A DUAL EYE TRACKING STUDY OF THE INFLUENCE OF COLOR AND GAZE CUES ON THE USE OF REFERRING EXPRESSIONS IN A SITUATED FARSI DIALOGUE ENVIRONMENT

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

BY

SARA RAZZAGHI ASL

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

SEPTEMBER 2015

A DUAL EYE TRACKING STUDY OF THE INFLUENCE OF COLOR AND GAZE CUES ON THE USE OF REFERRING EXPRESSIONS IN A SITUATED FARSI DIALOGUE ENVIRONMENT

Submitted by **SARA RAZZAGHI ASL** in partial fulfillment of the requirements for the degree of **Master of Science in Cognitive Science, Middle East Technical University** by,

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

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

Assist. Prof. Dr. Murat Perit Çakır _____
Supervisor, **Cognitive Science, METU**

Examining Committee Members:

Prof. Dr. Cem Bozşahin _____
Cognitive Science, METU

Assist. Prof. Dr. Murat Perit Çakır _____
Cognitive Science, METU

Assist. Prof. Dr. Cengiz Acartürk _____
Cognitive Science, METU

Prof. Dr. Deniz Zeyrek Bozşahin _____
Cognitive Science, METU

Assist. Prof Dr. Murat Ulubay _____
School of Business, YBU

Date: September 11,2015

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 wok.

Name, Last name: Sara Razzaghi Asl

ABSTRACT

A DUAL EYE TRACKING STUDY OF THE INFLUENCE OF COLOR AND GAZE CUES ON THE USE OF REFERRING EXPRESSİONS IN A SITUATED FARSI DIALOGUE ENVIRONMENT

Sara Razzaghi Asl
MSc. Department of Cognitive Science
Supervisor: Assist. Prof. Dr. Murat Perit Çakır

September 2015, 92 pages

The aim of this study is to explore the structure of Farsi referring expressions (RE) used during a collaborative Tangram puzzle solving activity, and investigate the role of different visual cue conditions on the types of RE and the degree of gaze coordination. A jigsaw task design was used which required participants to work as a team to solve Tangram puzzles in three conditions where (a) all pieces had the same color (normal condition), (b) all pieces were assigned a distinct color (color condition), and (c) all pieces had the same color but the partner's gaze information was visualized on the screen (gaze cueing condition). In this respect, two main aspects were under scrutiny: linguistic and dual eye-tacking analysis, while both are assumed to be enriched recourses for modulating joint attention. For this purpose, a corpus of Farsi REs in a situated dialogue environment is constructed to evaluate the frequency of specific RE's features and their length distribution. Descriptive statistics show that Mosallas (Triangle), Un (That) are the most frequently used RE words in the Farsi corpus. The RE feature distributions are compared with Turkish, Japanese and English RE corpora compiled with the same task to provide a cross-linguistic analysis. Conversational analysis of features of REs revealed the prominent role of color terms in identifying objects and the striking influence of shape and size in gaze cueing condition. Besides, cross-linguistic analysis results demonstrate that Farsi is distant to all languages in this respect. In case of dual eye-tracking analysis, results were not influenced significantly under different status and also along six trials. However, there was a significant interaction effect between conditions and trials especially for the color case.

Keywords: Dual Eye-Tracking, Referring Expressions, Tangram, Discourse annotation, Farsi Language Resources

ÖZ

FARSÇA DİYALOG ÇEVRESİNDE RENK VE GAZE ETKİSİNİN YÖNLENDİREN İFADELER KULLANIMINDA İKİLİ GÖZ İZLEME YÖNTEMİ İLE İNCELENMESİ

Sara Razzaghi Asl

Yüksek lisans Bilişsel Bilimler Bölümü

Tez Yöneticisi: Yrd. Doç. Dr. Murat Perit Çakır

Eylül 2015, 92 sayfa

Tez çalışmasının amacı iki kişinin işbirliği yaparak ortak bir Tangram görevini icra ederken kullandıkları Farsça gönderge ifadelerinin yapısal özelliklerini incelemek ve değişik görsel yardımcıların kullanılan gönderge ifade türleri ve göz koordinasyonu düzeyi üzerindeki etkilerini araştırmaktır. Bu amaçla iki kişinin ortak bir ekran aracılığıyla sadece ses ve ekran paylaşımı yardımıyla iletişim kurabildiği ve iki kişinin göz hareketlerini aynı anda takip edebilen bir deney düzeneği kurulmuş ve katılımcılardan bir birlerini sözel olarak yönlendirerek hedef şekli oluşturmaya çalışmaları istenmiştir. Katılımcılar parçaların renksiz olduğu, renkli olduğu ve bir birlerinin göz izlerini görebildiği üç farklı durumda ikişer tangram görevi tamamlamıştır. Bu ortamda ortak algının oluşumunda rol oynadığı varsayılan dilsel süreçlere ve ikili göz izlerine odaklanılmıştır. Yapılan deneyler sonucunda bir Farsça gönderge ifadeleri derlemi oluşturulmuş ve en sık kullanılan ifade türleri ve uzunluk dağılımları hesaplanmıştır. Elde edilen dağılımlar benzer bir deney ortamında oluşturulmuş Türkçe, Japonca ve İngilizce için oluşturulmuş derlemlerden elde edilen sonuçlarla kıyaslanmış, ve Farsça'nın bu dillerden gönderge ifadesi tür dağılımı bakımından farklılık gösterdiği gözlenmiştir. Gönderge ifadelerinin özellikleri incelendiğinde renk terimlerinin ağırlıklı olarak nesnelere gönderme yapmak için kullanıldığı, göz izi paylaşımı yapılan durumda ise şekil ve boyut özelliklerinin daha sık kullanıldığı görülmüştür. Göz izleri arasındaki örtüşmenin renksiz, renk ve göz izi paylaşım durumlarından anlamlı olarak etkilenmediği, ancak ilk ve ikinci denemeler ayrı ayrı analize dahil edildiğinde renkli parçaların olduğu durumda ikinci deneme sırasında diğer durumlara göre daha yüksek bir örtüşme oluştuğu gözlenmiştir.

Anahtar Kelimeler: Çift göz izleme, Yönlendiren İfadeler, Tangram, Söylem açıklama, Farsça Dil Kaynakları

To My Family …

ACKNOWLEDGMENT

Firstly, I would like to express my sincere gratitude to my advisor, Assist. Prof. Dr. Murat Perit Çakır, for the continuous support of my research, for his patience, motivation, and immense knowledge. His guidance helped me in all the span of research and conducting experiments.

Besides my advisor, I would like to thank METU Cognitive Science circle especially Prof. Dr. Cem Bozşahin, Prof. Dr. Deniz Zeyrek, Assist. Prof. Dr. Cengiz Acatürk and Assoc. prof. Dr. Annette Hohenberger for their precious supports and guides.

I would like to express my deepest thanks and appreciation to my parents for devoting their life to my progress with their endless love and spiritual support throughout my life, who taught me to love and scatter it.

My special thanks goes to Salar Vayghan Nezhad for standing by me and presenting his continuous helps and supports along writing this thesis. Who carried me long miles, when I thought I could not go on.

I am also thankful to Mr. Hakan Guler as a greatly valued member of Informatics institute for his continuous help and his kindly answers to my many questions during my study.

My sincere thanks go to my friends: Sanam Dehghan, Mahsa Ramiz for their kindness and encouragements.

I would also like to appreciate my home mate Mona Zolfaghari Borra For her patience, love and inspiring presence during writing my thesis.

I wish to acknowledge the help provided by Sardar Vayghan Nezhad.

I am indebted to Delasa and Razzaghi families for their various forms of support and motivation.

I am also grateful to my friends Kianoosh Ghasemi, Yasamin Dehghan, Arezu Hoseini, Navid Hoseini, Negin bagherzadi, Sona khaneshenas, Yalda hoseyni, for their supports and encouragements.

I am also thankful to METU COGS eye-tracking lab members especially Mehmetcan Fal for his instructions.

Last but not least, I would like to thank all my friends who kindly participated in my thesis experiment.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

AOI: Area of Interest

EOS: End of Sentences

GH: Givenness Hierarchy

LSD: Least Significant Difference

RE: Referring Expression

TT: Turn Taking

CHAPTER 1

INTRODUCTION

1.1. Motivation of the Study

Recognizing objects in conversational domain is embraced by the usage of referring expressions (RE's) as linguistic tools. RE's deal with the way participants lead or maintain each other's attention to specific entities (Spanger et al, 2012; Gundel & Heldberg, 2008). Thoughts and attention can be reflected in discourse and eye-movements; so that referring expressions hand in hand with eye-movements could be indicator of joint attention. Current study is aimed to evaluate eye-movements and the usage of Farsi's referring expressions for pairs in Tangram collaborative domain. These aspects are assessed under the influence of different cues like color and gaze.

1.1.1. Social Cognition

Analysis of social behavior among pairs, team members and larger scale societies have recently attracted increasing interest in cognitive science (Dale et al., 2011). Hutchins (1995) highlighted that in group activities participants become part of a united cognitive system and often act in a way which is different from their individual performances. Swarm cognition, swarm robots, dual-eye tracking methods are examples in which social cognitive processes play a fundamental role. Getting involved with collaborative issues may reveal promising breakthroughs for improving our understanding of disorders such as Autism and also it can inform many questions about social robotics and multi agent systems. Due to these reasons, there is a growing interest towards studies focusing on distributed cognitive systems.

As thinking may ripple through behavior, eye-movements and discourse mirror thoughts and attention. This study evaluates the integration of two aspects in collaborative problem solving: first one is dual-eye tracking analysis and the second one is linguistic analysis.

1.1.2. The Importance Of Joint Attention Analysis By Dual Eye Tracking

According to Richardson and Dale (2005), eye movements may serve as reliable resources for analyzing the status of mind and attention; therefore, in order to

understand a social cognitive system and observe cooperative processes in detail it seems necessary to evaluate how members of the system think, process and interact under different circumstances. In this path, eye tracking methods provide the opportunity to collect simultaneous eye-movements of participants in a collaborative context (Nussli & Jermann, 2012; Acartürk & Cakir, 2012). Recently the analysis of alignment in collaborative problem-solving tasks have attracted increasing interest from researchers (Janarthanam & Lemon 2009; Buschmeier et al. 2009). Some eye tracking studies are co-analyzed with fMRI, EEG, bodily movement and corpus data (Holmqvist et al., 2011). Regarding the shortage of related resources, despite their importance, dual eye tracking methods become preferable for empirical investigations. Besides, interpreting eye movements over previous findings have shown that the level of success in two person's communication is related to the coordination of their eye movements in dual-eye tracking experiments (Richardson and Dale, 2005).

The findings of previous studies suggest that characteristics of the environment and mental processes together affect eye movements (Richardson and Dale, 2005). The current study aimed to extend the dual-eye tracking paradigm by using different clues to see whether these factors may create differentiation in the level of gaze coordination and to explore gaze coordination dynamics in relation with linguistic phenomenon.

1.1.3. Conversational Analysis

There is an impartible interrelationship between language and cognition. Language has an omnipresent role in directing eye-movements as indicated in many studies. Referring to Nüssli (2011, p. 22) "Gaze is largely influenced by speech which is at the heart of collaboration". Meantime, as language is a complicated phenomenon that goes hand in hand with many different aspects such as gestures, bodily orientations, studying it in isolation cannot cover all questions about the organization of collaborative interaction.

Conversation involves substantial use of language and sometimes it may deviate from the rules or grammar of language (Clark and Wilkes-Gibbs, 1986). Furthermore, conversation is beyond creating chain of words distributed into a sequence of turns; rather it is a social activity in which contributors' struggle to reduce mutual comprehension effort (Clark and Schaefer, 1989). Also, it is claimed that participants take part in dialogues with their existing opinions, presumptions and information and contribute them into the conversation, which gradually forms a common ground or mutual knowledge during interaction (Cole, 1978; Clark and Schaefer, 1989).

Pairs interact via two phases: presenting intentions from the speaker's side and reflecting acknowledgements and acceptance from the listener's side. As the discourse moves forward, and the alignments in participants' communication increases, interlocutors establish and maintain a common ground of shared referents that accumulates and encodes changes as well as new information. In this manner, participants may even create their own common lexicon during conversation (Clark and Schaefer, 1989). Therefore, investigating the details of these procedures could

imply many points in understanding how interlocutors reach and maintain mutual understanding in interaction.

1.1.4. Referring Expressions

Apparently, mulling over related studies, in linguistic or specifically in situated dialogue settings, referring expressions operate like a fountain which is used for irrigating the ground for reaching efficient mutual comprehension. They equip conversation in a way to create intelligible collaborative environment for recognizing objects, directing and maintaining attention. Thus, it is worth to investigate how language or in lower level RE interferes with attention and cognition (Spanger et al, 2012; Gundel & Heldberg, 2008).

