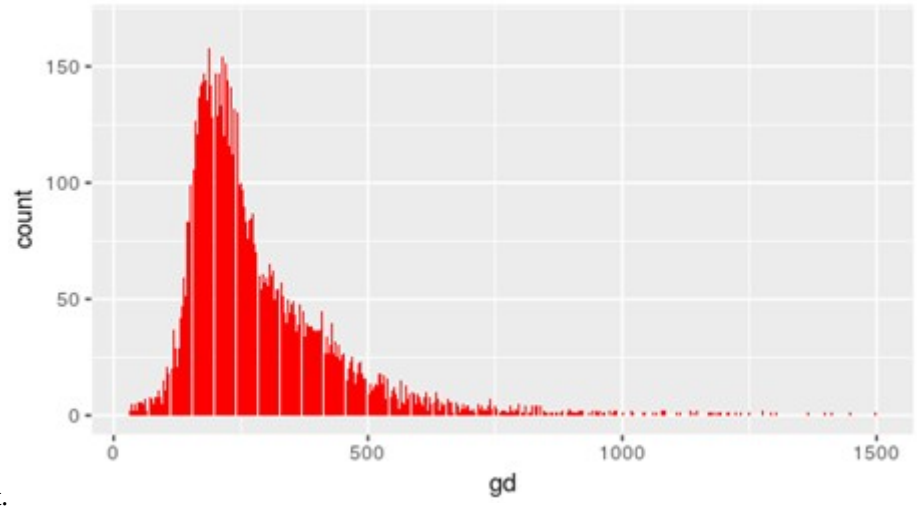


Figure C.31 Histogram of GD values before applying natural log transformation.

X axis shows GD values, y axis shows how many times each GD occurs in the dataset. The distribution is right-skewed which means the mean is on right side of the peak or median. We should apply natural log transformation on GD variable to resemble it a normal distribution.

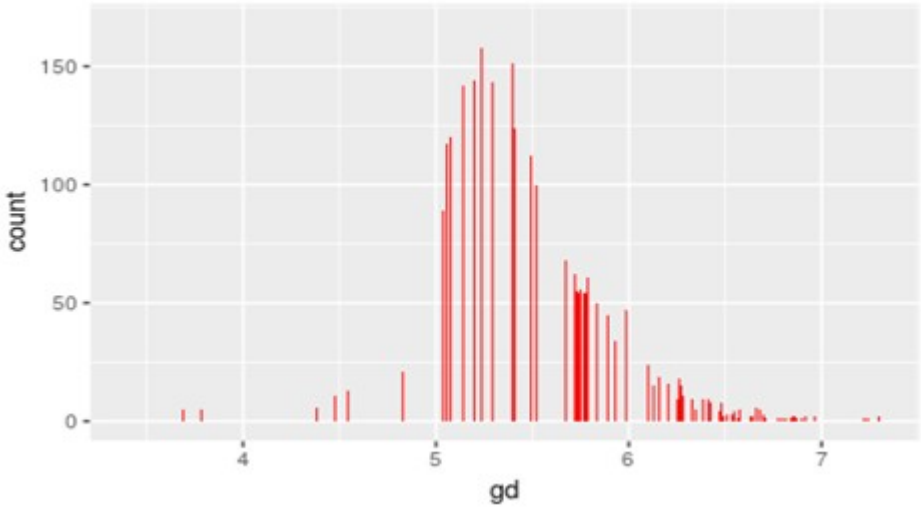

Figure C.32 Histogram of GD after natural log transformation.

