2017-04-01

Eye Behavior While Reading Words of Sanskrit and Urdu Origin in Hindi

Tahira Carroll
Brigham Young University

Eye Behavior While Reading Words of Sanskrit and Urdu Origin in Hindi

Tahira Carroll

A thesis submitted to the faculty of
Brigham Young University
in partial fulfillment of the requirements for the degree of

Master of Arts

Deryle Lonsdale, Chair
William Eggington
Dan P. Dewey

Department of Linguistics and English Language

Brigham Young University

ABSTRACT

Eye Behavior While Reading Words of Sanskrit and Urdu Origin in Hindi

Tahira Carroll
Department of Linguistics and English Language, BYU
Master of Arts

Hindi and Urdu are two branches of the same language sometimes known as Hindustani. They are divided by orthography and geography but when spoken are sometimes indistinguishable. Both have contributed loanwords that have now been completely assimilated into the language. The question of how the eye behaves during Hindi reading when it encounters Urdu loanwords has not been focused on extensively in prior research. The main purpose of this thesis is to document the eye behavior during reading Sanskrit-based words and Urdu loanwords in Hindi. We place fifteen word pairs consisting of one target Hindi Sanskrit-based word and its Urdu loanword equivalent in different sentences. Native Hindi speakers participate to read Hindi sentences containing either Urdu loanwords or the Sanskrit root word in Hindi. To quantify the differences in reading Hindi and Urdu loanwords in Devanagari (Hindi script) sentences we use an eye tracking methodology, which is used to measure eye movements of a participant during reading. We discover very distinctive eye behavior during reading of Urdu loanwords in comparison to reading Hindi Sanskrit-based words. Analysis also shows an interaction in eye behavior due to language and frequency.

Keywords: Eye tracking, reading, loanwords

ACKNOWLEDGMENTS

I would first like to thank my thesis advisor Dr. Deryle Lonsdale of the Linguistics and English Language Department at Brigham Young University. Dr. Lonsdale was there helping me every step of the way and taught me to navigate the whole process of thesis writing. He spent countless hours reviewing, encouraging and keeping me focused on the right track. I have been honored to have the opportunity to learn from a great mind and a true educator.

I must also express my very profound gratitude to my husband, Fernando Carroll for providing me with unfailing support and continuous encouragement throughout my years of study and through the process of researching and writing the thesis. My children, Adriel and Enaya and my mom, San San who were my own personal cheering squad supporting me all the way. This accomplishment would not have been possible without them.

Thank You

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

INTRODUCTION

This thesis addresses the difference in eye behavior of Native Hindi speakers when reading Hindi Sanskrit-based words and Hindi Urdu loanwords written in Devanagari (Hindi script). There has been debate for decades on how separate Hindi and Urdu truly are and whether the difference is only in orthography. For example, the word for ‘if’ in Hindi/Sanskrit is ‘yadi’ and the Urdu equivalent is ‘agar.’ Both words are found in spoken and written Hindi. Prior research and literature indicates that the eye behavior will differ when the eye encounters a word that is more familiar than one that is unfamiliar (K. Rayner, Ashby, J., Pollatsek, A., & Reichle, E. D., 2004; K. Rayner, Slattery, Drieghe, & Liversege, 2011; Husain, Vasishth, & Srinivasan, 2015). There are two possibilities of how an eye will behave when encountering a loanword during reading. The first possibility is that the eye movement will reflect the reader’s surprise, treating the loanword as a less familiar word. The other possibility is that there will be no measurable difference in eye behavior when encountering a loanword.

Historically Hindi and Urdu have been intertwined so closely that some people cannot even distinguish between the two languages. When speaking in Hindi many in India are not aware when they are using Urdu loanwords and may be more familiar with the loanwords than they are with the Sanskrit-based words. Urdu is completely mixed in with Hindi and even forms hybrids of Hindi/Urdu which are now found in everyday speech (Abbi & Hasnain, 1986).

In sociolinguistics a situation in which languages that are almost the same when spoken and have completely different writing systems is called digraphia. This unique digraphic interaction allows for a study that prior research has not focused on and starts to answer the question of how digraphic languages interact with loanwords. Our main hypothesis is that there will be distinctive eye behavior when Urdu loanwords are read in Hindi sentences. The null hypothesis is that there will be no difference in eye behavior.

The main two questions that are being investigated in this beginning step to understand the eye behavior when reading Hindi/Urdu loanwords are:

1. Is there a difference in eye behavior or movement when reading Hindi/Sanskrit root words and Urdu loanwords written in Devanagari?
2. What are the other interactions that are affecting the difference in eye behavior during reading?

To accomplish answering these questions we have chosen eye tracking as our methodology. Many researchers prefer eye tracking technology to observe the behavior of the eye during reading. Eye tracking is one of the accurate ways to measure reading processes and has a definite advantage over other methods in its ability to analyze the fractional movements of the eye during reading. An advantage of eye tracking is that it targets observation of the specific word or phrase that a researcher is interested in studying. It provides many types of measurements of eye behavior information to the researcher that then can be used to quantify the many different movements during reading (Roberts & Siyanova-Chanturia, 2013).

We decided to answer these questions with an eye tracking experiment to better understand the eye behavior in this digraphiac situation. In order to activate the reasons for this choice, we first discuss relevant research in reading and eye tracking, followed by a description of our methodology. We then report our analysis of the measurements from the experiment and conclude with a discussion of our results.

LITERATURE REVIEW

Studying eye behavior allows one to determine how the eye reacts when encountering different words, phrases, pictures or even whole sections of a passage. Eye behavior during reading registers surprise when encountering a specific word, phrase or region. In this thesis we will first discuss eye behavior during reading. We will solely focus on the eye behavior and how that can explain some of the reading processes. Any inferences about cognition not relating to eye behavior will not be discussed as this falls outside the scope of this experiment. Once we establish eye behavior and its relation to reading we will discuss how the eye behavior will be measured. Eye tracking allows one to measure the different eye behaviors. We will be defining and explaining the eye measurements that are used in this experiment. We will also discuss briefly how Hindi and Urdu are linked in history and continue to do so in present day India.

Interpreting the eye movement control in reading allows better understanding of the cognitive behavior that occurs in reading comprehension (Pollatsek, Reichle, & Rayner, 2006). Research has now proven that the eye does not just read blocks of words or chunks. Readers actually read every letter and word through the whole sentence whether they are aware of it or not (Rooy & Pretorius, 2013). A link between eye movement and cognition has been established and in simpler terms knowing where the eye is, and how it focuses and moves, reflects the comprehension that is occurring (Rayner, 1998).

Many studies have researched the reading processes and how they involve cognition. During reading the eye makes a series of pauses and movements from one point to another (Roberts & Siyanova-Chanturia, 2013). Rayner (1998) in discussing eye tracking and reading, stated that the eyes pause between movements for enough time to recognize a word, or in other terms, for comprehension to occur. Words that are very commonly used are not paused on and the eye travels over them without stopping. This does not mean that the eye is not reading the word; it just means that the eye does not need to stop on it directly to process it. The eye pauses or stops on content words 85% of the time when they are encountered. Function words receive less attention because they are recognized with more ease (Carpenter, Just, & Rayner, 1983).

Rayner (1998) also points out that reading, though progressive, does not always move forward and that around 10-15% of all movements that occur are regressive (Roberts & Siyanova-Chanturia, 2013). Regressive or backward eye movements to previously encountered text or words can be either short or long; the longer the movement the more processing difficulty (Rayner, 1998). Words that are hard to comprehend, that are not familiar or ambiguous can cause longer and more regressive eye movements (Roberts & Siyanova-Chanturia, 2013). The harder the overall text is to comprehend the larger the number of regressions (Pollatsek et al., 2006). Thus regression has been proven by many different studies as a good measurement of comprehension of words and sentences.

Research has shown that ambiguous sentences cause more regressive eye movement as the eye tries to maneuver or move to a previous word to try and understand the ambiguous portion of the sentence or word (Mitchell, Shen, Green, & Hodgson, 2008). This associates trouble in syntactic processes with more regressive eye movement, which shows that there is perhaps a comprehension issue.