There are several studies that focus on the categorization of REs in English based on corpus data gathered under different circumstances. For instance, as reported by Acartürk & Çakır, (2012), COCONUT corpus (Di Eugenio et al., 2000) is a pool of REs in English collected during a 2-D design task coordinated via text based communication, while QUAKE (Byron & Fosler-Lussier, 2006) and SCARE (Stoia et al., 2008) are based on interactions in a three dimensional environment. Although, these corpus studies have revealed many important aspects of REs, each of them constrained participants' activities in specific ways. For instance, the task used for the COCONUT study restricted the participants to the text-based interaction without supporting extra-linguistic aspects like gestures or prosodic features, whereas the tasks used for QUAKE and SCARE studies confined participants with some limited activities like picking up and dropping things. Spanger et al. (2011) provided a natural collaborative environment to eliminate some of those restrictions and as a result constructed the REX-J corpus for Japanese referring expressions. Acartürk & Çakır (2012) used the same situated dialog task designed by Spanger et al. (2011) in an effort to build a Turkish corpus of referring expressions.

The current study is motivated by the observation that there are not many studies focusing on the structure and type of referring expressions in Farsi. To address this gap, the current study employs the situated dialog task designed by Spanger et al. (2011) to build a corpus of Farsi Referring expressions corpus. Moreover, the study also investigates additional factors such as color and gaze cues on the distribution of RE types in Farsi, and thus aims to contribute new perspectives into the study of referring expressions in a situated dialog context.

1.2. Aim of the Study

The purpose of this study is to assess the coupling relationships between two Farsi speakers who collaboratively attempt a Tangram puzzle in a computerized collaborative problem-solving environment. The research scope is limited to the use of RE's and the alignment between the eye movements of the participants as enriched resources for estimating where peers are attending to, and for investigating the factors

which relate to directing and allocating attention. In this path, different cue conditions were included in the study; namely having access to uniquely colored puzzle pieces and to a visualization of the participant's eye gaze on the shared screen. This setup was used to pursue the following research questions.

1.3. Research Questions

This dissertation is motivated and directed by the following research questions:

1-How does the gaze alignment of directors and operators differ while solving Tangram puzzles in different visual cue conditions such as colored puzzle pieces and gaze cueing, in comparison with the normal condition?

2-How do Farsi, Turkish, Japanese and English languages compare to each other in terms of their percent distribution of Referring Expression categories observed in the same situated dialog setting?

3-How does the distribution of features like shape, color and size used in referring expressions change in different visual cue conditions?

4-Do the length of the Referring Expressions used and the number of turns taken change across different visual cue conditions?

5-Is there a relationship between the length and frequency of Farsi REs used and the degree of gaze overlap among different visual cue conditions?

6-What is the functional role fulfilled by referring expressions "in" (this) and "an/un" (that) in this situated dialog context? Does their usage change based on the role (i.e. instructor vs presenter) assumed by the speaker?

1.4. Thesis Outline

The next chapter contains brief theoretical background information about eye tracking, referring expressions hand in hand with the reviews over the mentioned topics. In the third chapter, the experimental setup, materials for dual eye tracking and linguistic analysis and the data acquisition process are explained. The fourth chapter presents the results of the study. The fifth chapter concludes the thesis with a discussion of the main findings.

CHAPTER 2

LITERATURE SURVEY

This chapter is divided into two parts and since the aim of the study is to integrate analysis of dual eye-tracking and referring expressions, the related background and reviews are emphasized hereunder. The first section provides information about the human eye, theories of color perception and the development of eye tracking research including the recently emerging dual-eye-tracking paradigm. Next, the Tangram puzzle is introduced, which will be the shared task used in this study. Finally a review of studies about referring expressions and referring expressions in Farsi are provided.

2.1. Background and Review of Eye Tracking

The current section is assigned for general survey of eye tracking as a rapidly growing technology and also it is preferred in this study. It contains concepts such as human eye, eye movements and eye-tracking technologies due to form a background for this survey and makes the study more intelligible.

2.1.1. Human Visual System

2.1.1.1. Human Eye

As eye trackers work based on light reflection from the pupil and the cornea; hereunder, a brief overview of the physiology of the human eye and the mechanism of human visual system are provided which will be used to describe the basic principles underlying recent eye movement analysis techniques.

The eye ball is composed of an aperture area and a photosensitive area (Nüssli, 2011). Light rays enter into the eye via the pupil pass through the lens, forming an inverted image over the retina at the back of the eye sphere (Holmqvist et al., 2011).

Retina is full of light sensitive cells which convert the entered light into electrical signals and dispatch them to the visual cortex through optic nerves for subsequent processes. The aperture part of the eye has many different parts; prominently, the lens and the ciliary muscles are responsible for focusing on a place. In this path ciliary muscles adjust lens curvature and manage focal distance; while pupil and iris tune the intensity and the rate of light which reaches the retina.

On these grounds, it is stated by Nüssli (2011) that dispersion of sensitive receptors' density is not the same in every part of the retina; hence, it causes different levels of precision and vividness in our vision. Fovea is located at the center of retina and because it has denser amount of receptors, it creates the most accurate sight in comparison with surrounding parts of the retina with two degrees range. So human beings move their eyes to bring the pictures into the central part of their retina to be able to see things at higher resolution.

Moreover, the human eye is endowed with three pairs of muscles containing vertical, horizontal and torsional directions, in order to control the eye movements in three dimensions. The brain is involved to govern these muscles to shift the direction of gaze towards specific locations in the visual scene (Holmqvist et al., 2011).

For the measurement of eye movements, reflection of both cornea and pupil play an important role and some devices provide their average reflections.

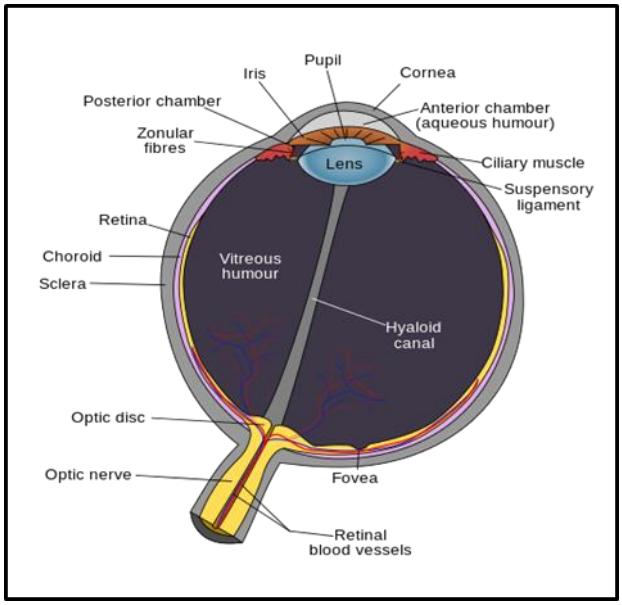


Figure 1 Schematic diagram of the human eye (Nussli, 2011, p. 16)

2.1.1.2. Color Perception

Since this study focuses on the role of color references in collaborative interaction, brief information about the neurobiology of color perception will be provided in this section. As mentioned above the retina is full of photosensitive cells which are called cones and rods. Cones distinguish color within small receptive fields in the visual field, whereas rods detect changes in light intensity over larger receptive fields that provide opportunities for sight in low lighting conditions as well as motion perception (Holmqvist et al., 2011). Duchowski (2007) stated that there are nearly 120 million rods and 7 million cones in the human retina. As the white light contains the entire spectrum, when it radiates over an object some of the spectrum is absorbed and some are reflected. The reflected colors form the observer's color realization. The human brain perceives color via a neural pathway that primarily involves input from the cone cells in the retina (Figure 2) as well as higher level visual processing in the brain in regions such as V2 and V4.

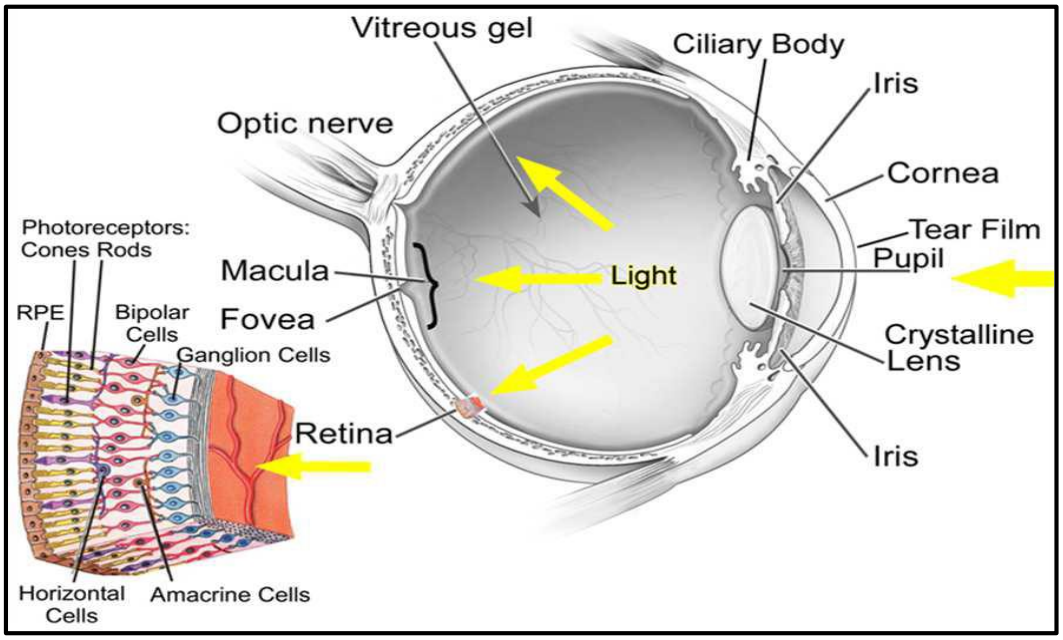


Figure 2 Conversion of light into electrical signals by rods and cones (Vera-Diaz & Doble, 2012, p. 120)

Thompson (2013) states that there are three kinds of cones, which are sensitive to red, green and blue light spectra, as well as short, medium and long wavelengths of light (Figure 3). Collection of signals from three of the cone cells form a color span which the eyes can detect.

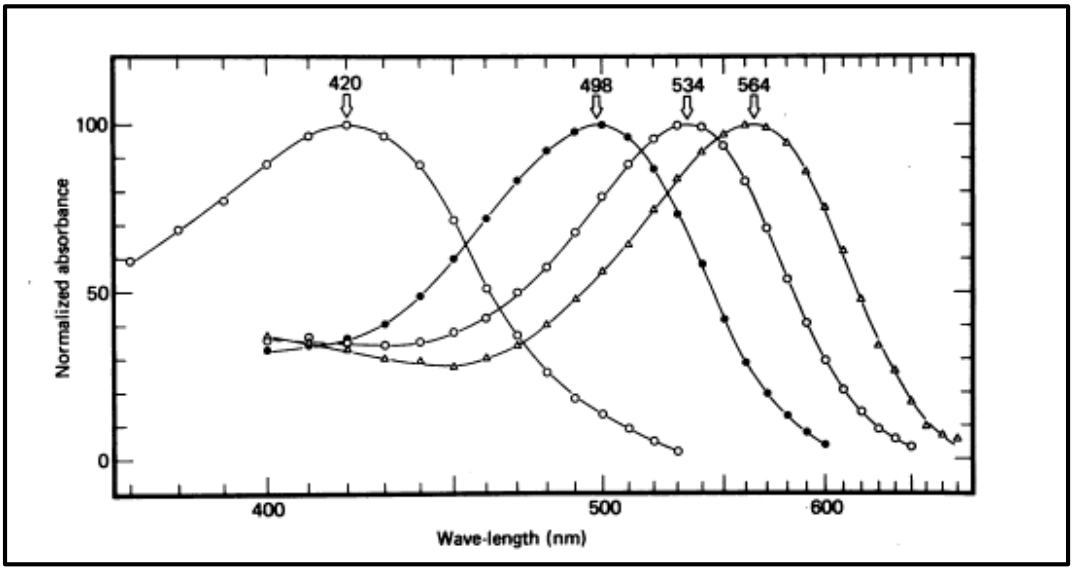


Figure 3 Relative perception of light absorbed.420nm is the mean wavelength of blue sensitive cones, 498nm is the mean for rods, 534nm is the mean of green cones, and 564nm is the mean of red sensitive cones (Bowmaker & Darnall, 1980, p.505).

2.1.1.3. Eye Movements

After a brief summary of basic physiological properties of the human eye, this section provides a summary of basic types of eye movements that are typically monitored by eye tracking techniques. According to Holmqvist et al. (2011) the data yielded from eye trackers mostly stand for eye fixation locations rather than its movements. Eye fixations refer to specific positions in the visual space where the eye stays put for a short span of time, which hints at where the subjects allocates his/her attention on the visual scene. Saccades are the prompt movements spanning between two fixations. It is also reported that saccades are one of the fastest human movements along which human beings remain sightless. In addition to fixations and saccades, there are additional types of eye movements such as the smooth pursuit, which is about following a moving object such as a bird in the sky. There are also micro-movements like tremors, drifts and micro saccades which are respectively involved with indefinite directing muscles, diversion from a fixed point and retaking eye to the fixed point (Nüssli, 2011; Holmqvist et al., 2011).