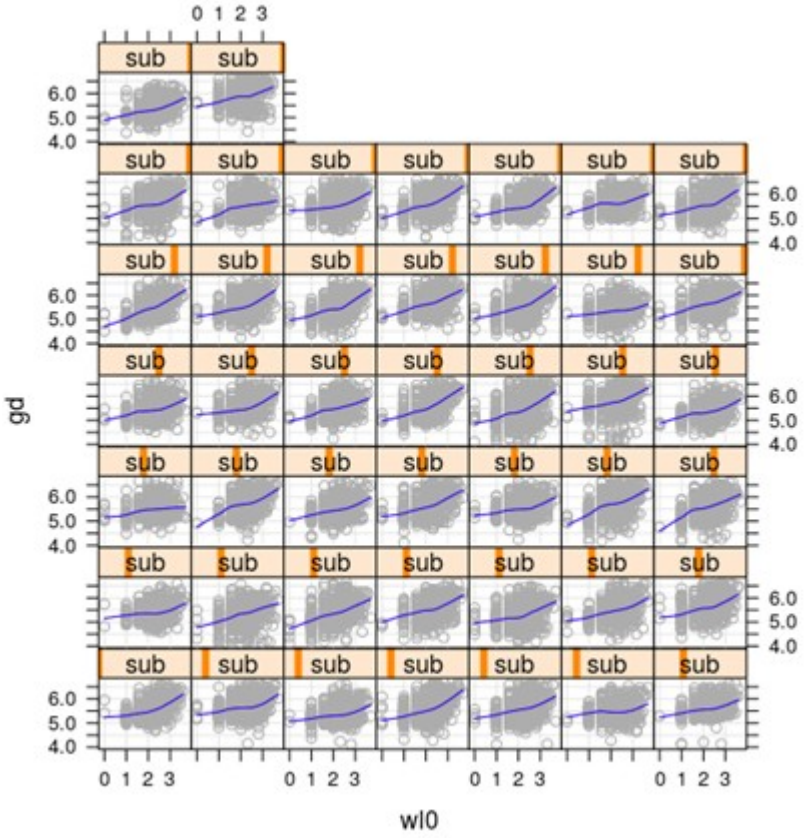


Figure C.33 Distributions of GD (after log transformations) among subjects.

It shows how each subject responds to the same task. We could say that there is a high variance between subjects. Therefore, we prefer Linear Mixed Model to Linear Regression since the former one helps us to assign unique intercept & slope to each subject

.

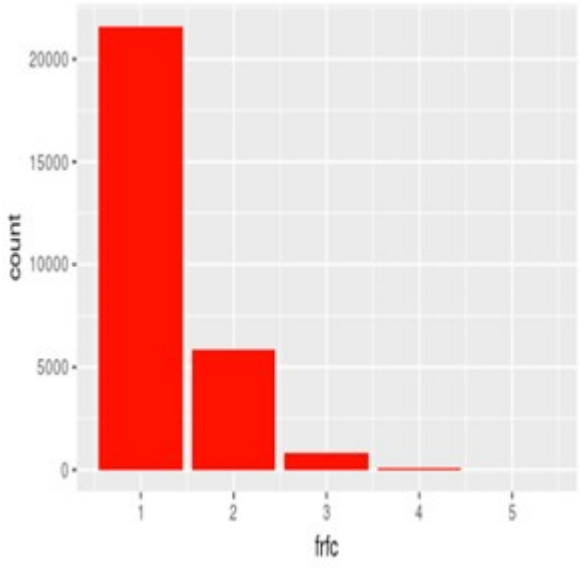


Figure C.34 X axis shows First Run Fixation Count,

y axis shows how many times each FRFC occurs in the dataset.

.

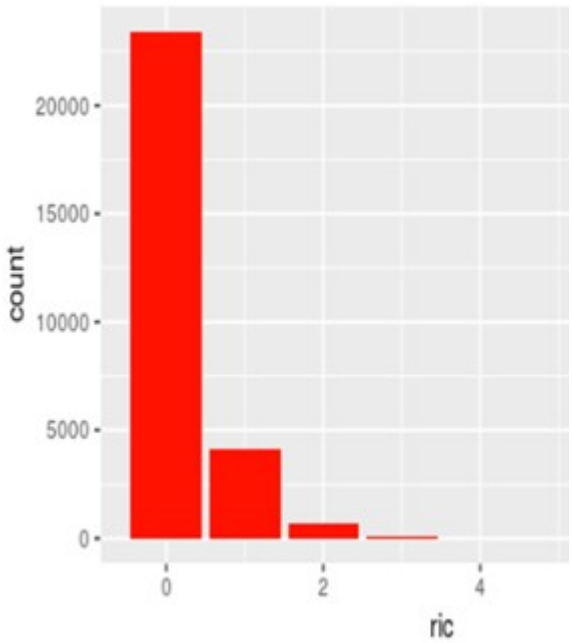


Figure C.35 X axis shows Regression in Count,

y axis shows how many times each RIC occurs in the dataset.

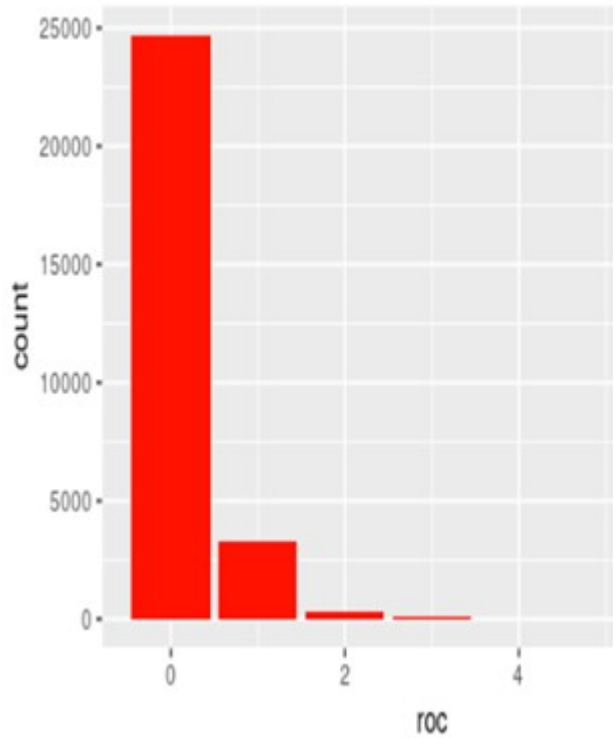


Figure C.36 X axis shows Regression out Count,

y axis shows how many times each ROC occurs in the dataset.