The movement of the eye from one point to another as a person tries to comprehend or read a word is called a saccade; the time when the eye fixates or stops is referred to as fixation. Eye tracking allows the observation of fixations and saccades. Fixations account for 85% of the reading time while saccades occur for about 15% (Bonhage, Mueller, Friederici, & Fiebach, 2015). Fast readers, in comparison to slow readers, tend to make longer saccades, shorter fixations, and fewer regressions (Anson & Schwegler, 2012). Thus fast readers tend to pause for a shorter time on words allowing them to have longer saccadic movements of the eye as it travels through the text. When a person is attempting to comprehend difficult text the fixation time is longer and the eye does short saccades or regression movements near the word. The average fixation time in reading is 200-250 milliseconds (Mishra & Singh, 2014).

Eye Tracking Measurements

Eye tracking is one of the methods to observe eye behavior. It provides the ability to measure different eye behaviors and research has correlated eye tracking measurements to explain some of what is occurring during reading.

Terms and concepts that will be important for this thesis are: Area of interest, fixation, saccades, first fixation duration, first run time, dwell time, regression in, regression out, skip and fixation count. The definitions and explanations are briefly defined below; (See Holmqvist et al. (2011) for a more detailed discussion).

Area of Interest (AOI) refers to the paragraph, sentence, word, picture or area that is of interest to the researcher. Specifying an AOI box is central to eye tracking and crucial to our experiment since we were looking to determine eye behavior of specific words. AOI boxes can be formed around a word to obtain exact data for what occurs when reading that specific word. *Fixation* is the time when the eye movement stops and is fixed on an AOI. *Saccades* are the movements the eye produces when progressing from one fixation to another. Once an AOI is established it allows a researcher to count the fixations and how long the fixations are on the specific area that the researcher wants to observe.

First fixation duration (FFD) begins when the eye first encounters a stimulus. The first fixation does not always materialize at the beginning of a word; sometimes the first fixation occurs even earlier, right after the previous saccade ends. The FFD reflects the duration of time taken for the word to be identified and recognition of the word to occur. The duration of the fixation is dependent on the processing of the information (Holmqvist et al., 2011). When this measurement was first developed FFD was linked with lexical activation processes (Holmqvist et al., 2011). First fixation is usually counted after the initial saccade and this becomes the first fixation point of the AOI. The properties of words that are directly correlated with FFD measurements are word frequency, morphological complexity, metaphorical status, orthographic properties, the degree of polysemy and other linguistic factors (Inhoff & Radach, 1998). It has become common practice to use FFD measurements in reading research.

According to prior research, if the initial fixation falls on the initial few letters then the fixation is short and the next FFD will be longer (Inhoff & Radach, 1998). If the initial fixation falls in the middle of the word, the fixation is longer and the subsequent FFD will be shorter. Thus FFD measurements can help predict other measurements in eye tracking. Prior research has proven that FFD has a direct correlation with the comprehension of a word.

FFD has been used in various visual studies. Henderson et al. (1999) used FFD to measure visual information acquisition. In their research, they found that FFD were longer for more informative areas of the pictures because more needed to be comprehended or it took longer to comprehend. They also found shorter FFD when there were simpler areas of the picture.

First Run Time (FRT) is the duration of the first dwell when the eye pauses or fixates on a word or the duration of the first fixation in an AOI such as a word, sentence or region. Reading research communities call FRT by many names. Some are: first pass dwell time, first pass gaze duration, first pass fixation time and duration of the first fixation (Holmqvist et al., 2011). If there is only one fixation then the FRT is the sum of the FFD and the saccades (Holmqvist et al., 2011).

Many researchers propose FRT as the measure for early processing and object recognition (Holmqvist et al., 2011). Researchers such as Henderson et al. (1999) have even further hypothesized that FRT and even second pass fixation times are longer for more complex sentences. Although Henderson et al. (1999) do state that total time spent on a stimulus is a better indicator for studying object recognition. A study conducted by Smith and Levy (2010) focused primarily on how first pass reading results of FRT measurements can determine if there is uncertainty in understanding a word. They used a Bayesian Model of word recognition to determine if word recognition occurs even before the complete word is processed and whether neighbor words aid in the comprehension through context. One of their main results was that higher frequency words tended to have shorter FRT than less frequent words. They also established a direct correlation between FRT and initial word recognition.

Dwell Time (DT) refers to the time from when the eye first encounters and then leaves the AOI. A higher DT can indicate uncertainty and problems in comprehension. DT is one of the most important measurements for any reading research. DT does not include all the repeat visits or regressions. DT is usually defined as all the fixation durations that occur in an AOI (Holmqvist et al., 2011). This time measurement includes fixations as well as non-fixations such as saccades and blinks (Holmqvist, 2015). Usually measurements over 3000 milliseconds are not used with ANOVA calculations because the variance disappears and those results are considered outliers, anomalies or mistakes and are not included in the results.

Many researchers use DT as their main measurement focus in reading research because it can show so much of the cognition process. The longer DTs indicate more interest in the AOI or more complex cognitive processes (Holmqvist et al., 2011). In a recent study by Rodd, Gaskell and Marslen-Wilson (2004) on modelling the effects of semantic ambiguity in word recognition they found that ambiguous words have slower lexical decision times in comparison to the unambiguous words.

A correlation between dwell times and comprehension have been established by various studies. Rayner (1998), in researching dwell times, found that longer dwell times relate to less frequent words and that the comprehension takes longer for those words than for frequently used words. In a subsequent study Pollatsek et al. (2006) further hypothesized that DT is used to study longer cognitive functions while FFD can be used for cognitive functions that are very fast, such as lexical activation and recognition. Henderson et al. (1999) found that there is a clear relationship between meaning and understanding and DT. In their study, items in the picture that were more interesting and complex required more processing time to understand and thus had a longer DT.

Regression in (RI) occurs when the eye backtracks in the opposite direction from the word or AOI. The saccade travels backwards. Sometimes this is also known as in-word regression in which the eye maneuvers to a previous section of the word within the same word. Regression occurs when the eye is not fixating but rather involves saccadic movement. Prior research has shown that in visual searches, saccades tend to occur towards the areas that contain the most amount of task relevant information (Pomplun, Reingold, & Shen, 2003). Thus many saccadic movements of regression could indicate that the participant is having difficulty with processing the word.

Regression out (RO) transpires when there is saccadic movement leaving the AOI or word to help process a problematic word. Usually this occurs when the eye travels back to a previous word to help with context. Researchers in previous studies have found that saccadic eye movement can occur cognitively when the eye is looking for hints or markers before and after to process the AOI. An increasing number of researchers are fascinated with the relationship between what is seen and the linguistic processes that occur to comprehend the picture or reading (Richardson & Daleb, 2005). Many studies focus on how words or contextual drawings can affect the eye behavior when encountering an AOI. In a study conducted by Smith and Levy (2010), they asked the question, "Can linguistic processing occur even before a word is completely identified?" They found that there is a relationship between the AOI and the words or visuals that give it context.

A scanpath is a visualization of the eye movements during reading. Figure 1 is an example of a scanpath of an English sentence. This example illustrates the behavior of the eye as it reads the sentence. For the AOI in Figure 1, 'windmill', there are three fixation points as well as multiple regressive movements out of the word and even a regression to a previous word. The regression to a previous word occurs because the eye registers surprise upon encountering the unexpected word 'windmill'. Regression in scanpaths like this example illustrate that regression occurs when the eye is surprised or when there is a comprehension issue. The eye compensates by either moving back within the word or jumping back to a previous word to help with context and comprehension. More regression equals more cognitive processes involved in comprehending the word (Holmqvist et al., 2011).

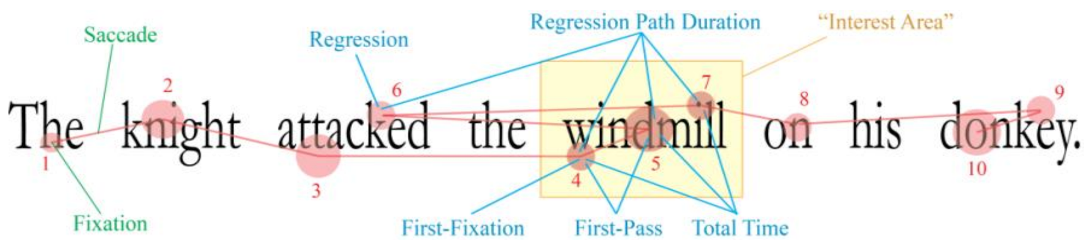


Figure 1: Scanpath of an English Sentence[1]

Regressions in and between words can give us a visual of the cognitive processes that are occurring. Many regression movements generally indicate a problem in comprehension. For example, Rayner (1998) shows that readers with dyslexia have more regressions. In elementary school, as reading skills improve and there is more comprehension of words as the children's vocabulary grows the number of regressions decreases (Holmqvist, 2015). Pollatsek et al. (2006) also found that older people once again start to increase the number of regressions, as their memories worsen and their cognitive processes take longer.