2.1.2. The Evolution of Eye Tracking Methods

According to Holmqvist et al. (2011) the first eye trackers were made in late 1800s. Rayner (1998) characterizes the historical development of eye tracking methods in three distinct periods. In the first period, Javal was the pioneer who was the first to consider the role of eye movements during reading in 1876. Along that period up to 1920, some related aspects like saccadic delay, its prevention and understanding interval revealed. The second period get involved with usage of eye tracking techniques such as bench-mounted and head-mounted eye trackers, but the dominant behaviorist paradigm at the time restricted eye tracking studies. After the 70's decade, the third period started and under the auspices of technological improvements mobiles enriched with eye trackers and omnipresence existence of eye trackers made cognitive studies easier.

Along these periods the most significance progresses were the usage of lens system with mirrors by Yarbus and Ditchburn between 1950s and 1970s, which could gather data accurately but the contact lenses used were bothersome for the subjects (Figure 4). Electromagnetic coil systems which evaluate the electromagnetic excitation in silicon contact lenses were another successful progress but anesthesia was needed and also lenses should be adjusted exclusively for each person's eye. Another breakthrough was Electrooculography (EOG) in which electromagnetic changes measured by muscle movements were measured with electrodes places around the eyes. Despite its affordability the EOG method had issues in precision due to drifts. The Dual Purkinje systems were expensive but precise and there was no need to enter it in to the eye, but afterwards it was understood that saccadic terminations scaled insufficiently (Holmqvist et al. 2011). As result, recently many devices are equipped with eye trackers and it is the most dominant technique for recording eye movements' data. Researchers measure fixations and saccades of people to find out the direction of gaze,

which mirrors the process of thinking and also presents the things which are in center of visual attention of people (Duchowski, 2007).

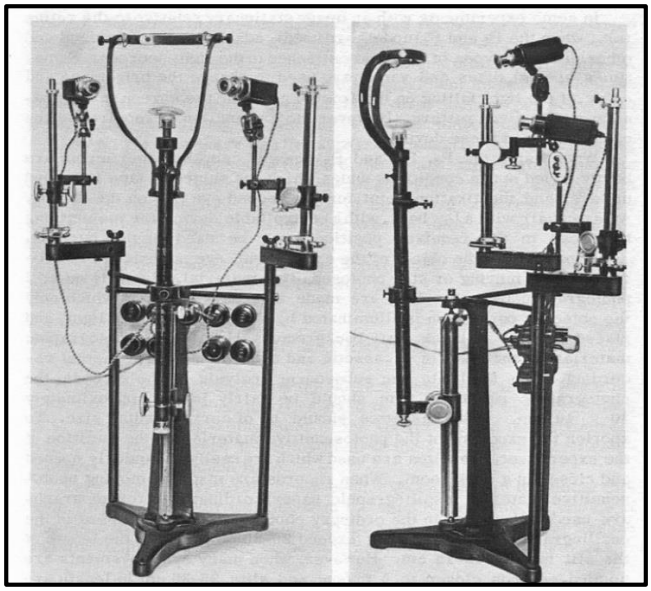


Figure 4 Eye tracker from the 1960s (Yarbus, 1967, p. 41)

Ensuring data accuracy is an important concern in eye tracking methods. Krash and Breitenbach (1983) stated that slight changes in adjustments create massive differentiations in estimations of fixation locations. Mostly eye trackers need a calibration for generating accurate estimates of gaze direction. During the experiments the quality of calibration may decrease due to changes in head position, which is one of the drawbacks of eye tracking methodology (Nüssli, 2011). This study aimed to reduce error prone points by attempting to calibrate participants' eye trackers whenever it is hindered and splitting the participants gaze spans precisely.

As eye trackers can help researchers decode many aspects of human thought and attention they are employed in many studies in the domain of psycholinguistics. In most cases eye tracking measures are coupled with other methodologies. In this dissertation, eye movements were studied hand in hand with verbal and linguistic analysis.

2.1.3. The Relationship between Eye Movements and Cognition

According to Nüssli (2011) it is not a secret that our eyes perform the leading role in perceiving the environment. Considering the fact that cognition rests largely on visual perception, it is rational to expect strong relationships between eye movements and cognitive processes. In fact, with the advent of eye-tracking methods, deciphering such relationships turned into an issue of broad interest to researchers. A famous study in this field is by Yarbus (1967) where a number of subjects were presented with a certain picture, and were asked to do a variety of tasks ranging from mere observation to more complex ones, which required them to make specific inferences such as guessing

people's age, economic status, or telling if they were relatives or not. Final results suggested that subjects' cognitive activities are closely linked with their eye movements, as shown in Figure 5. Yarbus thereby concluded that the viewer's eyes are driven by the cognitive process involved rather than by the visual content.

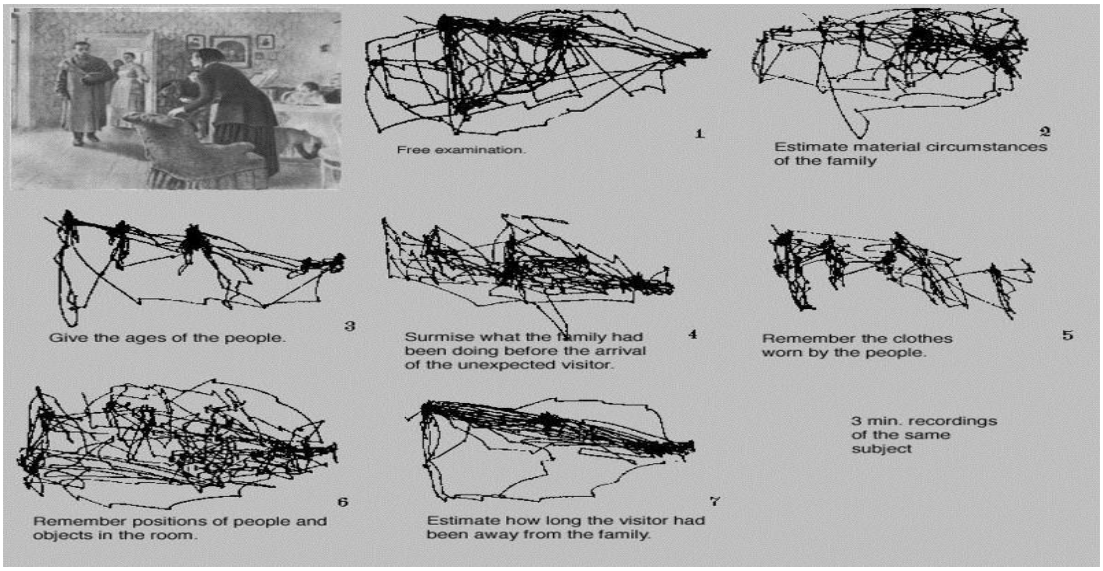


Figure 5 Differences of eye movements based on the asked task. Studied by Yarbus (1967) (Nussli, 2011, p. 21).

After reviewing the progress of eye tracking techniques and methods for analyzing eye movements, it seems necessary to highlight the existence breakthroughs in joint attention and dual eye tracking studies. These studies have offered key insights for understanding the mechanisms underlying joint attention and mutual understanding (Nussli & Jermann, 2012). This study aims to contribute to this line of work by investigating the role of referring expressions and visual cues in a dual eye tracking paradigm. In the next subsections relevant concepts for dual eye tracking paradigms are briefly explained.

2.1.4. Joint Attention

According to Butterworth et al., (1995) joint attention is the ability of sharing common focus on something among two or more people. It also involves with gaining, maintaining and drawing attention via verbal and non-verbal indications. Analyzing over alignment of attention helps bringing out people's intention, point of view and their social skills. Likewise, pairs' eye movements and gaze directions are influenced by visual characteristics of the world, what they hear, interact and process in their mind (Dale et al., 2005). As a result based on Hutchins (1995) both persons become a part of a collective or ensemble, and begin to act and react in a coordinated manner in such a way that is different than their individual performances in that domain. While human is a social creature it is beneficiary to understand the interplay of it with its situated

surrounding. It seems a large number of existing studies on dual gaze analysis focus on infants. In addition to this, Gustafsson et al., (2015) reports that there are also studies on gaze behavior of animals such as bird species and chimps, which tend to focus on differences between human and animal gaze following behavior. In the case of humans there are also numerous studies involving gaze tracking. Expressly, some studies support the importance of analyzing joint attention. For instance, the study of Sharma et al. (2015) provides, via a multiple eye-tracking method, information on the gaze distribution of a teacher during taping a Massive Open Online Course (MOOC) video, and uses the gathered data to show how student attention can be guided by teacher's actions. The findings suggest that the presentation of the gaze of the teacher to students helps them with pinpointing the intended content, which positively contributed to their understanding of the course material.

2.1.5. Review of Dual Eye Tracking Studies

Earlier research studies on the eye movement behavior during experiments that involve collaboration tasks provide convincing evidence that a speaker's gaze on a referent precedes by some time the oral mention of it. In other words, the point a speaker puts under their gaze gives clues as to what they are about to speak of shortly afterwards. This time gap between the fixation location and the mention of an object is called the eye-voice span. A voice-induced eye movement can likewise be found among listeners, occurring shortly after a subject is referred to by the speaker, which is called the voice-eye span (Nüssli, 2011).

The study of Richardson and Dale (2005) investigated the correlation between the eye movements of a speaker and their auditor. The speaker was shown a television show on which they were to present spontaneous comments. These comments were recorded and then re-played parallel with the televised show to a group of listeners. Having recorded the eye movements of both speakers and listeners, a cross-recurrence analysis of their ocular activity confirmed that the listener's eye movements trailed in good approximation behind that of the speaker by some two-second delay. In practice, the better this approximation was, the higher performance the listener exhibited on a comprehension test they were subsequently given. A following experiment conducted using low-level optical cues to guide the listeners' eye movements showed that these visual cues could influence the listener's latency in answering the comprehension test. Apparently, in the same way that an individual's shifts in attention can be monitored via their eye movements, the degree to which a bilateral communication is likely to succeed can be determined through the degree of coordination among the speaker's and the listener's eye movements.

Dale et al., (2011) provided another study in which a Tangram based shape ordering task was used for assessing the degree of cooperation among two teammates whose aim was to establish together the final position of a number of abstract geometrical

shapes. While the challenge lies in evolving an understandable way of reference to the shapes, the ultimate arrangement is revealed to only one of the participants, known as the director, and it is the second participant, the matcher, who is to produce the same ordering. In the digital design of Tangram, tracking participants' eye movements clarified that both time efficiency and eye-movement synchronization of the team improved through the three-round performance. To quantify this inter-personal harmony a cross-recurrence analysis was employed, which was later used to show that as the verbal discrepancies were resolved over time, the whole actions could be more perfectly modeled as an integrated system.

There are several other studies using the dual eye tracking paradigm in different joint task conditions. For instance, Sharma et al.'s (2013) dual-eye tracking study focused on the relationships between the discourse formed in the course of a pair programming comprehension task and the partners' eye movements in different timeframes. Four layers of interaction episodes are identified, each of which extends throughout the entire conversation. The purpose of this study is to find the links between different layers in different timescales. Outcomes pointed out the interaction between the level of realization and gaze parts but there was not a direct interplay between gaze and dialogue episodes while there was relation between gaze for level of operations and dialogue. In another study, Jermann et al., (2012) investigated how selection sharing among participants of program comprehension tasks can influence their visual navigation patterns. To this end, using a cross-reference analysis, the gaze patterns of forty couples were recorded while performing such tasks. The final result achieved shows a direct relation between gaze cross-recurrence and grounding efforts (including text selections) which the couple exerts to achieve a reference for mutual understanding. Selections in the form of broadcast, on the other hand, appear to act in place of indexing sites for the selector, since they, immediately after coming on the scene, draw the attention of non-selectors. Highest rate of gaze recurrence meanwhile is found when words are added to the selections.

Another study by Sharma et al., (2012) details, using dual eye-tracking analysis, how mutual understanding in a pair-programming experiment can facilitated by consistent and sequential gaze ordering between two speakers. With participation of forty pairs of programmers, the analysis was conducted on their gazes gliding on structural elements of the code, identifiers and expressions. As different from the code, the identifiers and the expressions draw more instances of tracing the data flow from successful communicators than from less harmonious programmers. Besides, moments when the partners' attention converges towards a single point coincided with more organized execution of the code and less switch of attention among identifiers and expressions.

Obviously, eye movements are influenced by the features of the environment and the nature of joint activity (Richardson and Dale, 2005). As this research evaluates the dual eye tracking and verbal comprehension assessments under different features of Tangram workspace; it is necessary to describe the characteristics of this game to clarify the reason why a Tangram task was selected for the current study.

2.2. Tangram Problem Solving

Referring to Sternberg (2004) and Solcum (2001) tangram is a traditional Chinese game which is one of the most famous dissection puzzles in the world. Tangram is composed of seven geometric pieces called “tans” that are used for creating various shapes, including 2 large right triangles, 1 medium right triangle, 2 small right triangles, 1 medium square and 1 parallelogram. Pieces are arranged in a way to form an outline without overlapping. Tangram game is used as a procedure for improving geometric spatial thinking by evaluating the characteristics of shape and relationship between its pieces (Scarlatos et al., 2002; Sedighian & Klawe, 1996). When tangram is used for teaching geometry among groups, the interactions often lead to profound thinking, reasoning and problem solving (e.g., Coleman, 2008) and also a deeper understanding of logical implications of specific visual configurations (Clements & Battista, 1992).