## Appendix D

### Sentences used in the predictability and eye tracking experiments

1تماشای برنامههای مذهبی ارتباط مثبتی با حضور در کلیسا داشت.
Watching religious programs linked positively with attendance at the church.

2.هفته فیلم جمهوری اسلامی ایران در مسکو فردا به پایان می رسد.

Islamic republic of Iran’s film week in Moscow comes to an end tomorrow.

3.فصل بهار امسال در اغلب ایالات آمریکا بسیار زودگذر بوده است.

Spring season in the United States of America fleeted this year.

4.تلسکوپ جدیدی برای مشاهده دقیق سیاهچالهها طراحی شده است.

A new telescope has been designed to observe black holes.

5.درمان ریزش مو در دوره یائسگی همان درمان علائم یائسگی است.

The cure for hair loss at menopausal period is the same as cure for menopausal symptoms.

6.از مصرف مقادیر زیاد شیرینیها به خصوص شکر خودداری کنند.

Avoid consuming excessive sweets especially sugar.

7.بیماریهای واگیردار ممکن است اثرات قابل توجهی داشته باشند.

Contagious diseases could have remarkable effects.

8.آب انگور از نظر خواص به شیر مادر بسیار نزدیک است.

Grapes juice is very close to breastfed milk in the terms of virtues.

9.پلیس خیلی زود مردم را از باند فرودگاه دور کرد.

Police removed people from the airport band quickly.

10.با حرص و اشتیاق به گلهای میخک صحرائی نگاه میکنم.

I am watching cross-country gillyflowers enthusiastically.

11.آمار بیماران قلبی در ایران در سطح بالایی است.

Heart disease patients’ statistics are high in Iran.

.12یکی از نشانههای بیماری پارکینسون لرزش شدید دستها است.

.One of the signs of Parkinson is drastic shaking hands

.13زلزله شدیدی یکی از نقاط دورافتاده اندونزی را تکان داد.

.A strong earthquake fluttered one of the remote areas in Indonesia

.14تنبیه بدنی کودکان به رشد روانی و عاطفی کودک آسیب میرساند.

Corporal punishment of children harms mental and emotional .growth of the kid

.15دود سیگار باعث جهش ژنتیکی منجر به سرطان میشود.

.Cigarette smoke causes genetically mutation which leads to cancer

.16دود سیگار اصلیترین عامل ایجادکننده سرطان در محل کار است.

.Cigarette smoke is the main cause for cancer at the work place

.17بسیاری از ایرانیان مقیم آمریکا تصورات نادرستی از ایران امروز دارند.

Most of the people living in the USA have false beliefs about .modern Iran

.18علمی کردن ورزش در دستور کار سازمان تربیت بدنی قرار دارد.

.Physical education organization is up to make sports scientific

.19اصلاحات به معنی تغییر و تحول در شیوههای حکومت کردن است.

.Reformation means to change the way of governing

.20غذاهای پخته و فاسدشدنی را بلافاصله بعد از مصرف داخل یخچال بگذارید.

Put the cooked and putrefying food in the fridge immediately after .consuming

.21پرخاشگری و عصبانیت در جوانان منجر به بروز تصلب شرایین میشود.

.Anger leads to arteriosclerosis among the young people

.22فشار کار به اندازه استعمال دخانیات برای سلامتی مضر است.

.Work pressure harms health as much as smoking tobacco

.23تدریس زبان فارسی و ایرانشناسی در آمریکا بنیانی استوار دارد.

Persian language and Iranology  tuition have an strong basis in the USA

.24زبانهایی مانند فارسی سالها در دانشگاههای آمریکا تدریس میشد.

Languages like Persian has been being taught in American .universities for years

.25کارشناسان علت سقوط هلیکوپتر را نقص فنی اعلام کردند.

.Experts declared technical defect is the cause of helicopter fall

.26سازمان بهداشت جهانی گزارشی درباره متوسط طول عمر در کشورها منتشر کرد.

World Health Organization published a report on lifetime in .countries

.27روابط نزدیک تهران و ریاض به سود صلح جهانی است.

.Close relationship between Tehran and Riyadh helps world peace

.28ضعف اعتماد به نفس یکی از متداولترین مشکلات روانی در جامعه است.

Low self-confidence is one of the common psychological problems .in the society

.29اشتیاق به وزنه برداری در بین جوانان ایران بسیار وجود دارد.

.Enthusiasm for weight lifting is high among Iranian youth

.30دوران بارداری و زایمان گورخر در ماه و فصل مخصوصی انجام نمیشود.