Skip occurs when a reader does not stop or fixate on a word and continues on to the next word even if they come back to it or regress back after skipping. Skip in eye tracking measurements is different from skipping that occurs in other linguistic processes. In eye tracking skip refers to a word with no fixations. The word is read but the eye just moves over it without pausing or fixating. Whether the eye skips the AOI is another observation to determine the difference in eye behavior in reading Hindi. As mentioned above a greater percentage of function words are usually skipped (Holmqvist et al., 2011). In Figure 1, the word 'the' is skipped and is only passed by with eye movement when there is regression from the AOI word 'windmill.' Rayner (1998) compared sentences with high frequency words (more than 10,000 occurrences per million-opm) and found that those words are skipped 67% of the time. Low frequency words

[1] From http://proswrite.com/2013/06/29/

(1 -10 opm) are skipped 10% of the time. Word length also plays a role in whether it will be skipped or not. The shorter a word the more likely it is to be skipped (Inhoff & Radach, 1998). Rayner (1998), in his study, also found that contextually predictable words are more likely to be not fixated upon or skipped. Levy et al. (2010) conducted a study in which they predicted that there is a neighborhood effect on words. Higher-frequency neighbor words mislead through interference and can inhibit the correct identification of the word.

Fixation Count (FC) is the total number of fixations that occur within an AOI (Henderson et al., 1999). It is the number of fixations from the beginning to the end of the AOI. FC is a critical measurement that is used for eye tracking studies dealing with reading. According to Henderson et al. (1999), the fixations that occur inside an AOI are called the *fixation density.* This does not measure the fixation duration like that of DT but the number of fixations that occurred in the AOI. Henderson et al. (1999) found that there were more fixations on semantically complex regions. In earlier studies of reading, fixation count was the most common measure where 11 out of 24 studies used it to determine whether there is a problem in comprehension (Holmqvist et al., 2011).

Studies have shown that a new task or a word that is unfamiliar receives more fixations. The FC is higher for tasks that take longer to process than ones that are more familiar. There is a correlation between FC and the difficulty of a task. Morris and Miller (1996) studied the behavior of pilots in simulation planes and found that fixation decreases as pilots spent more time on task. This study did not directly deal with just reading but it proved that FC is a good measure to determine the complexity or time taken to process a task. Thus as familiarity and comprehension grew for these pilots the fixation times decreased. Nakayama, Takahashi and Shimizu (2002) studied how task difficulty related to eye movement and found that fixation and task difficulty are negatively correlated. Van Orden, Limbert, Makeig and Jung (2001) conducted a study in which participants completed a task of a 2-hour mock warfare. They found that FC can predict the workload or track the performance of the participant. The higher the workload the more the FC, which could then be used as a measure of mental workload (Holmqvist et al., 2011).

Eye tracking has ecological validity and research has shown that it can be applied to real-world situations (Spinner, Gass, & Behney, 2013). Today this new technology allows us to better visualize and in consequence is changing our understanding of what is ensuing when we look at something or read a text (Anson & Schwegler, 2012).

One of the greatest advantages to using eye tracking methodology is its ability to show and record what is happening before, during and after reading. Eye tracking can also give late-time based measures along with fixation counts and regression (Roberts & Siyanova-Chanturia, 2013). Eye tracking gives us exact real time insight into what is happening in the eye and thus it becomes the ideal methodology to evaluate the fine-grained differences in eye behavior when reading Hindi and Urdu loanwords. For this experiment as mentioned before we needed to look at a specific AOI or the specific Urdu loanword and Sanskrit-based word. Eye tracking allows us to define those AOI's that we want to observe.

Hindi/Urdu

Hindi is the national language of India and is an Indo-Aryan language. It has its ancestry in the ancient Sanskrit family but it does have influences from Dravidian languages, Turkish languages, Arabic, Persian, Portuguese and English. Hindi is the fourth most spoken language after English, Spanish and Mandarin. It is written in Devanagari and is a SOV (subject-object-verb) language.

Historically, Urdu was the language of the royal courts of the Mogul Empire in India. Some have even claimed that Urdu was developed for the Muslim Mogul kings to help with ruling a nation that did not speak their Arabic language. Thus Urdu became a transliteration of a mixture of Hindi, Arabic and Persian. Urdu is written in a script similar to Arabic called the Nasta'liq alphabet. The grammar of Urdu is almost identical to Hindi.

Later in 1947 the division into Hindi versus Urdu became one of the catalysts for the division of India and Pakistan. Hindi became the national language of India and Urdu became the national language of Pakistan. The Partition leaves a memory of tension that still echoes in present day India.

The question about Hindi and Urdu being different languages has long been debated. Yet there are few studies that try to attempt to quantify these differences. Many studies involve bilinguals with languages that do not overlap orthographically. The only possibly similar studies are research into Serbo-Croatian speakers but there is a major difference because Serbo-Croatian speakers are considered monolingual and have an integrated lexicon even though their orthography consists of Latin and Cyrillic alphabets (Marian & Kaushanskaya, 2004).

Recent studies into bilingualism and reading have challenged the age-old idea that bilinguals can activate and deactivate the languages that they know. Instead, there is substantial evidence that both are activated simultaneously (Marian & Kaushanskaya, 2004). In reading where there is a shared or similar script, words that are similar tend to have shorter fixation and dwell time. Orthographic neighbors or similar orthographies in reading can trigger and facilitate understanding in target languages (Kaushanskaya & Marian, 2007). As mentioned these observations in bilingual reading are only relevant when the orthographies are similar.

Hindi and Urdu in many ways are almost indistinguishable when spoken and are only different when written. Hindi is written in the Devanagari script from left to right and Urdu is written in Arabic script that was derived from Persian and from right to left (King, 2006). Hindi and Urdu are the textbook example of a digraphic situation.

A study conducted by Husain et al. (2015) observed eye behavior relating to comprehension and parsing difficulties when deviating from the canonical SOV (subject-object-verb) major constituent order of Hindi. They studied the first fixation duration (FFD), first pass reading time (FPRT), rereading times (RRT) and looked at regression eye movement. They found that a long FFD and FPRT and more RRT and regression eye movement indicated a possible comprehension difficulty in Hindi. This was one of the first studies that observed the difficulties in reading Hindi. Their main results were that there were longer fixations for longer syllabic words and that high frequency words had shorter first fixation durations and less regression. They made a correlation between saccade length and fixation times and the cost of integration and storage cost and the effect it has on reading difficulty.

In a recent collection of studies into the orthographies of South and South East Asia, some important observations were recorded. Two main observations for the Hindi writing system were (Nag & Perfetti, 2014):

1. Higher frequency words had shorter fixation times and dwell times.
2. Words with more syllables had longer fixation, dwell times and regression.

These observations are the basis of our methodology. We predict that the eye movements will be different whether reading Hindi versus the Urdu word and that there will be an interaction due to the familiarity of the reader with the word. If the word is a high frequency word, there may be more familiarity of the word and thus the eye behavior would be different than when encountering a less frequent or less familiar word.

METHODOLOGY

For the experiment 30 native speakers of Hindi participated in the eye tracking study in which each read 30 Hindi Devanagari script sentences. The study was conducted at Brigham Young University in Provo, Utah, USA. All the participants in this study were native speakers of Hindi and could read the Devanagari script; they were also bilingual English/Hindi speakers. All the participants were also exposed to Hindustani (a mix of Hindi and Urdu) as was determined by a background survey which showed where each participant was from in India. There are geographical regions in India that predominately speak Hindustani and all the participants in this study were from areas that are familiar with Hindustani.

Each participant, after filling consent forms, was asked to fill out a questionnaire that dealt with basic background information. The questions were designed to eliminate any variables that could potentially affect the results like age, gender and educational background. For a list of all the questions found in the questionnaire see Appendix 1.