In the current study the visual joint attention and the delay of attention for two people while collaborating to solve tangram problems will be pondered; also the role of color clue over these factors will be investigated. Tangram puzzle game is shown in (Figure 6).

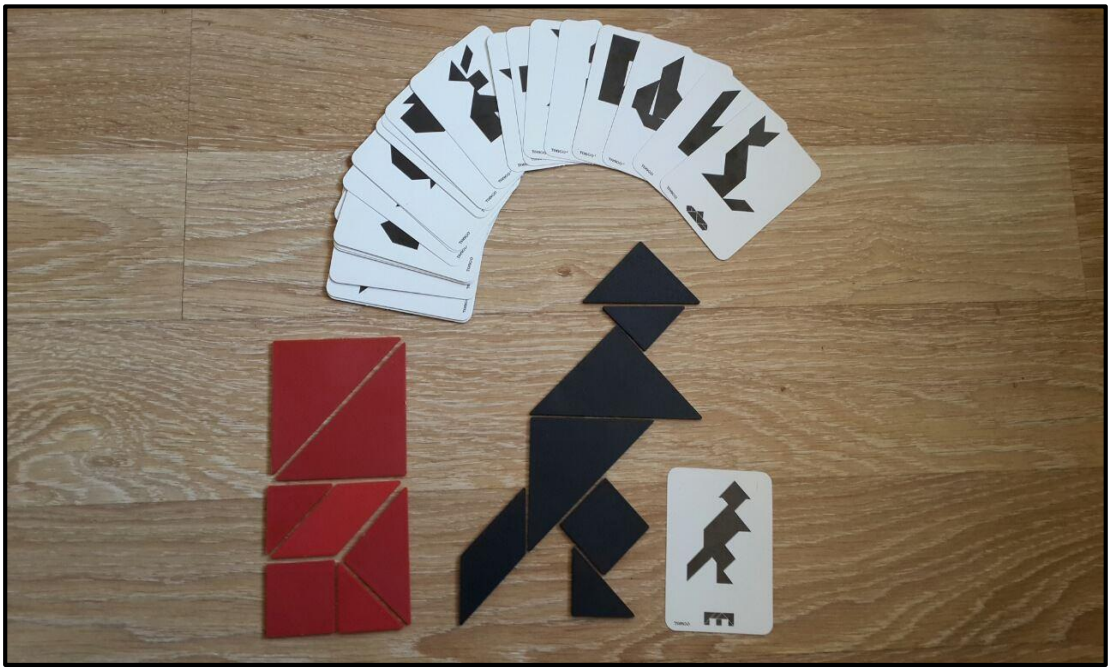


Figure 6 Tangram puzzle game

Lin et al. (2011) conducted an experiment based on collaborative Tangram problem solving in which children taught geometry in a virtual workspace over tablets. The environment provided by special learning and problem solving tactics for participants who were twenty five elementary students. The study revealed that student’s

capabilities in manipulating, rotating, spatial sensing and reasoning in social domains improved by their negotiations, conducting each other and receiving acknowledgements. Also the distance between better and worse ability children was diminished. Spanger et al. (2011), Richardson et al. (2011) and Acartürk & Çakır (2012) used Tangram puzzle in their studies. This game is also preferred in the current research.

Conversation effects gaze considerably, as stated by Nüssli (2011), meantime, language plays a fundamental role in the achievement of joint tasks such as collaborative tangram solving. According to Clark &Wilkes-Gibbs (1986) during a conversation speakers and listeners use various linguistic resources such as inserting repair sequences, questions and acceptance to establish and maintain a common ground to support their ongoing interaction. Referring expressions are tools of language to negotiate over identifying objects in the scope of conversation (Gundel & Heldberg, 2008). The prominent effect of referring expressions in recognizing objects and shortage of studies in Farsi for RE domain motivated the study to go to this direction. Meantime, omnipresent role of some referring expressions for recognizing objects put forward the idea to evaluate which factors of Tangram domain remain in conversation, in the path of reaching mutual understanding under the existence of different cues. For this purpose, some striking aspects of REs are described in further parts.

2.3. Referring Expressions

Recently many studies have focused on the classification of referring expressions, particularly in English. One of the studies in the case of evaluating the role of referring expressions in a collaborative domain was conducted by Clark &Wilkes-Gibbs (1986) in which participants conversed about ordering complicated shapes (Tangram shapes). During the experiment participants created a common ground based on the context and their own beliefs and with time their common ground changed and expanded (Cole, 1978; Clark and Schaefer, 1989). Two main phases for interplay are presentation and acceptance phases. Along conversation adjacency pairs (two sequenced utterances) form the contribution tree based on these two phases (Clark and Schaefer, 1989). As a result, along trials the number of words and the number of turn taken decreased, and participants could solve the task with progressively less conversational effort. Regarding the important role of referring expressions in leading mind and attention hand in hand with aim of this research, some striking points for referring expressions, and reviews over them are declared underneath.

2.3.1. Cognitive Vision Over Features of Referring Expressions

Some references are dominance in the conversations for representing things and permanent and temporary features are defined for objects. Referring to Clark &Wilkes-Gibbs (1986), enduring features are constant characteristics like shape, size and color and the temporary ones are like location, orientation which can be changed by operations. As people tend to use the identification terms that they have used before repeatedly to refer to the same thing, the role of permanent properties in reaching

mutual comprehension becomes important. It is claimed that when pairs struggle to reduce joint effort over referents they should pick and maneuver over permanent characteristics, which was confirmed and selected in the study of Clark &Wilkes-Gibbs (1986) by 90 percent of abundance. In object recognition scope Braje et al. (1999) commented that temporary cues has less impact over object recognition. They declared that degrading or removing temporal factors does not influence referent recognition. Also, changing sharpness, texture and so on could not deteriorate the flow of recognition while the shape was constant. This study is settled in a way to evaluate how the usage of terms over permanent factors of tangram pieces fluctuate under the existence of cues such as Color and partner's eye movements.

2.3.2. Reviews of Studies on Referring Expressions

As stated in Spanger et al. (2011), TUNA (Van Deemter, 2007) is one of the largest corpus of English referring expression including about two thousand REs, but it is restricted to REs produced by single person. GRE3D3 (Dale and Viethen) is the corpus for individual's relational expressions and it contains less amount of expressions in comparison with TUNA. As discussed by Acartürk & Çakır (2012) COCONUT corpus (Di Eugenio et al., 2000) has a pool of referring expressions, the conversation is mediated by text messages with obliged turn takings and the shared environment involves a 2-D design task in which participants buy and organize things in 2 rooms. Based on Spanger et al. (2011) the COCONUT corpus is similar to the TUNA corpus as it persuades participants for producing rather simple statements. Also it is stated that COCONUT covers three kinds of features: problem solving speech features, speech and entity features without considering extra linguistic aspects. Such restrictions can be considered as a drawback for the COCONUT corpus. QUAKE (Byron & Fosler-Lussier, 2006) and SCARE (Stoia et al., 2008) are the names of other existing RE corpora in English. SCARE is an improved version of QUAKE and both involve communications in a 3-D environment that require the use of location based references.

All of these studies had the disadvantage of being far from real dialogues. In order to bridge that gap, nowadays studies are involved with evaluating referring expressions in situated dialogues. For example, referring expressions coupled with pointing movements. Also the relationship between visual information and referring expressions deliberated. But still there was lack of resources in the respect of the convergence of referring expressions with contributors' operations. Eriksson (2008) worked on the act of RE over the face to face interaction of language and bodily movements simultaneously in which particularly demonstrative expressions are scrutinized, it revealed that the mix of demonstrative expressions and bodily gestures like pointing cannot be adequately satisfactory for participants, the rate of existence repairing sequences, is the evidence for that. To fill the gap Spanger et al. (2011) conducted Tangram simulator to produce REX-J corpus due to gather Japanese and English REs. The same simulator is used afterwards with Perit Çakır & Acartürk (2012) for constructing Turkish collection of REs and here in this research it is used for preparing Farsi corpus of REs.

2.3.2.1. Theoretical Frameworks Related to Referring Expressions

Hereunder, the role of referring expressions based on different frameworks is studied. Firstly, Givenness hierarchy is mentioned. Secondly, the Centering Framework is considered.

2.3.2.1.1. Givenness Hierarchy Framework

The Givenness Hierarchy Methodology which was offered by Gundle et al. (1993) defined six levels of cognitive statues for referring expressions in language discourse and for bringing something in the focus of attention. Protocols are assigned and determiners and pronouns restricted information for allocating the referent clarification in one of the status. The statuses are shown in Table 1. These statues over the hierarchy indicate the state of memory and attention from the most narrowed (in focus) to the least narrowed (type identifiable) (Gundle et al., 2003). The introduced hierarchy was also supported by experiments for research of dispersion of referring expressions in five languages containing (English, Japanizes, Mandurian Chinese, Russian and Spanish).

Table 1 Givenness hierarchy and associated forms in English obtained from (Gundle et al. 2003).

In focus	>	**Activated**	>	**Familiar**	>	**unique identifiable**	>	**Referential**	>	**type identifiable**
It		this, that, this N		that N		the N		indefinite this N		a N

Likewise, another cross linguistic study by (Gundle et al., 2010) was conducted to develop the Givenness theory for referring expressions prediction over Eegimma, Kumyk, Ojibwe and Tunisian Arabic. The results revealed three points in this respect: firstly, language can address differentiations with higher levels of hierarchy, if it can address differences in two adjoining levels. Secondly, two higher levels of hierarchy are distinguishable by all languages. Finally, there are not special formations for languages to address the differentiations between two levels.

2.3.2.1.2. Centering Framework

Another studied framework in purpose of clarifying the place of referring expression is Centering Framework. Grosz et al. (1995) asserted that, some entities are in the focal point of conversation in comparison with other entities, and it enforces restrictions over the usage of different kinds of referring expressions. Also, it is stated that the coherence of speech is influenced by the adaptability between the usage of referring expressions and centering attitudes. Linguistic structure is composed of parts there can be local coherence (coherence in the same part) and global coherence (coherence with other parts). Centering framework cares about local coherence.

Yoshida, E. (2008) explored how center transition patterns in the centering framework changed with the type and distribution of referring expressions. A unified interpretation is proposed in the study to understand the behavior of referring expressions in spoken language by looking at the connection between referential selection and local and global coherence of discourse. In a broad view, the research seeks: (1): to depict through a contrastive analysis an outline of semantic and pragmatic referring expressions commonly used in English and Japanese natural discourses, (2) to analyze the way in which anaphoric and deictic expressions can determine the discourse structure and can underline specific part of discourse segment and (3) to review how referring expressions have been selected and distributed in Map Task Corpus, and to shed light on how participants work together to decide the chief referents against their widely accepted common ground.

Needless to say, Yoshida, E. (2008) claimed that the two languages diverge from one another when the form of reference is considered from a grammatical standpoint. However, the process through which topic entities are suggested, formed and altered to following topic entities appears largely alike, hence comparable in the two languages. Setting side by side the choice and the distribution of referring expressions of the four different transition patterns of centers led the study to key factors involved in the corresponding relations between Japanese and English referring expressions. These key elements show that in discourse, topic chains of noun phrases are created and dealt with like proper names. This, in turn, indicates that when the topic entity is formed as the conversation develops, a full noun phrase has a major part to play. This is, in the main, because the existing centering model fails to cover noun phrases topic chain in anaphoric relations as far as the local focus of discourse is looked at. Therefore, to include both pronouns and those full noun phrases which used for continuations across segment boundaries, the centering needed to be incorporated with a model of global focus. It can be derived from Walker's cache model that anaphors do not always necessarily appear in shorter forms. Likewise, as opposed to (zero) pronouns, the line of noun phrases help keep the attention focused both within a discourse segment and when overstepping segment boundaries. These processes are expected to regulate other applications of language as well. As a result, diluted reference forms should not be readily taken as a clue to the degree of focus of attention, nor should using full noun phrases necessarily be seen as a sign of a shift in focus. What is more, when moving across segment boundaries, the anaphoric relations link with deictic expressions thanks to expansion to global coherence of discourse. Finally, this writer believes the selection and arrangement of reference expressions in the Map Task Corpus is influenced by the way participants work jointly to judge the weightiest entity in the current discourse against their common ground.

Yoshida (2011) also investigated the connection between discourse entities on the one hand, and topic chaining and discourse coherence on the other, by showing how referential choice and distribution can define the center transition patterns in the centering framework. The application of English and Japanese referring expressions frequent in a variety of real-life settings has also been studied, along with theoretical frameworks applied and developed with a view to explain local and global discourse coherence. The methodology adopted by Yoshida mainly centers around a discourse-

based integrated method of anaphora resolution where integrated criteria for reference expression usage is suggested.

While the study is over referring expressions distribution in Farsi corpus, it seems necessary to explain Farsi briefly.

2.4. Persian Language

According to the Persian literature Encyclopedia by Bruijn (2015) Persian is a language in an Indo-Iranian branch of the Indo-European languages. The literary form of New Persian is known as Farsi in Iran, where it is the country's official language. There are approximately 110 million Persian speakers Worldwide. The language is spoken in Iran, Afghanistan and Tajikistan and its affiliation can be seen in Figure 7.