Fertility and childbirth of Zebras does not happen in especial month .and season

.31رسیدگی در سازمان قضایی بر اساس آیین دادرسی کیفری سابق است.

.Adjudication follows the former rules of penalty

.32دانشمندان اسلامی با مطالعه متون طب قدیمی یونان آنها را کامل کردند.

Islamic scientists completed the ancient Greek medical texts by .studying them

.33در واقع طب اسلامی مبانی طب نوین غربی را پایهریزی کرده است.

In fact Islamic medicine created the modern western medicine .basis

.34استعمال سیگار مهمترین عامل قابل پیشگیری بیماریهای لثه و دهان است.

Smoking cigarettes is one of the preventable causes of mouth and .gum diseases

.35بیشتر مردم بدون یارانههای غذایی دولتها حتی نمیتوانند نفس بکشند.

Most of the people cannot survive without government nutritive .subsidies

.36علم پزشکی همیشه با شکستن تابوهای اجتماعی پیشرفت کرده است.

.Medical science has improved with breaking social taboos

37.رییسجمهوری مساله امنیت را مساله مهمی برای کشورهای منطقه دانست.

The presidency declared the safety issue as an important issue for the region’s countries.

38.ازبکستان خواهان ارتقای روابط با جمهوری اسلامی ایران است.

Uzbekistan wants to promote relationships with Iran.

39.انقلاب ارتباطات تقریبا سراسر جهان را تحت پوشش قرار داده است.

Revolution of communication influenced all of the world.

40.ایران وارد مرحله جدیدی از روابط خود با کشورهای جهان شده است.

Iran has trod to a new level of its relationships with other countries.

41.خاخامهای آمریکایی برای متهمان به جاسوسی تقاضای عفو میکنند.

American rabbis have demanded amnesty for the accused people for spying

42.فشار خون بالا یا بیماری بیعلامت چندان هم بیعلامت نیست.

High blood pressure or the unsigned disease is not that unsigned.

43.طرح ممنوعیت ورود نیروهای مسلح به دانشگاهها تقدیم مجلس میشود.

The idea of prohibition of entering armed forces to the universities offered to the parliament.

44.متاسفانه در برخی مناطق خانوادهها مانع تحصیل فرزندان خود میشوند.

Unfortunately in some regions families hold their children up from education.

45.آمیزش ایرانیها با نواحی آسیای صغیر از زمان کوروش آغاز شده بود.

Hybridizing of Iranians with the region of Anatolia started from Cyrus era.

46. تئاتر ایران نسبت به تئاترهای کشورهای مختلف به هیچ وجه عقبمانده نیست.

Iranian theatre in comparison with other countries’ theatre is not retard nowise.

47.ادبیات باستانی در قرون وسطای اروپا به فراموشی سپرده شده بود.

Ancient literature was called off in medieval era of Europe.

48.در یونان و روم باستان ماه را الهه جوانی میپنداشتند.

They considered moon as youth goddess in ancient Greece and Rome.

.49بقیه داستان زندگی بلال را در کتاب صدای زلال بلال بخوانید.

Read the rest of the story of Bilal’s life in the book of limpid voice of .Bilal

.50خورشید از مشرق طلوع و در مغرب غروب میکند.

.The sun rises from the east and sets in the west

.51ماموران پلیس علت تصادف را ترکیدن لاستیک اتوبوس ذکر کردند.

The policemen indicated the burst of bus’s tire as the cause of .crash

.52این انقلاب یک تجربه جدید و یک حرف نو در دنیاست.

.This revolution is a new experience and utterance in the world

.53همین قانون اساسی مجلس را رکن مهم این نظام قرار داده است.

This constitution makes the parliament an important pillar of the .regime

.54مطامع استعماری در زیر پوشش مسیحیت پنهان شده است.

.Colonization greediness hides in cover of Christianity

.55شمار دیدارهای مسابقات فوتبال قهرمانی باشگاههای اروپا کاهش مییابد.

Number of European football clubs’ championship matches is .decreasing

.56 بانک جهانی یک سرطان در حال رشد در جهان سوم است.

.World Bank is a growing cancer in third world countries

.57مردم به جان آمده در جستجوی روزنهای برای آزادی بودند.

.Exhausted people are looking for an eyelet for freedom

.58مسافران نجات یافته هواپیما زندگی خود را مدیون خلبان هواپیما میدانند.

The surviving airplane passengers owe their lives to the airplane’s .captain

.59کاراتهکارهای دانشجوی ایران در پیکارهای قهرمانی کشور شرکت میکنند.

Iranian student karate players are going to take part in national .championship competitions

.60چین و کره از لحاظ سرعت پیشرفت میتوانند غافلگیرکننده بشوند.

China and Korea in the terms of improvement rate could be .surprising

.61انبوه جمعیت شرق اگر به تکنیک دسترسی پیدا کنند سیلآسا میگردند.