We formed 30 sentences with the use of the EMILLE (Enabling Minority Language Engineering) Corpus which contains sentence pairs in Hindi and Urdu (Baker et al., 2004). The EMILLE Corpus was commissioned by the British Government. This was a collaborative effort of Lancaster University in England and Central Institute of Indian Languages in India. There are three components to this corpus monolingual, parallel and annotated corpus. The parallel corpus consists of 200,000 words of text in English and its accompanying translations in Hindi, Bengali, Punjabi, Gujarati and Urdu. The parallel corpus allowed us to find Sanskrit-based words and their Urdu equivalents. Some of the sentences from the EMILLE contain Hindi Sanskrit-based words that we replaced with Urdu loanwords. Due to font encoding issues with the eye tracker we had to use words with no conjunct characters or Sanskrit characters. This forced us to modify some sentences from the EMILLE. We were limited by the Hindi characters that we could use but we were able to circumvent the issue and manipulate 30 sentences with a range of varying differences, some words that were studied being more frequently used and some less frequent. We chose sentences that could be found in everyday speech in India as well as sentences that were similar to ones found in online newspaper articles. The sentences were of varying grammatical difficulty but none were ambiguous to cause a false eye surprisal due to ambiguity. For this same reason of eye surprisal no sentences with unusual contexts were chosen. All the Urdu loanwords that we chose were ones found in vernacular Hindi but are still considered Urdu loanwords. We transliterated the Urdu words into Devanagari script, in order to not elicit eye surprisal due to orthography.

A sampling of the sentences is provided in Figure 2. The words that were studied are bolded for convenience; in the experiment they did not appear in bold. The AOI words which were either the Urdu loanword or the Sanskrit-based words were placed in different places in the sentence: some at the beginning, some in the middle, and some at the end so that we could observe the participant encountering the specific word or stimulus. The full list can be found in Appendix 2.

अगर आप के पास कोई मुशकिल है हम को बताइए।
Agar ap ke paas koi mushkil hai ham ko bataie.
If you have any problems please tell us.

हम सब साथ **भोजन** बहुत देर से बनाते हैं।
Ham sab sath **bhojan** bahut der se banate hain.
All of us together prepare the **food** very late.

बहुत लोगों के पास बुखार के लिए अलग अलग **इलाज** हैं।
Bahut logon ke paas bukhar ke liye alag alag **ilaaj** hain.
Lots of people for fever have different **treatments**.

Figure 2: Sample stimuli sentences for the experiment

Table 1 lists the specific AOI words, their linguistic roots and their applicable frequency data. The frequency data was collected from the Prasant Hindi Corpus and was listed in Wiktionary[2]. The Prasant Hindi Corpus is made up of 100,000 words collected from various online newspaper articles, containing both Hindi and Urdu loanwords and giving the frequency of words that are frequently used. The Prasant Hindi Corpus did not list the frequency of non-frequent words.

Table 1: AOI Words with their linguistic root and relevant frequency data

Meaning	**Word Sanskrit/Urdu**	**Frequency**
Year	संवत sanvant	
	साल saal	621
Livelihood	जीविका jivika	
	रोजगारी rojgari	
Treatment	उपचार upcaar	
	इलाज ilaaj	
Law	विधि vidhi	375
	कानून kanoon	179
Schools	पाठशालाएं pathshalayen	
	मदरसा madrasa	
Help	सेविका sevika	401
	मदद madad	
Traveler	राही rahi	

[2] See http://www.wiktionary.org

	मुसाफिर musafir	
Food	भोजन bhojan	
	खाना khana	277
If	यदि yadi	1335
	अगर agar	665
Language	भाषा bhasha	1162
	ज़बान zabaan	
News	समाचार samachar	135
	खबर khabar	130
Regret	खेद khed	
	अफ़सोस afsos	
Reason	कारण karan	
	वजह vajah	141
Marriage	विवाह vivah	422
	शादी shadi	243
Important	विशेष vishesh	982
	खास khas	170

We made an effort to find word pairs of varying familiarity or frequency. Some word pairs had similar word frequencies like ‘khabar/samachar’ (news) to detect whether there is a difference in eye behavior for a Sanskrit-based word and Urdu loanword when both are frequent words. Words like ‘yadi/agar’ (if) are both commonly used but had different frequencies, and we wanted to see whether there is a variation in eye movement. We also observed word pairs like ‘khana/bhojan’ (food) in which ‘khana’ the Urdu loanword is commonly used, whereas ‘bhojan’ is not a common word. We tracked the eye behavior of word pairs like ‘vidhi/kanoon’ (law) in which the Sanskrit-based word is more common than the Urdu loanword. Finally we also looked at the eye behavior for word pairs in which both were not commonly used words as in ‘jivika/rojgari’ (livelihood).

In the experiment, we used the Eyelink 1000 Plus, which has the ability to record and process the eye movement at specific moments. AOI boxes were formed around the specific Hindi words and their Urdu loanword counterparts to observe the eye behavior during the reading of the target words for the sentences.

The EyeLink 1000 Plus is commonly used by reading researchers, especially for word-level measurements, because of the amount of data it can collect and process. The EyeLink 1000 requires correct calibration for each individual participant. According to Winke (2013) even though the Eyelink 1000 accomadates participants wearing glasses, there is interference which causes skewed or incorrect data. Glasses also make it very hard to set up the correct calibration right from the start. Thus the data from two of the participants had to be dropped due to glare from glasses, which caused skewed results.

After the questionnaire was completed participants were asked to seat themselves in front of the EyeLink 1000. They were asked to get comfortable and to place their head in the tower mount with their forehead resting against the frame and their chin supported by the chin rest for better stability. Once they were in a position that they could hold comfortably they were asked to push the space bar on a keyboard placed in front of them. The camera was adjusted and focused to be able to see the eye clearly.

In the global view of the screen, the image of the pupil was clicked on to determine the auto threshold and the camera was refocused if necessary. At the start, the EyeLink 1000 was calibrated with each participant. The calibration involved having the participants look at nine fixation points on a grid. Once they were able to fixate on the grid points they were asked to look at different fixation points on the screen to establish where their eyes were fixating. Calibrations were accepted and validated once the participant was able to fixate accurately on the fixation points. See Figure 3 for a visual of the calibration.

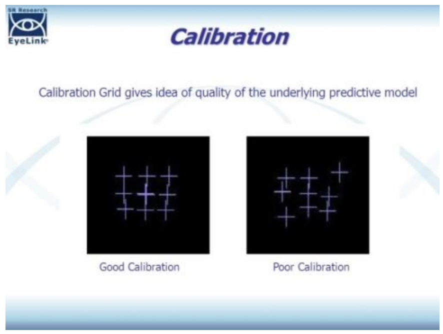


Figure 3: Calibration grid[3]

At this point, calibration and validation of the calibration were complete and the experiment could begin. Participants were asked to read 30 Hindi Devanagari sentences which were presented in random order for each. When a participant was ready to move on from one sentence to another, they pressed the space bar in front of them. The participants read 30 sentences and their eye movements were recorded.

The eye movements measurements we observed were: first fixation duration, first run time (the time it took for the eye to first pass over the word), the dwell time (how long the eye stayed on the word), regression in (did the eye jump back to the previous word), regression out (did the eye jump to the next word), skip (the eye jumping or skipping over the word) and the fixation count (how long the eye fixated or stopped on the word).

---

[3] See http://www.fishmanmarketing.com/projects/thompson-mcmullan/#sthash.YsBM0IMJ.dpbsWeb.

RESULTS

We employed both a one-way ANOVA and a factorial ANOVA to process the data for each word pair. We first conducted a one-way ANOVA to compare mean fixation times and regression movements between Hindi Sanskrit versus Urdu loanwords to determine whether they were statistically significant. The independent variables were language with two levels: Hindi and Urdu. The dependent variables are: FFD, FRT, DT, RI, RO, Skip and FC. The mean values of each measurement are compared for both languages and a $p < 0.05$ determined whether there was a significant difference in eye movement for that measurement.

One-Way ANOVA

For some of the word pairs and some of the measurements we detected a difference in eye movement. Table 2 below shows the p values for the word pairs for the seven different measurements from the eye tracking data. The significant values are bolded and an asterix appears next to every p value that is significant. Of the 105 different data points, 35 data measurements are significantly different between Hindi and Urdu.