Figure 7 Persian speaking area retrieved from:
http://www.iranchamber.com/literature/articles/persian_language.php

Indeed, word order in written Persian is SOV although it can be different in conversational dialects. In Noun phrases (NP) and Propositional phrases (PP), Persian language acts like head initial. (Amtrup et al., 2000). To clarify, in short syntactic categories (verb, noun, adjective, and adverb) in Farsi are demonstrated here:

Nouns are head of noun phrases such as:

"Khorshid-e derakhshan" (The Shiny Sun- خورشید درخشان)

Sun Shiny

Verbs are mostly located in final position of sentences: (e.g., Raftam is verb means I went)

"Man be ketabkhane raftam" (I went to the library) من به کتابخانه رفتم

I to Library Went refers to I

Adjectives modify nouns and they can appear before and after nouns. (e.g., Bozorg (Big) بزرگ - koochak (Small) کوچک are adjectives)

Adjectives which come after the noun take genitive particle «e» or «ye»:

pesar e khub (good boy) پسرخوب - khane ye bozorg (big house) خانه ی بزرگ- mosallas e bozorg (big triangle) مثلث بزرگ

Adverbs modify verbs, adjectives or other adverbs similar to English. Adjectives can appear in adjectival phrase (AP) or alone

(e.g., Kheili (Very) خیلی- Hamishe (Always) همیشه Hargez (Never) هرگز are adverbs)

Man hargez sigar nemikesham. (I never smoke) (من هرگز سیگار نمی کشم)

Kheili bozaorg (very large) خیلی بزرگ

2.5. A Review over Referring Expression in Farsi

Considering Farsi, there were not many resources in respect of referring expressions. The only research is about the role of null referring expression in a conversational context. Shokouhi (1996) claimed that English conversations start with full NP and continues with pronominal forms, while Farsi conversations starts with NP but continues with null referring expressions. Null RE defined as "non-occurrence of overt nominal or pronominal form". In this study, which is conducted over normal everyday dialogues, two Persian conversational contexts are discussed. In both, apparently, null referring expressions best work in an unmarked form to track referents. In one context, a referent acts as the central figure of the discourse, when in the other, a general schema explored by Fillmore (1975), Prince (1981) and Chafe (1987) is involved. Contrary to Persian speakers who typically favor null referring, English speakers lean towards a pronominal form. Below, part of dialogue which was used by Shokouhi (1996) is selected.Null REs are marked inside parenthesis with bold.

B: rästi äqä-ye Mehrabän dige raft/?

Really Mr. Mehraban yet went

'By the way, did Mr. Mehraban leave?

A: oun ham,

He also

'Well, he'

B: dige ne-mi-yäd//,

Anymore doesn't **(he)** come back'

A: na dige raft,

No just went

'No **(he)**'s left for good'

General schema based on Fillmore (1975) is composed of schemas which connect and form a structure that can be at the same time part of the other frame works. Afterwards, it was declared by Fillmore that frames activate each other. Meantime, from Prince's (1981) outlook, general frame is the series of smaller schemas which are brought out from the main one. For Chafe (1987) the former attitude, create a cognitive view point as it get involves with new information and the subsequent one regarded accessible.

Overall, Shokouhi (1996) underlined that whether a general schema exists in the context or a referent is cast as the protagonist, null referring expressions are what a Persian speaker is expected to prefer. The English speaker, on the other hand, tends to use pronominal in the first case, and pronouns in the latter. Finally, it seems that applying a cross-linguistic approach to discourse structures proves more promising than when the study is confined to merely syntactical analysis of sentences and terms based on the previous studies. As mentioned in the review of RE, there are researches over the evaluation of RE features in Japanese, English and Turkish but there is a gap for Farsi in this domain, in order to bridge that gap this study is formed to address the percentages of distribution of referring expression and to assess the cross-linguistic analysis over them.

CHAPTER 3

METHODOLOGY

This research is conducted over a dual-Tangram problem solving experiment in order to analyze two main aspects related to the achievement of joint attention in a collaborative domain, including linguistic analysis of referring expressions in a Farsi corpus and a dual-eye tracking analysis of participants under different cue conditions such as using color and gaze cues. Also it is aimed to observe whether and how these two aspects relate to each other. In this path, after pondering over the background of the study, an experiment is designed to reach these objectives. This chapter introduces the research questions, design of the study and all the materials and procedures used for collecting and analyzing the data in our corpus.

3.1. Research Questions

This thesis study aimed to investigate the following research questions:

1-How does the gaze alignment of directors and operators differ while solving Tangram puzzles in different visual cue conditions such as colored puzzle pieces and gaze cueing, in comparison with the normal condition?

2-How do Farsi, Turkish, Japanese and English languages compare to each other in terms of their percent distribution of Referring Expression categories observed in the same situated dialog setting?

3-How does the distribution of features like shape, color and size used in referring expressions change in different visual cue conditions?

4-Do the length of the Referring Expressions used and the number of turns taken change across different visual cue conditions?

5- Is there a relationship between the length and frequency of Farsi REs used and the degree of gaze overlap among different visual cue conditions?

6-What is the functional role fulfilled by referring expressions “in” (This) and “an/un” (That) in this situated dialog context? Does their usage change based on the role (i.e. instructor vs presenter) assumed by the speaker?

3.2. Design of Study

This study employs mixed methods (Clark & Creswell, 2011) to pursue the research questions listed above. Participants’ eye movement coordination, the percent of referring expressions used, the number and the length of the turns taken during conversation comprise the quantitative data, whereas excerpts that illustrate the functional use of pronouns constitute the qualitative data in this study.

3.2.1. Participants

Five pairs (2 male and 8 female) among Middle East Technical University students were recruited for the experiment. Those participants’ native language was Farsi and eight of them knew Azeri as well. They were mostly masters or PhD students except for one person who was at the undergraduate stage. Participants were majoring in similar fields such as engineering, informatics and physics and they were acquainted with basic mathematical and geometrical topics. They were grouped into same gender pairs who knew each other in order to eliminate pleasantries between them. Among these dyads there wasn’t any one who uses thick glasses or contact-lenses with special filters (there was one more group which was excluded because of this problem which caused inconsistency in gaze data) and they didn’t have any eye disorders like color blindness. Two roles were defined for each pair of participants including the instructor and the operator. The roles were switched after each trial. Each participant assumed both roles in all three conditions.

3.2.2. Apparatus

In order to record participants’ eye movements synchronously, two identical Eye Tribe trackers with a sampling rate of 60 Hz were used (Figure 8). The eye trackers were mounted at the base of two identical HP Pavilion laptops with Intel Core i7-4510U processors and USB 3.0 compatible mother-boards. A third desktop was used to record the problem solving moves, gaze visualizations and the sound. A pair of microphones and head phones were also used so that, peers could clearly hear and communicate with each other.

Figure 8 The Eye Tribe Tracker

3.2.3. Software

In order to work with the Eye Tribe tracker, Eye Tribe SDK was installed which contains Tracker SW and Tracker UI programs. A custom Java program developed at the METU COGS Eye Tracking Lab was used to connect to the eye trackers, stream gaze data across two clients, and to visualize and record dual gaze data. In order to facilitate collaboration via screen and mouse sharing, the TeamViewer software was used. Tangram Simulator and Player software (Spanger et al., 2009, 2010; Tokunaga et al., 2010) were used to host the puzzle solving sessions and for analyzing the collected data. For transcribing Word documents and for annotation Excel documents were used. To simplify the counting process in Excel, the Kutools software was used. A custom Java program developed at the METU COGS Eye Tracking Lab was again used to compute the distribution of raw gaze data on specified areas of interest on the screen, and to produce scarf and gaze recurrence plots to assist gaze coordination analysis. In addition, for screen and voice recording CamStudio Recorder software was used. Statistical analyses were conducted with SPSS v.22.

3.2.4. Location and Positioning of Participants

The experiments were conducted at the METU COGS Lab in a quiet atmosphere. Pairs sat in the same room back to back and they could not see each other. Contributors sat in front of the monitor at a distance of approximately 60 cm.

The Eye Tribe UI is equipped with a track box which is beneficiary for positioning participants appropriately; it has a model eye which is mirroring the current state of both eyes. Each eye should be place in each side of diagonal line and if the color of condition is green it indicates that the participant's location is acceptable and one can proceed to the calibration stage. Figure 9 shows acceptable and un-acceptable instances.

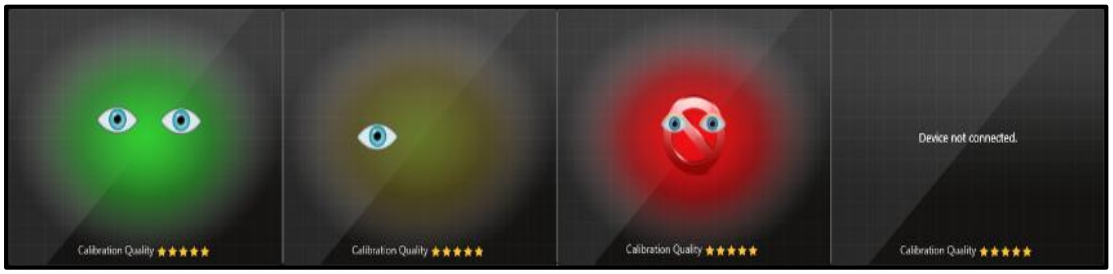


Figure 9 Quality of tracking which indicate good eye tracking, limited tracking and error message conditions. Retrieved from http://dev.theeyetribe.com/start/

3.2.5. Pilot Study

Two Turkish pairs from the METU Informatics Institute were recruited for pilot tests before performing the main tests to correct deficiencies in the experimental design and observe probable difficulties. It was examined to see if the time interval for playing and for hints' appearance was enough. We also checked if participants use color terms while solving the problems. They reported some misunderstanding due to the difference between the scales of the target image and the pieces used, so the main pairs were informed about the size mismatch between target shape and working environment shapes. The pilot study suggested that the time span seemed sufficient and they used color references such as "the blue" along collaborative trials instead of using long phrases such as "one of those big triangles". The participants declared that the task would became very difficult if they weren't given any hints. They also they said that the activity was fun and they feel good because of knowing their peers during game. Audio and screen records and their quality were checked, then the groups' eye tracking data was evaluated and there wasn't any specific problem. The gaze data of both participants were visualized on the shared screen simultaneously in the gaze cuing condition. The visibility of the gaze information of both participants were reported to cause distractions. Therefore, only the partner's gaze was visualized in the main experiment with an improved smoothing algorithm to reduce the distraction caused by the real-time gaze cursor visualization.

.

3.2.6. Experimental Setup

In this experiment dual eye tracking method was applied. Participants were composed of 10 students (2 male and 8 female) grouped in 5 pairs with same gender. Pairs collaboratively worked on solving 6 Tangram tasks through the shared screen via Team Viewer software. Tangram puzzle is an ancient Chinese dissection puzzle in which seven pieces form a target shape (Sternberg, 2004; Solcum, 2001). The collaborative sessions were conducted with the Tangram Simulator Software (Spanger et al., 2009, 2010; Tokunaga et al., 2010) which gives the opportunity to move, flip and rotate the geometric pieces with the mouse. Figure 10 indicates the screenshot of simulator environment in the normal condition.

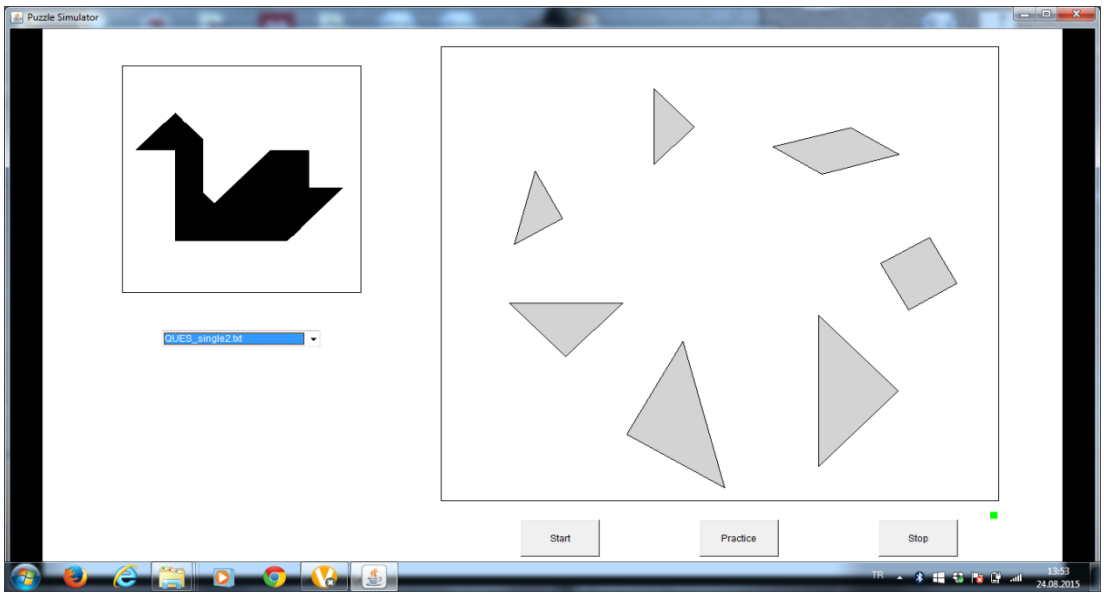


Figure 10 Screenshot of tangram simulator environment.