Plenty of east crowd could turn to be forceful if they reach technique.

.62سوخت کشتی ها از پالایشگاه های بندرعباس و آبادان تامین خواهد شد

Ships’ fuel will be provided from Bandar Abbas and Abadan refineries.

.63 سوخت رسانی به کشتی ها لطمه ای به محیط زیست وارد نخواهد کرد.

.Fuelling the ships will not harm the environment

.64رنگ قرمز در نظرِ مگس به صورت سیاه ظاهر میشود.

.The red colour will be shown as black for a fly

.65درصد بسیار ناچیزی از چاقی ناشی از غدد مترشحه داخلی هستند.

.Negligible percent of obesity emerge from parathyroid

.66جوشانده سبزی باعث جلوگیری از تهوع و سرگیجه میشود.

.Sodden vegetable prevent nausea and vertigo

.67عامه همواره با یک مشت اندیشههای محدود آسوده تر زندگی میکرده است.

.The public has always lived comfortable with narrow thoughts

.68طبیعت بیجان برای اولین بار بر صفحه تلویزیون جان گرفت.

The breathless nature was animated for the first time on the TV screen.

.69مقامهای نفتی کویتی در پالایشگاه حالت فوقالعاده اعلام کردهاند.

The petrol officials of Kuwait declared extraordinary mode in the refinery.

.70اروپا در قبال پدیده مهاجرت سیاست واحدی اتخاذ میکند.

.The Europe has taken united politics against emigration issue

.71امروزه در جراحی لیزری بیشتر به لایه برداری پوست توجه میشود.

Nowadays in laser surgeries mostly they pay attention to skin exfoliating.

.72قوه قضاییه میتواند مظهر عدالت در نظام اسلامی باشد.

.Judicature could be the symbol of justice in Islamic regime

.73نظام قضایی جمهوری اسلامی ساختاری مخصوص به خود دارد.

.The Islamic Republics judicature has a unique structure

.74لیموترش ها را در یک ظرف پر آب قرار دهید تا تازه بمانند.

.Put the limes in container full of water to keep them fresh

.75همیشه گلابی را نیم ساعت قبل از صرف غذا میل کنید.

.Always have the pear half an hour before the meal

.76انتخاب کلمات مناسب از مهمترین رموز موفقیت در سخن گفتن است.

The choice of decent words plays an important role in giving a .speech

.77اقتصاد ورشکسته رهبران کره شمالی را وادار به تجدید نظر کرد.

.Bankrupted economy made the North Korean leaders revise

.78بسیاری از مسائل بشر با گفت و شنود حل شدهاند.

.Most of human’s problems has met a solution by dialogue

.79بهترین راه پیشگیری از سرطان مثانه نزد سیگاریها ترک سیگار است.

.Quitting smoking could prevent bladder cancer in smokers

.80آقای خاتمی توانسته چهره ایران را در سطح بینالمللی مطلوب نشان دهد.

Mr. Khatami has demonstrated a good portray of Iran on .international levels

.81آثار تاریخی اصفهان را میتوان از عجایب جهان به حساب آورد.

Ancient monuments of Isfahan could be counted one of world’s .wonders

.82فاکتورهای منفی موجود در شرایط زندگی زنان پرشمار است.

.Negative factors of women’s lives are numerous

.83به طور متوسط حقوق زنان نصف حقوق مردان است.

.On average women salaries are half of the men

.84مردم لبنان هرگز حمایتهای همه جانبه ایران را از خود فراموش نمیکنند.

.The Lebanese people will never forget Iran’s supports

.85قطعاً ما از بیثباتی در خاورمیانه نفعی نخواهیم برد.

.We definitely will not profit from instability in the Middle East

.86آدمیان در شهرها یا در چادرها کنار یکدیگر خانه دارند.

.Humans inhabit in cities or tents besides each other

.87هر زنی میخواهد که برای همسرش مهمترین شخص باشد.

Every woman wants to be the most important person for her husband.

88.مردان و زنان نسبت به چیزهای متفاوت حسادت میورزند.

Men and women get jealous because of different issues.

89.مرد نسبت به مسائل خانواده و زناشویی حسادت میورزد.

Man gets jealous for family and marital issues.

90.معنا و هدف اصلی هدیه برای زنان اهمیت دارد.

The main aim of giving gift is important for the women.

91.رضایت خاطر زنان در گرو ارضای نیازهای عاطفی آنان است.

Women’s satisfaction is in pawn of their emotional gladden.

92.هیچوقت به عنوان هدیه به همسرانتان وسایل آشپزخانه ندهید.

Never give kitchen ware as a gift to your wife.

93.نظر مرد در مورد وضع ظاهری همسرش خیلی اهمیت دارد.

Man’s opinion on his wife’s appearance is very important.

94.زندگی مردم آمریکا به وامهای سنگین متکی شده است.

American peoples’ lives are dependent on heavy loans.