Table 2: One-way ANOVA Data

Hindi/Urdu Loanword	FFD	FRT	DT	RI	RO	Skip	FC
vajah/karaan	0.398	0.505	0.369	0.225	0.647	**0.020***	0.556
khed/ahfsos	0.375	0.142	**0.036***	0.498	0.926	0.88	0.231
samachar/khabar	0.066	**0.042***	0.295	0.077	0.579	**0.001***	**0.003***
bhasha/zabaan	0.375	0.142	**0.036***	0.498	0.926	0.88	0.231
sanvant/saal	**0.018***	0.015*	**0.000***	**0.019***	0.936	0.474	0.147
bhojan/khana	0.28	0.812	0.702	0.865	0.59	0.983	0.529
raahi/musafir	**0.027 ***	0.158	**0.003***	**0.006***	**0.001***	0.053	**0.000***
sevika/madad	0.09	0.459	**0.000***	**0.000***	0.055	0.827	**0.001***
pathshalayen/madrasa	0.551	**0.028***	**0.002***	**0.009***	0.313	**0.000***	**0.003***
vidhi/kanoon	0.488	0.404	0.125	0.357	0.963	0.194	**0.042***
upchaar/ilaaj	0.586	0.876	**0.035***	**0.000***	0.052	**0.002***	0.171
jivika/rojgari	0.742	0.538	**0.017***	0.313	**0.047***	0.762	**0.012***
yadi/agar	0.926	0.859	0.814	**0.003***	0.24	**0.011***	0.223
vishesh/khaas	**0.047***	0.201	**0.001***	0.9	**0.024***	0.063	**0.027***

vivah/shaadi	0.809	0.72	0.178	0.063	0.132	0.555	0.227

First Fixation Duration (FFD)

Table 2 shows that most word pairs do not have eye behavior that is significantly different when first encountering the stimulus or AOI in the FFD. Three out of the fifteen pairs show FFD measurements that are significantly different. The pairs that are statistically different are ‘sanvant/saal’, ‘raahi/musafir’ and ‘vishesh/khaas’. This was surprising because FFD is linked to first recognition and one would assume that there would be a difference in recognition of loanwords and root words.

‘Sanvant/saal’ is a pair in which the Urdu loanword is more commonly used. The word ‘saal’ is listed in the top 1900 words frequently used in Hindi. The average time for ‘sanvant’ is 334 ms. and the average time the eye first fixated on ‘saal’ is 233 ms. ‘Sanvant’ the Sanksrit-based word definitely shows a difference in eye behavior because it falls outside the normal FFD times. The FFD showed that in the case of this word pair when encountering a more frequently used word the FFD is shorter than when encountering a word that is less familiar.

Both ‘raahi/musafir’ are not high frequency words. ‘Raahi’ has an FFD of 196 ms. and ‘musafir’ has a FFD of 249 ms. They have a p value of 0.027, making it significant because the p value is less than 0.05. ‘Raahi’ is a Sanskrit-based word and ‘musafir’ is an Urdu loanword. For this word pair, both the results fall close to the normal range for FFD.

‘Vishesh/khaas’ also has data that is significant. ‘Vishesh’ is the Sanskrit-based word and is a more frequently used word. It has a mean time of 267 ms. which is higher than that of ‘khaas’ which is 204 ms. This discrepancy is interesting because even though one would expect ‘vishesh’ to have the shorter FFD it is ‘khaas’ the Urdu word which has the faster time. ‘Khaas’ also has a fixation time that was within the normal range for fixation times.

First Run Time (FRT)

Three of the word pairs have significant differences in FRT. ‘Samachar/khabar’ (news) have similar frequencies but significantly different mean FRT. ‘Samachar’ has an average FRT of 334 ms. while ‘khabar’ has a mean time of 560 ms. which is almost double the time of Urdu word. One could extrapolate that this result possibly indicates a longer time to process or recognize the Urdu word versus the Sanskrit word.

‘Sanvant/saal’ are again significantly different in the FRT mean times. This is as expected because ‘saal’ is a frequently used word and ‘sanvant’ is not and thus one would expect to see different eye behavior for this word pair. ‘Saal’ has a FRT mean time of 357 ms. whereas ‘sanvant’ the Sanskrit word has more than double the time of FRT of 756 ms. As prior research has indicated, this could mean that there is a longer cognitive process for the word ‘sanvant’ or in other words the eye fixates longer on this word because it takes longer to process it since it is not as familiar as the common word ‘saal’.

‘Pathshalayen/madrasa’ (school) have results that are significant. Both words are not frequently used words. ‘Pathshalayen’ the Sanskrit-based word has a mean FRT of 624 ms and ‘madrasa’ has a mean FRT of 1115 ms. which is higher than most of the other results. This conclusion is supported by previous research done by Henderson et al. (1999) and it is now considered acceptable that longer FRT indicates a problematic or more complex process occurring to comprehend the word or phrase or even picture.

Dwell Time (DT)

Dwell time is perhaps the one measurement that shows the most or indicates clearly the difference in eye behavior between the word pairs. Of the fifteen word pairs, eight pairs have mean times that are significantly different. Many researchers have found that DT is the best measurement to determine if there is a difference in how the brain processes an AOI.

‘Khed/afsos’ (regret) have significantly different eye behavior for DT. Neither words are in the top 1900 frequently used words in Hindi. In India one hears the word ‘khed’ most often when there is a cancellation or change in programming on the television. ‘Afsos’ is heard mainly in North India to convey regrets. Their p value is 0.036 for the DT. The mean time for ‘khed’ was 723 ms., and for ‘afsos’ the mean time was 991 ms. The eye takes a longer time to dwell on the Urdu loanword ‘afsos’.

‘Bhasha/zaban’ (language) also have p values less than 0.05. The Sanskrit-root word ‘bhasha’ is a commonly used word with a frequency of 1162 and an average DT of 599 ms. while ‘zaban’ the Urdu loanword has an average DT of 894 ms. The other word pairs that have eye behavior that are different are ‘sanvant/saal’ (year), ‘raahi/musafir’ (traveler), ‘sevika/madad’ (help), ‘pathshalayen/madrasa’ (school), ‘upcaar/ilaaj’ (treatment), ‘jivika/rojgari’ (livelihood) and ‘vishesh/khaas’ (important). Interestingly, if the word in the word pairs is more frequent, then they have the shorter dwell time in comparison to the word that is less frequent. There is no clear indication that Sanskrit-based words or the Urdu loanwords have shorter DT. The only correlation that is evident seems to occur due to word frequency, and whether there is word recognition.

Regression In (RI)

The data for regression in when the eye travels backward within the word as the word is being processed also shows marked differences in eye behavior for some of the word pairs. The pairs of words that have a difference in saccadic RI movements are ‘sanvant/saal’ (year), ‘raahi/musafir’ (traveler), ‘sevika/madad’ (help), ‘upcaar/ilaaj’ (treatment) and ‘yadi/agar’ (if).

No patterns are established with these results except for that there is a difference in eye movement between some of the word pairs. There is no obvious interaction with frequency, linguistic root and length of word for these measurements.

Regression Out (RO)

RO is part of the saccadic movement and differs from RI in that the eye travels out to the word prior or after the word or AOI. There is difference in saccadic RO behavior between the word pairs. The word pairs that showed a significant difference are ‘raahi/musafir’ (traveler), ‘jevika/rojgari’ (livelihood) and ‘vishesh/khaas’ (important). It is important note here that all the Urdu loanwords of these pairs had less saccadic RO movement than that of their Sanskrit loanword pair.

Skip

A skip occurs when the eye completely skips or does not fixate on the target word or AOI. Words with a higher frequency were most likely to be not fixated upon and skipped. This is true for the word pair ‘yadi/agar’ (if). There are more skips for both these words in comparison to any of the other words. The number of skips is significantly different for each. Yadi’ is in the list of most frequent words in Hindi and has more number of skips than ‘agar’ which has a lower frequency. ‘Samachar/khabar’ (news) are both frequent words and both are skipped a few times, but on average ‘samachar,’ the more common word, is skipped more than ‘khabar’. ‘Vajah/karan’ (reason) and ‘pathshalayen/madrasa’ (school) all seem to follow this pattern in which the more common word tends to be skipped more than the less common word.