It should be regarded that as well as screen sharing through Team Viewer software, peers could communicate verbally via headphones and microphones. Participants were deliberately placed back to back, to eliminate the effects of communicating via face, eye and body gestures. Two roles are defined for the participants which are director and operator. The director or instructor is a person who can see both the target and the content of game. He/she can lead via communication but cannot use the mouse. The operator is a person who can manipulate the game area via the mouse but cannot see the target. Therefore, both parties need to work together as a team to produce the goal shape by using the 7 Tangram pieces. After each trial peers' role were switched. Tasks were categorized in three groups, so by switching roles, each of the peers had both roles (director and solver) under each condition. Game states included Normal condition, Color condition and Gaze cueing condition and the order of conditions were counterbalanced and randomly assigned to one of the 6 puzzles, for instance one group faced with two colored game at first then two normal games and finally two gazed

games while for another group the condition's sequences was like Gaze, Color, Normal. In the Normal condition participants attempt to solve puzzles by using pieces that have the same color. In the colored condition the Tangram pieces are colored with unique colors. In the gaze cueing condition peers can see their counterparts' eye movements as a small circle on their screen. The second and third conditions are conducted as cues in order to observe differentiations in joint attention of couples and usage of Referring Expressions for distinguishing pieces. Screenshots from colored and gaze cueing conditions are shown in Figure 11.

Selected colors for Tangram colored games in Farsi are ("آبی"“Abi” (blue), "زرد "“zard” (yellow), "بنفش"”banafsh” (purple/violet) "طوسی"“tosi”(gray), "سیاه/مشکی"”siyah/meshki” (black),"سبز" ”sabz” (green) and"قرمز" “Ghermez” (red)). Selected colors term's length in terms of pronunciations in Farsi are composed of three or four letters.

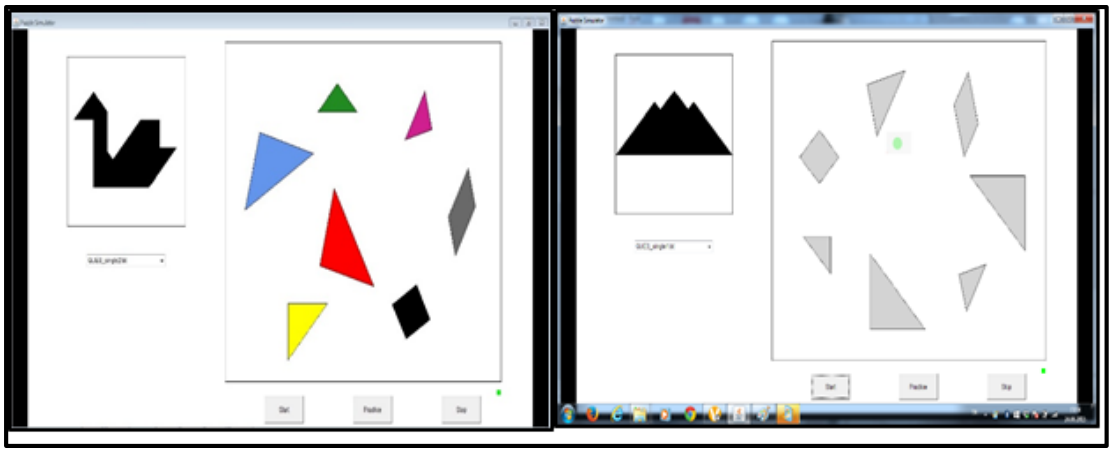


Figure 11 Using color and Gaze cues over Tangram Game

Six goal shapes which were used along the experiments are indicated hereunder they contained both symmetric and asymmetric abstract shapes with different levels of difficulty (both Geometrical and detailed abstract ones (Figure 12). Pairs had at most eight minutes for solving each puzzle and if they couldn't solve it they encountered with a "time over" message. It should be considered that while playing, every 3 minutes one hint appears on the screen which indicates the right place of one of the pieces.

It should be regarded that based on differentiations over programming protocols, for one of the groups (G5) the gaze condition was slightly different and they could see their own gaze movements beside their partner's eye movements. In other groups, participants in gaze condition just see their counterpart's eye motions.

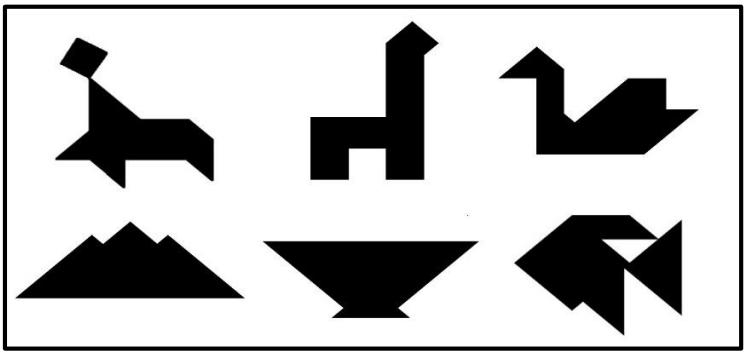


Figure 12 Target shapes of the experiments

3.3. Procedure of Data Collection and Calibration

After participants' placement calibration was done, during calibration players were asked not to move their head and hands as much as possible and asked to follow a moving circle on the screen in order to calibrate the device as explained below:

Calibration is done in order to teach computer the state of participants' eyes when they are fixed in a special places on the screen. For calibrating, the Eye Tribe UI was used by selecting calibration part after adjusting the location of candidate (Figure 13).

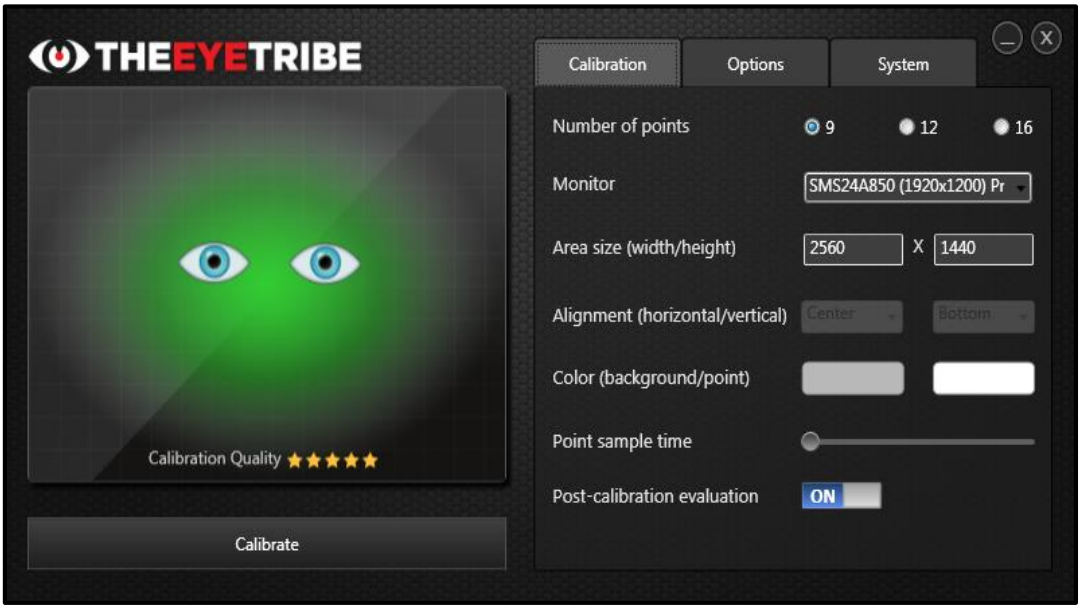


Figure 13 UI retrieved from http://dev.theeyetribe.com/start/

When the calibration process starts, contributors are asked to follow dots which appears one by one on the screen. 9 points part is selected for this purpose to be shown in the screen. That takes 20 seconds and result of it which is displayed by number of stars should be at least good to go through further stages. The calibration page is shown in Figure 14.

Figure 14 Eye tribe's calibration with 9-points obtained from http://dev.theeyetribe.com/general/

After calibration, participants were trained over a demo exercise to try working in the Tangram environment by altering shapes to feel comfortable about main tasks. Then they started to solve puzzles and while they were collaborating to form target shapes, the third computer recorded their movements, gaze visualization and audios in order to use them for constructing the Farsi RE corpus. Besides, the eye tribe server collected eye movements of pairs' members and as a result data were represented by (x, y) coordination of gazes of the screen.

3.4. Type of Data and Data Analysis

After collecting data among Farsi speaker pairs, two types of data produced and analyzed afterwards. First type is dual eye tracking method in which the coordination of eye fixations over the screen was gathered for both directors and operators. For measurement cross recurrence analysis used to represent the alignment of participants' eye motions by passing through the rounds of different operating circumstances. The second data type is the recorded voices and recorded screen movements in the intention of building a Farsi corpus. 30 dialogues were generated which were transcribed and annotated to clarify the role of conversational RE and its features scattering in absorbing and maintaining attention over the objects when endeavoring upon solving the problems quickly and in collaborative way.

3.4.1. Cross Recurrence Analysis

The study was focused on raw gaze coordinates extracted from the eye trackers. The shared screen was organized into 16 equal rectangular AOIs (Area of interests) and target area to observe where eye fixations of pairs land on the screen, considering time factor. The eye movements' data was equipped with time stamps exported from Eye tribe software. Coordination of two time series for director and operator in the scope of tasks chunked based on start and finish time of each trial and two stream of gaze data compared with cross recurrent analysis which quantitatively measuring dual eye fixation data as it is done by Richardson and Dale (2005) and Richardson et al. (2007). As mentioned above, over the screen 16 AOI defined on working area beside the target area to allocate where each peer focus at each time. The split of workspace into 16 parts are indicated in Figure 15.

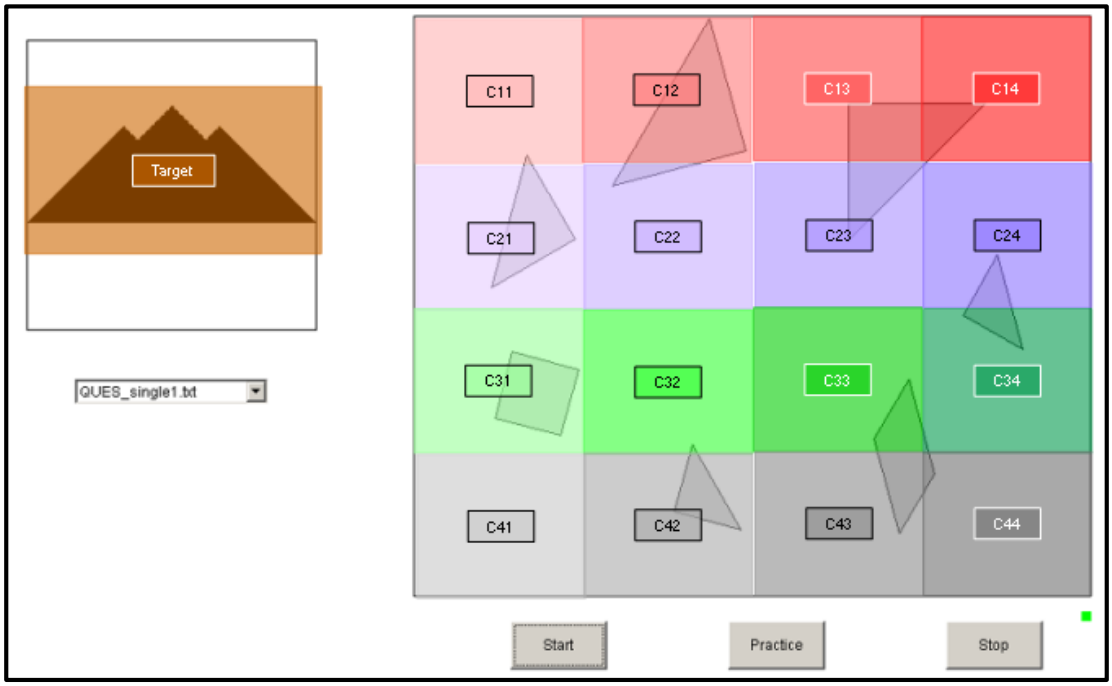


Figure 15 The division of working space into 16 equal AOI to show pairs fixation overlap Çakır & Acartürk, (2013)

This information is input and the output is a scarf plot which indicates the areas among 16 parts in which they had a gaze match. It is composed of three rows, row 0 shows dispersion of eye gaze for first member of pair and in same way row1 shows it for second member, while, row 3 is their gaze overlap. For each AOI one color is assigned for example in picture above C11 is presented with light pink also it is shown by same color in gaze plot of Figure 16. T is for target and C indicates eye movements out of the area of interest these colors are shown as the protocol under each scarf plot. Also this program gives the opportunity to zoom in and out to extract more qualitative details.