95.جهت تماس با سایر شهرستانها گرفتن کد شهر الزامی است.

The city code is mandatory to call other cities.

96.رسیدگی به معشیت مردم از وظایف مهم همه مسئولان کشور است.

Organizing peoples’ livelihood is one of the crucial duties of the country officials.

97.مردم در گذران زندگی با سختیهای فراوانی روبرو هستند.

People meet very adversities in thriving their lives.

98.متاسفانه بخش عمدهای از نیازهای ما از صدور نفت تامین میشود.

Unfortunately major parts of our needs are supplied from exporting petrol.

99.اوج حماسه و معجزهگری جوانان ایرانی در دوران دفاع مقدس بود.

The miracle and the epic of Iranian youth was on its’ peak in the Holy Defense era.

**Yes/no questions used in the experiment**

در جمله ای که خواندید، در مورد ارتباط تماشای برنامه های مذهبی با حضور در کلیسا بحث شده بود. بله

در جمله ای که خواندید، از آغاز هفته دفاع مقدس بحث شده بود. نه

در جمله ای که خواندید، در مورد فصل بهار در ایالات آمریکا بحث شده بود. بله

در جمله ای که خواندید، از طراحی تلسکوپ جدیدی بحث شده بود. بله

در جمله ای که خواندید، از درمان علائم یائسگی بحث شده بود. بله

در جمله ای که خواندید، مصرف شیرینی ها بحث شده بود. بله

در جمله ای که خواندید، در مورد بیماری های واگیردار بحث شده بود. بله

در جمله ای که خواندید، در مورد خواص سبزیجات بحث شده بود. نه

در جمله ای که خواندید، در مورد عکس العمل پلیس در تظاهرات بحث شده بود. نه

در جمله ای که خواندید، از نگاه به صفحات کتاب بحث شده بود. نه

در جمله ای که خواندید، از روشهای درمان بیماری های قلبی بحث شده بود. نه

در جمله ای که خواندید، در مورد بیماری پارکینسون بحث شده بود. بله

در جمله ای که خواندید، از طوفان اخیر در آمریکا بحث شده بود. نه

در جمله ای که خواندید، از آمار تنبیه بدنی کودکان بحث شده بود. نه

در جمله ای که خواندید، گفته شده بود که دود سیگار باعث جهش ژنتیکی می شود. بله

در جمله ای که خواندید، از عامل اصلی بیماری های روانی بحث شده بود. نه

در جمله ای که خواندید، از تصورات مردم آمریکا از ایران امروز بحث شده بود. بله

در جمله ای که خواندید، از تاثیرات ورزش بحث شده بود. نه

در جمله ای که خواندید، از معنای اصلاحات بحث شده بود. بله

در جمله ای که خواندید، از گذاشتن غذاهای پخته در یخچال بحث شده بود. بله

در جمله ای که خواندید، در مورد اثرات مصرف مواد مخدر در جوانان بحث شده بود. نه

در جمله ای که خواندید، از اثرات فشار کار بر روی سلامتی بحث شد. بله

در جمله ای که خواندید، در مورد تدریس زبان چینی بحث شد. نه

در جمله ای که خواندید، تدریس زبان فارسی در دانشگاههای آمریکا بحث شده بود. بله

در جمله ای که خواندید، در مورد هواپیماربایی بحث شد. نه

در جمله ای که خواندید، از گزارش سازمان ملل در مورد جمعیت کشورها بحث شده بود. نه

در جمله ای که خواندید، از روابط تهران و واشینگتن بحث شده بود. نه

در جمله ای که خواندید، از اثرات خشونت بحث شد. نه

در جمله ای که خواندید، از اشتیاق جوانان به بازیهای کامپیوتری بحث شد. نه

در جمله ای که خواندید، از اثرات زایمان بر سلامتی مادر بحث شد. نه

در جمله ای که خواندید، در مورد آیین دادرسی کیفری بحث شد. بله

در جمله ای که خواندید، از تاثیرات دانشمندان غربی بر روی متون طب بحث شد. نه

در جمله ای که خواندید، از مبانی طب غربی بحث شد. بله

در جمله ای که خواندید، از پیشگیری بیماریهای لثه و دهان بحث شده بود. بله

در جمله ای که خواندید، نیاز مردم به یارانه های دولت بحث شد. بله

در جمله ای که خواندید، از تاثیر شکستن تابوهای اجتماعی بر پیشرفت علم پزشکی بحث شد. بله

در جمله ای که خواندید، رهبری به عنوان یکی از مهم ترین مسایل کشوری مطرح شده بود. نه