Fixation Count (FC)

The FC is significantly different for more than half of our word pairs. The word pairs that demonstrated different FC are, ‘samachar/khabar’ (news), ‘raahi/musafir’ (traveler), ‘sevika/madad’ (help), ‘pathshalayen/madrasa’ (school), ‘vidhi/kanoon’ (law), ‘jivika/rojgari’ (livelihood) and ‘vishesh/khaas’ (important). One identifiable pattern with all these word pairs is that if the word is a common or more frequent word it has a smaller fixation count than that of its less common counterpart. The linguistic roots of the word did not affect the outcome but frequency of the word tended to affect the eye behavior.

A scanplot of the word pair ‘sanvant/saal’ is provided in Figure 4 to diagram the differences in the results and to visually compare the differences in eye movement of ‘sanvant/saal’. Figure 4 is the eye scanplots of the word pairs and it shows how the eye fixates longer on ‘sanvant’ the Sanskrit word which is also the less frequent word. ‘Sanvant’ is barely visible behind all the fixations and saccadic movements and the fixations all seem to be occurring at the beginning of the word.

‘Saal’ scanplot (2nd word) ‘Sanvant’ scanplot (2nd word)

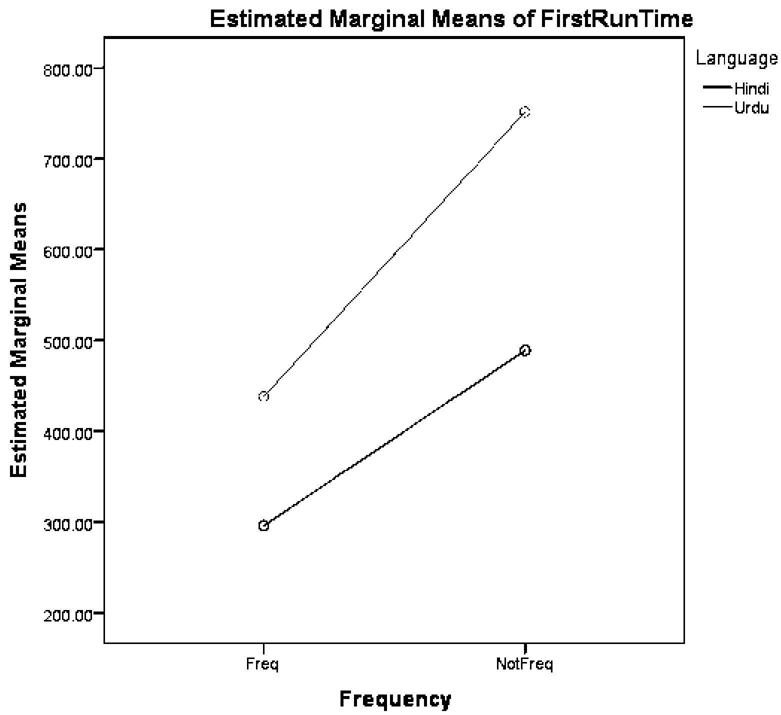
Figure 4: Comparison of ‘saal’ and ‘sanvant’ eye plots

Factorial ANOVA

The second statistical analysis that we ran is a factorial ANOVA, or three-way ANOVA, to determine the interaction between the three variables: language, frequency and eye tracking measurements. For the results from the one-way ANOVA that show significant differences we ran a factorial ANOVA to see if there is an interaction between languages: Hindi versus Urdu, frequency of the word, and the eye tracking measurements. There are two independent variables and each have two levels. Language is the first independent variable with two levels: Hindi and Urdu. The second independent variable is frequency with two levels: frequent words and infrequent words. The dependent variable is the eye measurements: FFD, FRT, DT, RI, RO, Skip and FC. Each measurement is individually processed to see whether there is an interaction with language and frequency. A $p < 0.05$ indicates a significant interaction between the variables.

The one-way ANOVA results show that there is a noticeable difference in eye behavior but the interaction is unclear. Frequency seemed to be a variable in many of the word pairs that are significantly different. In order to determine the interaction between the three variables (language, frequency and eye tracking measurements) we ran a factorial ANOVA on the word pairs that showed significant differences.

The Factorial ANOVA results show that FFD and FRT low frequency words elicit statistically different eye responses from Hindi and Urdu. Figure 5 shows the graph and the statistical analysis results of FRT. The analysis found a difference in eye behavior for Sanskrit-based words versus Urdu loanwords. The frequent Sanskrit-based words and Urdu loanwords are not significantly different yet the graph shows that the eye behaves differently when encountering these words. The FRT of Urdu words are longer than those of the Sanskrit-based words. Crucially there is a significant difference in eye behavior for infrequent Sanskrit-based words and the infrequent Urdu loanwords. We initially started this experiment to see if there is a difference in eye movement when reading Urdu loanwords in Hindi and these results reveal that there is a difference. The interaction of frequency further increases that difference in reading Urdu loanwords in Hindi.

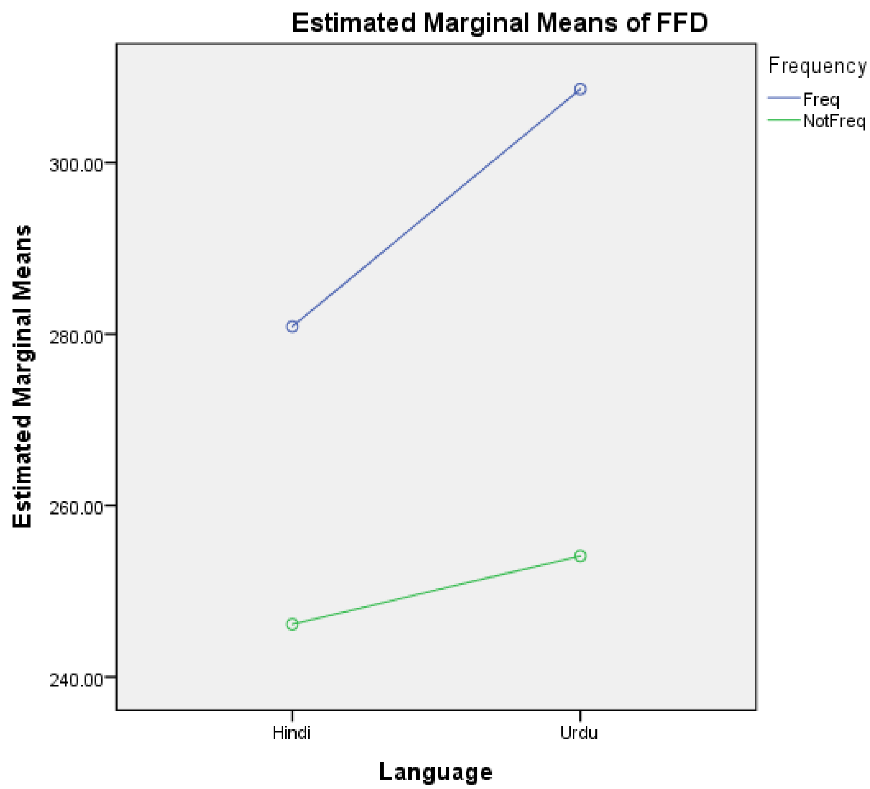


Dependent Variable: FirstRunTime

Frequency		Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Freq	Contrast	334875.249	1	334875.249	2.294	**.132**	.017
	Error	19854959.010	136	145992.346			
NotFreq	Contrast	1010450.008	1	1010450.008	6.921	**.010**	.048
	Error	19854959.010	136	145992.346			

Figure 5: Estimated marginal means of FRT interaction with language and frequency

Figure 6 presents the FFD factorial ANOVA results. The FFD results also demonstrate that there is a significant difference in low frequency Urdu loanwords and Hindi Sanskrit-based words. The p value = 0.001 reflects an even more significant difference than the FRT results of p = 0.01. The high frequency word pairs are not significantly different just as in the FRT results. Interestingly, both FRT and FFD have links to early processing and initial word recognition.