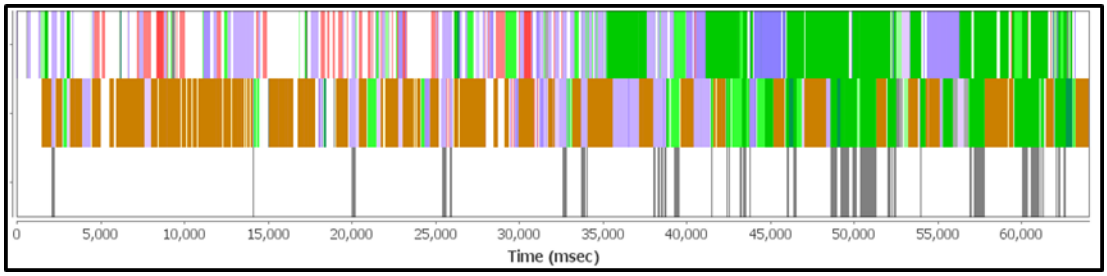


Figure 16 Scarf plot for AOIs retrieved from poster of Perit Çakır & Acartürk, (2013)

Software also provide a summarized diagram to observe the scattering percentage of recurrence between game couples' gaze patterns with the delay span which is adjusted between -2 and +2 seconds for each trial Figure 17. According to Richardson and Dale (2005) there is a delay of two seconds between directors and operators gaze match so it seems logical to assign the delay time in same way. Zero point indicates the percentage of recurrence in accurate gaze overlap. This diagram (Figure 17) also represents shuffled condition in which the order of gazes is shuffled randomly and it arranges a baseline for comparison with actual recurrence dispersion. It should be regarded that when the order of gazes are changed the period of gazes remained same. This graph (Figure 17) also demonstrates if director and matchers' gaze are balanced or dominated by one of them. Based on Nüssli (2011) the schematic explanation of cross-recurrence quantification is shown below in Figure 18.

The rate of recurrences produced for diagonals around the main diagonal and the amount of them entered in to the diagram and it represent the recurrence value regarding the delay between two flows (Nüssli, 2011).

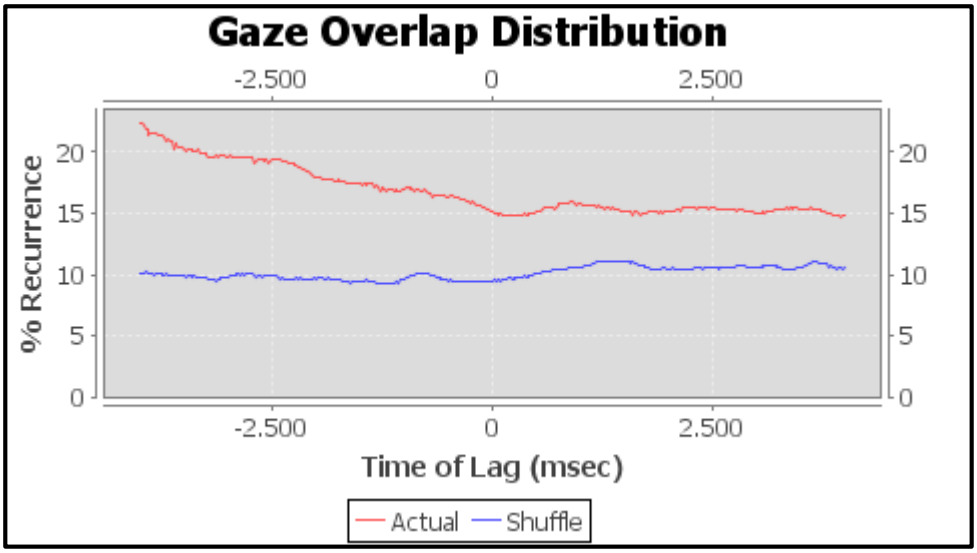


Figure 17 Gaze Overlap Distribution

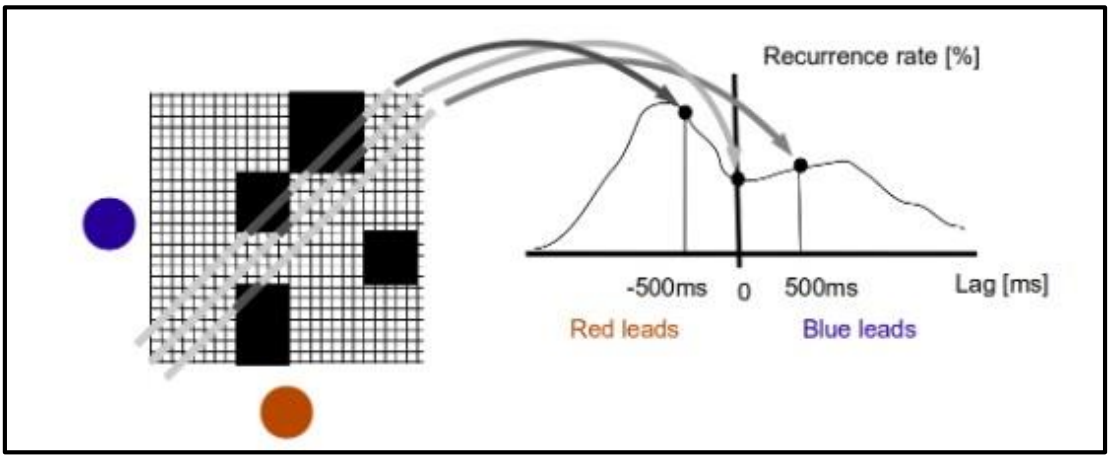


Figure 18 Interpretation of cross recurrence analysis (Nussli, 2011, p. 86)

The rate of recurrences produced for diagonals around the main diagonal and the amount of them entered in to the diagram and it represent the recurrence value regarding the delay between two flows (Nüssli, 2011).

3.4.2. Linguistic analysis

In order to scrutinize over dispersion of REs features in Farsi, their length distribution and probable reduction of turn takings under different circumstances like using color and observing gaze of the counterparts, the underneath method was used. A total of 30 conversations were transcribed and annotated as unveiled below to clarify how conversation or to some extent how referring expressions affect attention.

.

3.4.2.1. Transcription

Along solving puzzles, for each pair, six Farsi dialogues are recorded in which each talk corresponded to each Tangram trial. The length of each dialogue is at the most eight minutes, but it could take less than that based on participants' speed of finding a solution. For five tests, pairs produced 30 dialogues which are transcribed by two native speakers in word documents. Transcription is done simultaneously by listening to audio files and following movies which were captured from screen recordings during games. Turns taken during conversations are quoted by their names or first letter of their names. Figure 19 indicates a screenshot of the transcription environment. The aim is to be able to evaluate the rate of usage of Referring Expressions features in an annotation phase. So, Farsi words are written with Latin alphabet which is nominated Fingilish, in order to create general readability Table 2

Table 2 Finglish examples

Fingilish	Persian	English
Mosallas	مثلث	Triangle
In	این	This

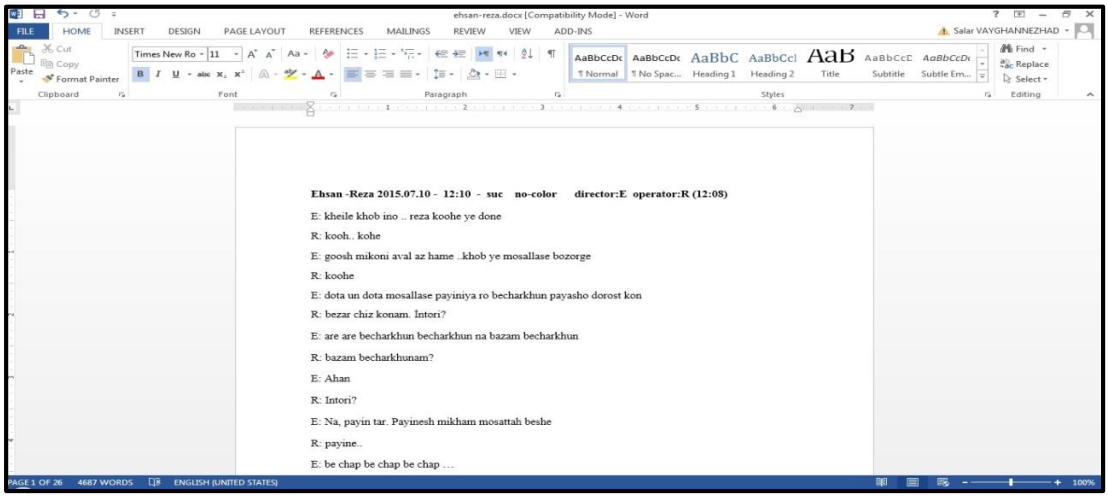
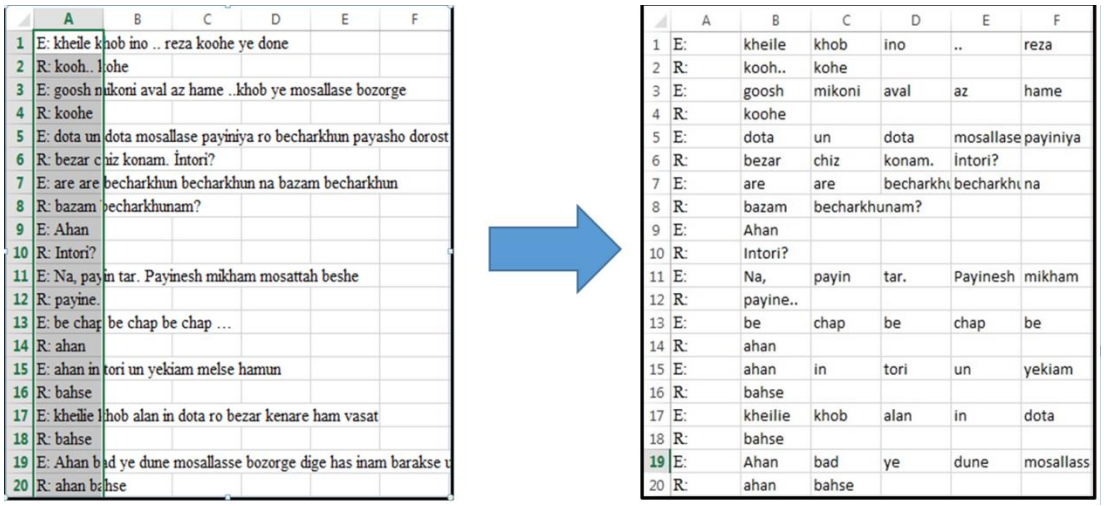


Figure 19 Dialogue **t**ranscription **e**nvironment

3.4.2.2. Annotation of Referring Expressions

Transcribed dialogues are entered in the Excel 2013 environment to split the longer phrases/utterances into words. Below the screenshots from original text and text split are shown in **Figure** 20



	A	B	C	D	E	F
1	E: kheile k	hob ino .. reza koohe ye done				
2	R: kooh.. l	ohe				
3	E: goosh n	ikoni aval az hame ..khob ye mosallase bozorge				
4	R: koohe					
5	E: dota un	dota mosallase payiniya ro becharkhun payasho dorost				
6	R: bezar c	hiz konam. İntori?				
7	E: are are	becharkhun becharkhun na bazam becharkhun				
8	R: bazam	becharkhunam?				
9	E: Ahan					
10	R: Intori?					
11	E: Na, pay	in tar. Payinesh mikham mosattah beshe				
12	R: payine.					
13	E: be chap	be chap be chap …				
14	R: ahan					
15	E: ahan in	tori un yekiam melse hamun				
16	R: bahse					
17	E: kheilie l	hob alan in dota ro bezar kenare ham vasat				
18	R: bahse					
19	E: Ahan b	ad ye dune mosallasse bozorge dige has inam barakse u				
20	R: ahan ba	hse				

	A	B	C	D	E	F
1	E:	kheile	khob	ino	..	reza
2	R:	kooh..	kohe			
3	E:	goosh	mikoni	aval	az	hame
4	R:	koohe				
5	E:	dota	un	dota	mosallase	payiniya
6	R:	bezar	chiz	konam.	İntori?	
7	E:	are	are	becharkh	becharkh	na
8	R:	bazam	becharkhunam?			
9	E:	Ahan				
10	R:	Intori?				
11	E:	Na,	payin	tar.	Payinesh	mikham
12	R:	payine..				
13	E:	be	chap	be	chap	be
14	R:	ahan				
15	E:	ahan	in	tori	un	yekiam
16	R:	bahse				
17	E:	kheilie	khob	alan	in	dota
18	R:	bahse				
19	E:	Ahan	bad	ye	dune	mosallass
20	R:	ahan	bahse			

Figure 20 split of original text into words in new columns

As the main goal is harvesting referring expressions, their length and features dispersion over noun phrases, the work is done in two stages. In first stage, referring expressions are colored based on their length in the first sheet; in other words, it is referring expressions' recognition phase, then along second stage selected referring expressions are moved to another sheet and categorized in different feature groups by defining different colors for each specification, this phase is about identifying the expression's referent category. Below screenshots for both stages are shown in Figure 21 and Figure 22, for first stage it should be considered so:

2Word /Cell=38/ RE=19 means the number of 2word referring expression is 19, which is composed of 38 words or cells.