در جمله ای که خواندید، از قطع روابط کشورها با ایران بحث شده بود. نه

در جمله ای که خواندید، از گسترش انقلاب ارتباطات بحث شد. بله
در جمله ای که خواندید، از روابط ایران و کشورهای جهان بحث شد. بله
در جمله ای که خواندید، در مورد کشیشهای مسیحی بحث شد. نه
در جمله ای که خواندید، در مورد بیماری سرطان بحث شد. نه
در جمله ای که خواندید، ممنوعیت ورود نیروهای مسلح به دانشگاهها مطرح شد. بله
در جمله ای که خواندید، در مورد منع فرزندان از تحصیل در برخی مناطق بحث شده بود. بله
در جمله ای که خواندید، آمیزش ایرانی ها با نواحی آسیای صغیر بحث شد. بله
در جمله ای که خواندید، در مورد عقب ماندگی علمی ایران بحث شد. نه
در جمله ای که خواندید، قرون رنسانس اروپا بحث شد. نه
در جمله ای که خواندید، از الهه ی آب بحث شد. نه
در جمله ای که خواندید، از کتاب صدای زلال بلال بحث شد. بله
در جمله ای که خواندید، از طلوع و غروب خورشید بحث شد. بله
در جمله ای که خواندید، علت تصادف اتوبوس ذکر شد. بله
در جمله ای که خواندید، جدید و نو بودن انقلاب مطرح شد. بله
در جمله ای که خواندید، در مورد ارکان خانواده بحث شد. نه
در جمله ای که خواندید، مطامع استعماری بحث شد. بله
در جمله ای که خواندید، در مورد کاهش جمعیت بحث شده بود. نه
در جمله ای که خواندید، در مورد بانک جهانی بحث شد. بله
در جمله ای که خواندید، در مورد جستجوی آزادی بحث شد. بله
در جمله ای که خواندید، در مورد کشته شدگان هواپیما بحث شده بود. نه
در جمله ای که خواندید، در مورد کاراته کارها بحث شده بود. بله
در جمله ای که خواندید، از سرعت پیشرفت چین و کره بحث شده بود. بله
در جمله ای که خواندید، از جمعیت شرق بحث شد. بله
در جمله ای که خواندید، در مورد تامین سوخت کشتی ها بحث شد. بله
در جمله ای که خواندید، در مورد محیط زیست بحث شد. نه
در جمله ای که خواندید، در مورد گربه ها بحث شد. نه
در جمله ای که خواندید، در مورد چاقی بحث شد. بله
در جمله ای که خواندید، از مضرات سبزیجات بحث شد. نه
در جمله ای که خواندید، در مورد اندیشه های محدود عامه بحث شد. بله
در جمله ای که خواندید، در مورد تلویزیون بحث شد. بله
در جمله ای که خواندید، در مورد صادرات نفت بحث شد. نه