Dependent Variable: FFD

Frequency	(I) Language	(J) Language	Mean Difference (I-J)	Std. Error	Sig.[b]	95% Confidence Interval for Difference[b]	
						Lower Bound	Upper Bound
Freq	Hindi	Urdu	20.871	38.846	**.592**	-56.189	97.931
NotFreq	Hindi	Urdu	137.050*	38.456	**.001**	60.763	213.336

Based on estimated marginal means

*. The mean difference is significant at the .05 level.

b. Adjustment for multiple comparisons: Least Significant Difference (equivalent to no adjustments).

Figure 6: Estimated marginal means of FFD interaction with language and frequency

DISCUSSION

When we began this experiment there were two possible outcomes. The first was that we would see no obvious differences in reading Urdu loanwords and Hindi Sanskrit-based words. The second outcome was that the eye would register surprisal upon encountering an Urdu loanword. Our results demonstrated that in FRT and FFD measurements there is a quantifiable difference when reading Hindi Sanskrit-based words versus Urdu loanwords. Our results reinforce the second outcome, which was also the basis of our hypothesis that there would be a difference in eye behavior.

This thesis supports that assumption and further demonstrated the differences in eye behavior due to language and frequency interaction. The second question we asked was: If there is a difference in eye behavior what are the interactions that are evident in the results? The factorial ANOVA results demonstrate that the word frequency, along with linguistic root, is a variable that also affects the eye behavior when reading Hindi. The infrequent Urdu words' eye behavior is statistically different compared to the Sanskrit-root word counterpart. From our factorial ANOVA results we find that frequency is interacting with our variables of language and eye tracking measurements of FRT and FFD.

Both of these measurements, FRT and FFD, have been proven by prior research to have correlations to word recognition and early comprehension. In our case early lexical processing in word recognition and comprehension appear to be different for Urdu loanwords than for Hindi Sanskrit-based words. There is definitely a difference in eye behavior when reading the Urdu loanword and Hindi Sanskrit-based word especially when they are not frequently used words.

The eye tracking measurements indicate that the linguistic root, in this case Hindi Sanskrit-based words and Urdu loanwords elicit unique eye behavior. The more frequent words tend to have shorter FFD, FRT and DT. They also have less RI, RO and FC and some that are function words are even not fixated upon or skipped with regards to eye tracking. These results can direct further research to better answer these questions in this unique digraphic situation.

FUTURE WORK

We conclude with ideas and directions one might take to further this study into understanding better the differences in Hindi and Urdu loanwords.

Creating a Hindi corpus of word frequencies and usage including Hindi and Urdu loanwords would be invaluable to future research. Many prior corpora separate Hindi and Urdu into two different languages. It would be useful to develop a corpus of Hindi words along with the Urdu loanwords. A corpus including frequency data for both frequent and infrequent words would aid in gaining a better understanding of the interaction of frequency in the language. Another useful element in this corpus would be to also include colloquial Hindi sentences because they contain the most usage of Urdu loanwords.

A study into word length and its effect on eye tracking measurements would help determine the other variables that could impact the eye behavior during reading. One can assume that shorter words would have shorter DT and FC but how does that correlate to the word frequency and linguistic roots? Will there be a difference in eye behavior for a short Urdu loanword versus a longer Hindi Sanskrit-based word? A similar study that has paved the way for this study was conducted by Smith and Levy (2010). They studied the fixation durations in first-pass reading and how that correlated to word recognition or uncertainty.

The obvious next step to our experiment would be using minimal sentence pairs where the only difference would be in the placement of Hindi versus its Urdu loanword counterpart. This would enable to better contrast the exact eye movement differences between the two languages. Bonhage et al. (2015) conducted a similar study to see how reading is effected by predictive eye behavior. They introduced a new predictive eye gaze reading task that could be modified for a study on minimal pairs of Hindi/Urdu words.

Another area to focus on would be an in-depth literature review of eye tracking measurements and its ecological validity when correlating eye tracking measurements to comprehension and other cognitive functions. In our research we found very unique behaviors in regards to FRT measurements. A study explaining exactly how FRT correlates to comprehension would give us a better ecological answer to the differences in cognition during Hindi reading. Smith and Levy (2010) conducted a study linking First Pass Reading or FRT with word uncertainty and their methodology could be duplicated for a Hindi/Urdu study.

We also recommend further study into the digraphic situation. One of the questions that can be addressed is: How does the difference in script or orthography of the two languages correlate to eye behavior? A further step would be to duplicate our thesis with the Urdu orthography and study the Hindi loanwords in Urdu. This would determine whether Hindi loanwords in Urdu have similar eye behavior as Urdu loanwords in Hindi.

There is much more to be done to truly understand the interactions of the differences in Hindi and Urdu loanwords. Urdu is integrally a part of Hindi yet during reading Hindi there are distinctive eye behaviors for Urdu loanwords. We encourage continued research into quantifying the differences in reading Hindi and Urdu loanwords in Devanagari.

APPENDIX 1

Survey Questions

1. What is your gender?

2. Please choose your age range

□ 18 – 25

□ 26 – 35

□ 36 – 45

□ 46 – 55

□ 56 -65

□ 70+

3. Where in India are you from?

4. Where in India did you go to school?

5. Were you taught in English or Hindi at school?

6. What is your highest level of education?

APPENDIX 2

Sentences used in the experiment

हर **साल** हम मसूरी घूमने के लिए जाते।
Har **saal** ham Mussoorie ghumne ke liye jate.
Every **year** we go to Mussoorie to visit.

पिछले **संवत** बहुत बारिश पड़ा।
Pichle **sanvant** bahut barish pada.
Last **year** a lot of rain fell.

अपनी **जीविका** के लिए कोई ऐसा काम कीजिए जिससे आप पसंद करते हैं।
Apni **jivika** ke liye koi aisa kaam kijiye jisse ap pasand karte hain.
For your **livelihood** do a work that you like to do.

भारत में बच्चों को भी **रोजगारी** के बारे में सोचना पड़ता।
Bharat main bachhon ko bhi **rojgari** ke bare mein sochna parta.
In India children also have to think about their **livelihood.**

मेरा **उपचार** कौन करेगा?
Mera **upchar** kaun karega?
Who will do my **treatment**?

बहुत लोगों के पास बुखार के लिए अलग अलग **इलाज** हैं।
Bahut logon ke paas bukhar ke liye alag alag **ilaaj** hain.
Lots of people for fever have different **treatments**.

कानून अँधा नहीं है ।
Kanoon andha nahin hai.
The **law** is not blind.

विधि हम सब को सलामत रखती है।
Vidhi ham sab ko salamat rakhti hai.
The **law** keeps all of us safe.

आपके बच्चे कौन सी **पाठशालाएं** जायेंगे?
Apke bacche kaun si **paathshalayen jayenge?**
Which **school** will your children go to?

मदरसा में बहुत सी बातें सिखाई जाती हैं।
Madrasa mein bahut si baaten sikhai jati hain.
In **school** lots of things are taught.

आपको पुलिस से **मदद** कैसे मिल सकती है।
Apko pulis se **madad** kaise mil sakti hai.
How you can get **help** from the police.

उसकी **सेविका** से हम भारत पहुँच सके ।
Uski **sevika** se ham Bharat pahunch sake.
With his **help** we were able to reach India.

मैं एक **मुसाफिर** हूं।
Mein ek **musafir** hoon.
I am a **traveler**.

मेरी जीवन **राही** के साथ मैं चला।
Meri jivan **raahi** ke saath mein chala.
With my life **traveler** I went/ left.

हम सब साथ **भोजन** बहुत देर से बनाते हैं
Ham sab sath **bhojan** bahut der se banate hain.
All of us together prepare the **food** very late.

उसने बहुत **खाना** हमारे लिए पकाया
Usne bahut **khana** hamare liye pakaya.
She cooked a lot of **food** for us**.**

अगर आप के पास कोई मुशकिल है हम को बताइए।
Agar ap ke paas koi mushkil hai ham ko bataie..
If you have any problems please tell us.

वह यहां **यदि** आएगा तो में बहुत खुश हो जाऊंगा।
Voh yahan **yadi** ayega to mein bahut khush ho jaunga.
If he comes here then I will be very happy.

हमको बचपन से मां **ज़बान** सिखाया जाता है।
Hamko bachpan se maa **zabaan** sikhaya jata hai.
From childhood we are taught the mother **language**.

हिंदी **भाषा** उतर में बोली जाती है।
Hindi **bhasha** uttar mein boli jati hai.
The Hindi **language** is spoken in the north.