در جمله ای که خواندید، در مورد سیاست اروپا در قبال پدیده مهاجرت بحث شد. بله

در جمله ای که خواندید، در مورد جراحی قلب بحث شد. نه

در جمله ای که خواندید، در مورد قوه قضاییه بحث شد. بله

در جمله ای که خواندید، در مورد نظام قضایی بحث شد. بله

در جمله ای که خواندید، در مورد فواید آب بحث شد. نه

در جمله ای که خواندید، در مورد سبزیجات بحث شد. نه

در جمله ای که خواندید، در مورد رموز سخن گفتن بحث شد. بله

در جمله ای که خواندید، در مورد سیاست غرب بحث شد. نه

در جمله ای که خواندید، در مورد مسائل بشر بحث شد. بله

در جمله ای که خواندید، در مورد داروهای ضد سرطان بحث شد. نه

در جمله ای که خواندید، در مورد اقتصاد کشور بحث شد. نه

در جمله ای که خواندید، در مورد آثار تاریخی اصفهان بحث شد. بله

در جمله ای که خواندید، از خشونت خانگی بحث شد. نه

در جمله ای که خواندید، در مورد حقوق بشر در غرب بحث شد. نه

در جمله ای که خواندید، در مورد جنگ داخلی بحث شد. نه

در جمله ای که خواندید، از ثبات اقتصادی در کشور بحث شد. نه

در جمله ای که خواندید، در مورد زندگی انسانها در کنار هم بحث شد. بله

در جمله ای که خواندید، در مورد زنان بحث شد. بله

در جمله ای که خواندید، در مورد حسادت بین فرزندان بحث شد. نه

در جمله ای که خواندید، در مورد حسادت مردان بحث شد. بله

در جمله ای که خواندید، در مورد معنا و مفهوم هدیه برای زنان بحث شد. بله

در جمله ای که خواندید، در مورد نیازهای جنسی زنان بحث شد. نه

در جمله ای که خواندید، از اهمیت آرامش در خانواده بحث شد. نه

در جمله ای که خواندید، از وضع ظاهری زنان بحث شد. بله

در جمله ای که خواندید، در مورد زندگی مردم آمریکا بحث شد. بله

در جمله ای که خواندید، در مورد اهمیت تلفن در زندگی بشری بحث شد. نه

در جمله ای که خواندید، از وظایف مسئولان کشور بحث شد. بله

در جمله ای که خواندید، از آسایش زندگی مردم بحث شد. نه

در جمله ای که خواندید، در مورد صادرات خشکبار بحث شد. نه

در جمله ای که خواندید، در مورد راهپیمایی روز قدس بحث شد. نه

## Appendix E

### Bilgilendirilmiş Onay Katılım Formu
### (Informed Consent Form)

Bu çalışma, ODTÜ Enformatik Enstitüsü Bilişsel Bilimler Programı bünyesinde yürütülmekte olan yüksek lisans tez çalışma kapsamında düzenlenmektedir. Çalışma danışmanı Yrd. Doç. Dr. Cengiz Acartürk ve çalışmanın gerçekleştirildiği yüksek lisans öğrencisi Fatemeh Soleymani'dir.

Çalışma, Farsça okuma alışkanlıkların incelenmesi amacı ile gerçekleştirilmektedir. Çalışma boyunca gösterilecek materyal genel olarak kişisel rahatsızlık verecek içeriğe sahip değildir. Sizden beklenen, cihaz ekranında gösterilen yönergeleri takip

etmektir. Ancak, katılım sırasında gösterilen materyalden ya da herhangi başka bir nedenden ötürü kendinizi rahatsız hissederseniz çalışmayı yarıda bırakıp çıkmakta serbestsiniz. Böyle bir durumda çalışmayı yürüten kişiye, çalışmayı tamamlamadığınızı söylemek yeterli olacaktır.

Çalışmaya katılım bilgilendirilmiş onay (informed consent) esasına dayanmaktadır. Çalışma boyunca, sizden istenecek kimlik bilgileri verilerle eşleştirilmemektedir. Cevaplarınız tamamıyla gizli tutulacak ve sadece araştırmacılar tarafından değerlendirilecektir; elde edilecek bilgiler bilimsel yayınlarda kullanılacaktır. Çalışma sonunda, varsa çalışmayla ilgili sorularınız cevaplanacaktır.
Katılımınız için şimdiden teşekkür ederiz.

(Formu doldurup imzaladıktan sonra uygulayıcıya geri veriniz).

***Bu çalışmaya bilgilendirilmiş olarak katılıyorum ve istediğim zaman yarıda kesip çıkabileceğimi biliyorum. Verdiğim bilgilerin bilimsel amaçlı yayınlarda kullanılmasını kabul ediyorum.***

İsim Soyad Tarih İmza

----/----/-----

## Appendix F

## Dilbilgisel Artalan Anketi

| **Kişisel Bilgiler**<br>(Bu formdaki kimlik bilgileri verilerle eşleştirilmemektedir.) | | **Kod:** |
|---|---|---|
| Soyadı | Adı | Bugünün Tarihi |

| | | | |
|---|---|---|---|
| Doğum Yılı | Kadın ( ) | Erkek ( ) | |
| Telefon Numarası | E-posta Adresi | | |
| Şu anki mesleğiniz? | | | |
| En yüksek tahsiliniz (veya muadili)<br>(lütfen yuvarlağa alınız) | Ortaokul | Lise | Üniversite Derecesi |
| | Mesleki Eğitim | Diğer? | |
| Fakülteniz | | | |
| Bölümünüz | | | |
| Sınıfınız | Hazırlık ( ) 1. Sınıf ( ) 2. Sınıf ( )<br>3. Sınıf ( ) 4. Sınıf ( ) | | |
| Lisede hazırlık okunuz mu? | Evet ( )<br>Hayır ( ) | | |
| Üniversitede hazırlık okudunuz mu? | Evet ( )<br>Hayır ( ) | | |

| **Genel Sağlık Durumunuz** | | | |
|---|---|---|---|
| Yazarken hangi elinizi kullanıyorsunuz? | Sağ ( ) | Sol ( ) | |
| Tanısı konmuş herhangi bir dil bozukluğunuz var mı (disleksi, kekemelik gibi)? | Hayır ( ) | Evet ( ) | Varsa, lütfen ayrıntılandırınız. |
| Çalışma sırasında gözlük kullandınız mı? | Hayır ( ) | Evet ( ) | |

| | | | |
|---|---|---|---|
| Çalışma sırasında lens kullandınız mı? | Hayır ( ) | Evet ( ) | |

| **Hangi dil(ler)i, hangi sırayla öğrendiniz?** (anadiliniz dahil) | | | |
|---|---|---|---|
| Dil | Hangi yaştan itibaren? | Ne kadar süreyle? | Öğrendiğiniz yer? (evde, okulda, başka) Lütfen belirtiniz. |
| 1. | | | |
| 2. | | | |
| 3. | | | |

| İran ve Türkiye dışında başka ülkelerde yaşadınız mı? | Ne kadar süreyle? | Hangi sebeple? (okul, eğitim, vs.) |
|---|---|---|
| 1. | | |
| 2. | | |
| 3. | | |