जब हम वहां पहुंचे तो सब से पहले हम ने हमारी **खबर** दी।
Jab ham vahan pahunche to sab se pehle ham ne hamari **khabar** di.
When we arrived there we first gave our **news**.

उसकी **समाचार** के लिए हम इंतज़ार करते रहे।
Uski **samachar** ke liye ham intazar karte rahe.
We kept on waiting for his **news**.

हम को **अफ़सोस** है की आज हम नहीं जा सकते हैं।
Ham ko **afsos** hai ki aaj ham nahin ja sakte hain.
We **regret** that we cannot go today.

नहीं आने के लिए यह मेरी बड़ी **खेद** है।
Nahin aane ke liye yeh meri badi **khed** hai.
For not coming this is my great **regret**.

उसकी **वजह** से हम काम ख़तम नहीं कर सके।
Uski **vajah** se ham kam khata nahin kar sake.
Because of his **reason** we were unable to finish our work.

तुम उसको सच **कारण** बताओ।
Tum usko sach **karan** batao.
You tell her the true **reason**.

हम सब मुंबई जा रहे हैं मेरी बहिन की **शादी** के लिए।
Ham sab Mumbai ja rahe hain meri behin ki **shaadi** ke liye.
We are all going to Mumbai for my sister's **marriage**.

बहुत समय बीत चूका है उसकी **विवाह** हो गई है।
Bahut samay biit chuka hai uski **vivah** ho gai hai.
A lot of time has passed her **marriage** has already happened.

आज मेरी **खास** सहेली आ रही है मेरे साथ खाने के लिए।
Aj meri **khaas** saheli a rahi hai mere saath khaane ke liye.
Today my **important** friend is coming to eat with me.

अब बहुत समय बाकी नहीं है बस **विशेष** काम कीजिये।
Ab bahut samay baki nahin hai bas **vishesh** kaam kijiye.
Now there is not a lot of time left only do the **important** work.

REFERENCES

Abbi, A., & Hasnain, S. I. (1986). *Lexical Modernization and Its Socio-Linguistic Effects: A Case from Indian Urdu*: Institution of Education Sciences.

Anson, C. M., & Schwegler, R. A. (2012). Tracking the mind's eye: A new technology for researching twenty-first-century writing and reading processes. *College Composition and Communication, 64*(1), 151-171.

Baker, P., Hardie, A., McEnery, T., Xiao, R., Bontcheva, K., Cunningham, H., . . . Leisher, M. (2004). Corpus Linguistics and South Asian Languages: Corpus Creation and Tool Development. *Literary and Linguistic Computing, 19* (4), 509-524.

Bonhage, C. E., Mueller, J. L., Friederici, A. D., & Fiebach, C. J. (2015). Combined eye tracking and fMRI reveals neural basis of linguistic predictions during sentence comprehension. *Cortex, 68*, 33-47.

Carpenter, P., Just, M., & Rayner, K. (1983). What your eyes do while your mind is reading *Eye movements in reading: Perceptual and language processes*, 275 - 307.

Henderson, J. M., Weeks, P. A., & Hollingworth, A. . (1999). The effects of semantic consistency on eye movements during complex scenic viewing. *Journal of Experimental Psychology: Human Perception and Performance, 25(I)*, 210-228.

Holmqvist, K., Nystrom, M., Andersson, R., Dewhurst, R., Jarodzka, H., & van de Weijer, J. (2011). *Eye Tracking: A comprehensive guide to methods and measures.* . Oxford: Oxford University Press.

Husain, S., Vasishth, S., & Srinivasan, N. (2015). Integration and prediction difficulty in Hindi sentence comprehension: Evidence from an eye-tracking corpus. *Journal of Eye Movement Research 8* (2), 1-12.

Inhoff, A. W., & Radach, R. . (1998). Definition and computation of oculomotor measures in the study of cognitive processes. *Eye Guidance in Reading and Scene Perception. In G.M. Underwood (Ed).* , 29-53.

Kaushanskaya, M., & Marian, V. (2007). Bilingual Language Processing and Interference in Bilinguals: Evidence From Eye Tracking and Picture Naming. *Language Learning, 57*(1), 119-163.

King, R. (2006). The poisonous potency of script: Hindi and Urdu *International Journal of the Sociology of Language* (pp. 43).

Marian, V., & Kaushanskaya, M. (2004). Mapping Written Input onto Orthographic Representations: The Case of Bilinguals With Partially Overlapping Orthographies. *Proceedings of the Annual Meeting of the Cognitive Science Society, 26*.

Mishra, R. K., & Singh, N. (2014). Language non-selective activation of orthography during spoken word processing in Hindi-English sequential bilinguals: an eye tracking visual world study. *Reading and Writing, 27*(1), 129-151.

Mitchell, D. C., Shen, X., Green, M. J., & Hodgson, T. L. (2008). Accounting for regressive eye-movements in models of sentence processing: A reappraisal of the Selective Reanalysis hypothesis. *Journal of Memory and Language, 59*(3), 266-293.

Morris, T. L., & Miller, J. C. (1996). Electrooculographic and performance indices of fatigue during simulated flight. *Biological Psychology, 42*(3), 343-360.

Nag, S., & Perfetti, C. A. . (2014). Reading and writing: Insights from the alphasyllabaries of South and Southeast Asia,. *Writing Systems Research, 6*, 1-9.

Nakayama, M., Takahashi, K., & Shimizu, Y. (2002). *The act of task difficulty and eye-movement frequency for the 'Oculo-motor indices'*. Paper presented at the Proceedings of the 2002 Symposium on Eye Tracking Research & Applications, New Orleans, Louisiana.

Pollatsek, A., Reichle, E. D., & Rayner, K. (2006). Tests of the E-Z Reader model: Exploring the interface between cognition and eye-movement control. *Cognitive Psychology, 52*(1), 1-56.

Pomplun, Marc, Reingold, Eyal M., & Shen, Jiye. (2003). Area activation: a computational model of saccadic selectivity in visual search. *Cognitive Science, 27*(2), 299-312.

Rayner. (1998). Eye movements in reading and information processing : 20 years of research. *Psychological Bulletin, 124*(3), 372-422.

Rayner, K., Ashby, J., Pollatsek, A., & Reichle, E. D. (2004). The effects of word frequency and predictability on eye movements in reading: Implications for the E-Z Reader model. *Journal of Experimental Psychology: Human Perception and Performance, 30*(4), 720-732.

Rayner, K., Slattery, T., Drieghe, D., & Liversege, S. (2011). Eye movements and word skipping during reading: Effects of word length and predictability. *Journal of Experimental Psychology - Human Perception and Performance, 37*(2), 514-528.

Richardson, D. C., & Daleb, R. (2005). Looking to understand: The coupling between speakers' and listeners' eye movements and its relationship to discourse comprehension. *Cognitive Science 29*, 1045-1060.

Roberts, L., & Siyanova-Chanturia, A. (2013). Using eye-tracking to investigate topics in L2 acquisition and L2 processing. *Studies in Second Language Acquisition, 35*(2), 213-235.

Rodd, J. M., Gaskell, M. G., & Marslen-Wilson, W. D. (2004). Modelling the effects of semantic ambiguity in word recognition. *Cognitive Science, 28*(1), 89-104.

Rooy, B. V. , & Pretorius, E. (2013). Is reading in an agglutinating language different from an analytic language? An analysis of isiZulu and English reading based on eye movements. *Southern African Linguistics and Applied Language Studies, 31*(3), 281-297.

Smith, N. J., & Levy, R. (2010). Fixation durations in first-pass reading reflect uncertainty about word identity. *In Proceedings of the 32nd Annual Meeting of the Cognitive Science Society*, 1313–1318.

Spinner, P., Gass, S. M., & Behney, J. (2013). Ecological validity in eye-tracking: An empirical study. *Studies in Second Language Acquisition, 35*(2), 389-415.

Van Orden, K. F., Limbert, W., Makeig, S., & Jung, T. P. (2001). Eye activity correlates of workload during a visuospatial memory task. *Hum Factors, 43*(1), 111-121.

Winke, Paula M. (2013). Eye-Tracking Technology for Reading *The Companion to Language Assessment*: John Wiley & Sons, Inc.