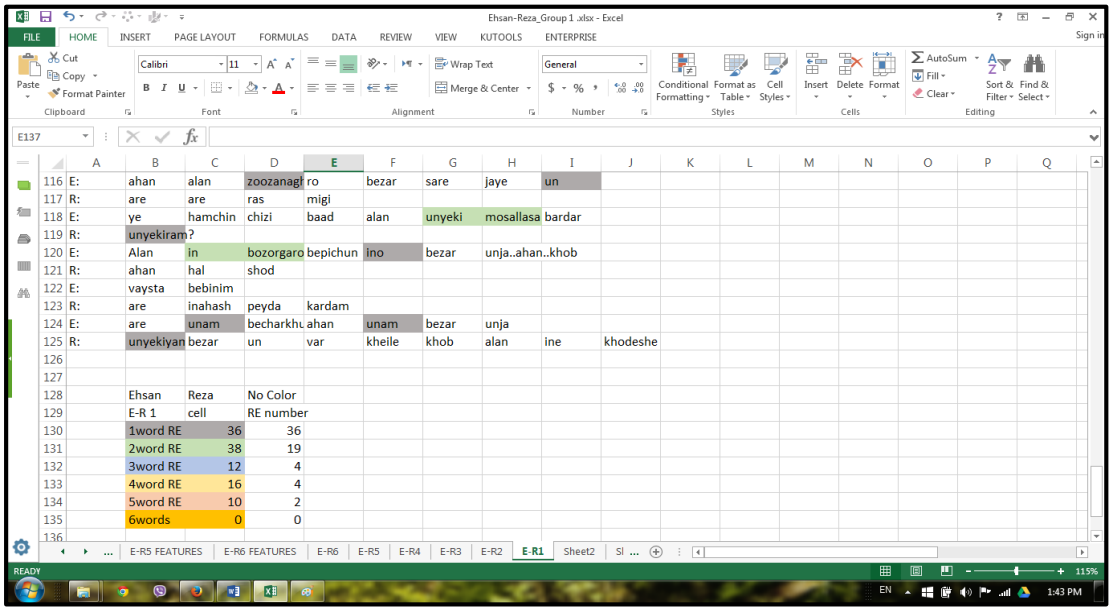


Figure 21 Screenshot of identification of referring expressions and categorizing them based on their length

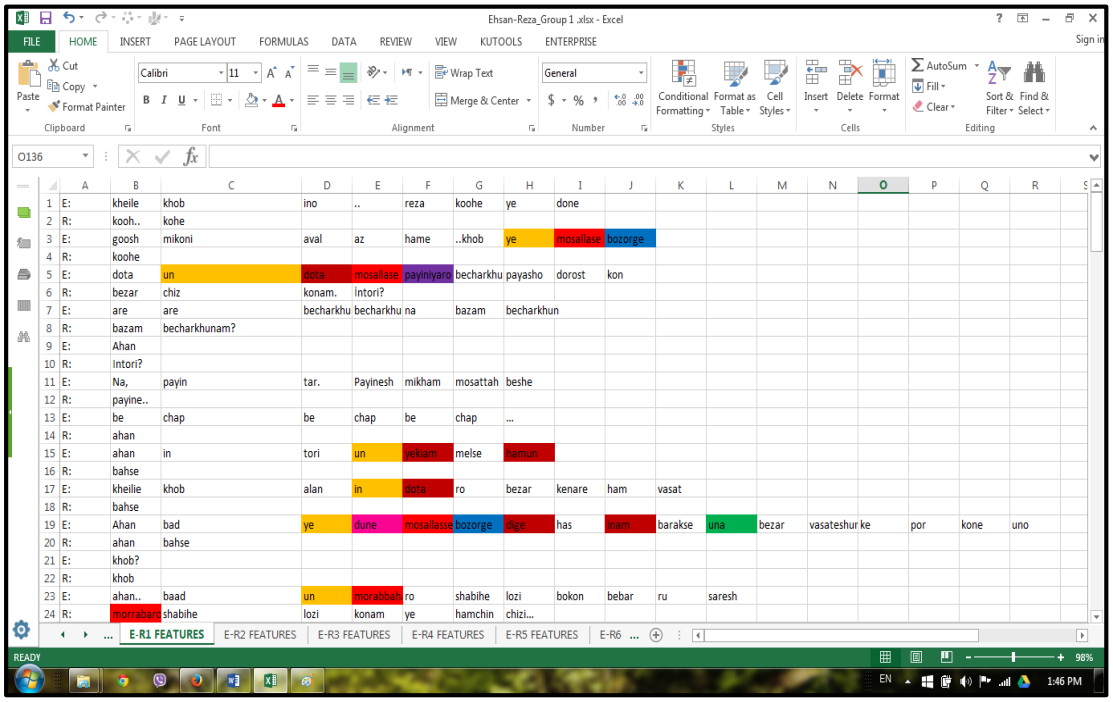


Figure 22 Screenshot of categorization of Referring expressions with different colors

Colors are defined in this way for categorizing RE's features and counting them **Figure** 23:

Figure 23 RE's Categorization color definition

It should be regarded that, for counting cells by color of their feature, which is required for scrutinizing over data, Kutools add-in software is preferred which is a powerful Excel toolbox and enhances Excel functions. In this respect, used path is:

Kutools>Functions>Statistical & Math> COUNTBYCELLCOLOR

Then the address area and color defined as bellow and it produces the numbers of cells in same color.

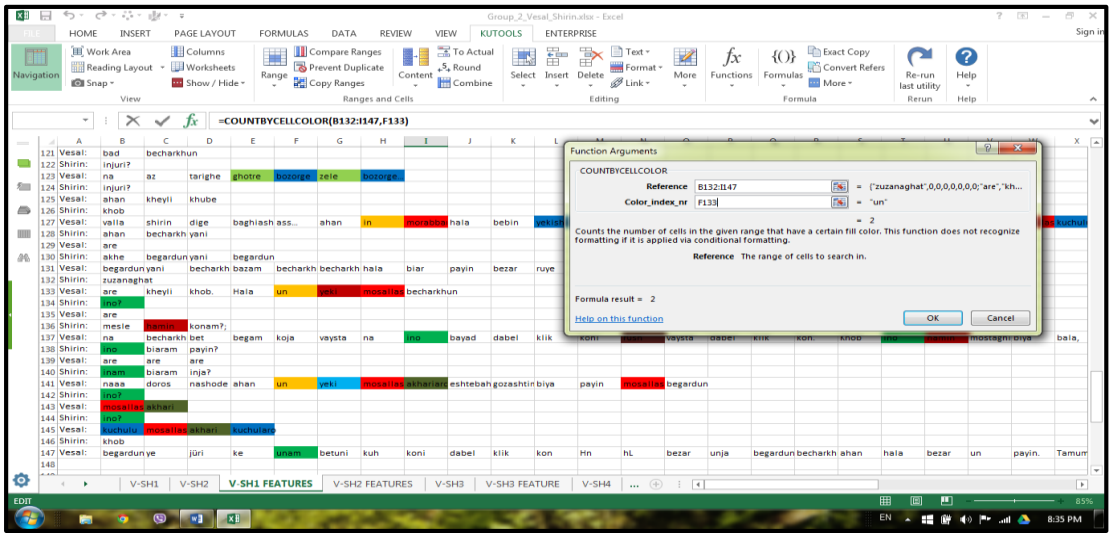


Figure 24 Process of counting referring expressions' features via Kutools

After counting the cells by their correspondent color the percentages of RE's features are computed to be able to compare with Turkish, English and Japanese version and have a cross-linguistic analysis. For finding the similarities between languages in using RE features in their languages Euclidean distance is used it will be explained in Calculation methods further. Meantime, the most frequent words in REs are reported in a table based on counting the words manually.

3.4.2.3. Annotation Guideline

30 dialogues are annotated based on the following assumptions:

- Noun phrase referring expressions are marked.
- Selected referring expressions' referents stand for definite object or set of objects (puzzle pieces in the Tangram working space) or part of the pieces.
- Referent expressions which lead to the location are disregarded.
- Single-word and multiple-word noun-phrases are considered.
- Metaphoric referents are not marked.
- Negative referring expressions are not studied but correcting ones are observed. For example in cases like this:"na, un mosallas, na. Motavaziolazlae" (No, not that triangle, the parallelogram) that triangle is not marked while the parallelogram is counted.
- For each word in a referring expression just one feature (e.g. color, shape) is defined.
- Ambiguous situations in which it is not clear which object is intended, are not marked.
- If one of the participants used something wrongly but the referent was clear, then it is included in the corpus. For example, when the only blue piece in the working environment is a triangle and they use it square and they both understand that they are talking about triangle. So in this situation, it is counted in the shape group.
- Muttering to self is rejected along annotation.
- If the word is interrupted by another utterance and if it was clear it is included. Such as "un mosal…" (That train…)
- If a word is not covered with any of features it is entered in "other" category.
- Interjections which are not written in Farsi but pronounced are not annotated. (Farsi has interjections such as (a/e/o) they are added to the letters and give them new meanings. Mosallas e bozorg (بزرگ مثلث) (big triangle) e is not written in Farsi but pronounced. These interjections are used in first grades of primary school but afterwards they are used without writing.
- Some words are pronounced different in written (formal) Farsi and conversational one. For example for that we say Un (is conversational) while we write An is written way.

- Same in Farsi has two words same this (hamin) and same that (hamun/ haman) they are used instead of this and that too.
- For length analysis, RE's with length between one and six are included. (There was observation of one seven word and eight word RE in the data and they are categorized in annotation part but they are not entered in word length analysis.

All referring expressions which are used in the way of reaching mutual realization are categorized word by word as described as follows. Examples of the selected features for Farsi referring expressions are tabulated inTable 3, which is designed in the same way as the Turkish version which was produced by Acartürk & Çakır, (2012) and Japanese corpus in comparison with English corpus yielded by Spanger et al. (2009).

Table 3 Examples for features of RE in Farsi

Features	Examples
Demonstrative	
Adjective	*in mosallas* 'this triangle' این مثلث
Pronoun	*in* 'this', *anlun* 'that' این آن(اون)
Nominalized form	*Kuchika* 'small-PLU' کوچک‌ها (کوچیکا)
Partitive	*Yeki az* 'one of …' یکی از
Determinative	*Dige* 'other', *hamin* 'same' for near, *hamun/haman* 'same' for far دیگه، همین، همان (همون)
Pronominal Quantifier	*In chiz* 'this thing' این چیز
Attribute	
Size	*Mosallase bozorg* 'large triangle' مثلث بزرگ
Shape	*Mosallase bozorg* 'large triangle' مثلث بزرگ
Direction	*Un ke be chap nega mikone / ru be chapiye* 'the one facing to left' اون که به چپ نگا می‌کنه / رو به چپیه
Spatial relation	
Projective	*Rastiye* 'the one on the right' راستیه
Topological	*biruniye* 'the one outside' بیرونیه
Overlapping	*Ruyish/ ruiyeh* 'the one on the top' رویش/روییه
Action mentioning	*Un ke charkhundi* 'the one you turned' اون که چرخوندیش
Time adverbial	*Vapasin/ akhariye* 'the one a moment ago' واپسین / آخریه
other	as 'of' -dune 'piece' از- دانه/ دونه

Another native speaker is trained based on these assumptions to classify the referring expressions of two of the transcribed dialogues, independently, in order to evaluate the reliability of annotation and consider rate of disagreement over the issue. For this reason formula below used and it produced 88.97 of reliability for one of the dialogues and 90.09 for another one.

$$\frac{\boldsymbol{Agreement}}{\boldsymbol{disagreement + agreement}} \times \mathbf{100}$$

3.4.2.4. Functional Study of ‘In’ (This) and ‘An/ Un’ (That)

To have a short glance over the use of This and That in Farsi Ten first pronoun (In(this) / An/Un (that)) from Normal trials are selected and evaluated to see if they are used by director or operator. And what is the feature of the word after them.

3.4.3. Statistical Methods

For evaluating the reliability of annotation “Interrater Reliability” is performed which compares the annotation of writer with another native speaker’s annotation for specific parts of transcription. In order to indicate the dramatic differentiations over experimental conditions, repeated measure ANOVA is used. Besides, one-way ANOVA conducted to bring out mean gaze recurrence percentages for each group. For correcting the degree of freedom of distribution, Greenhouse- Geisser is conducted. Meantime, Due to finding the similarities between languages over the use of referring expressions category distribution, Euclidean distance is applied. Also for validating repeated measure ANOVA, Mauchly’s Sphericity Test was applied over eye gaze data. Finally, Pairwise LSD comparison method is conducted to find the differentiation among pairs’ eye gaze performance.

CHAPTER 4

RESULTS

This chapter consists of two main sections. In the first section, the results related to dual-eye tracking data analysis including the scarf plots for AOIs and gaze recurrence distributions are presented. The second section reports the results obtained from our annotated corpus of Farsi referring expressions, including the percentage distribution of referring expression features and their comparison with other languages such as Turkish, English and Japanese. Next, our findings on the length distribution of Farsi referring expressions and the functional use of *in* 'this' and *an/un* 'that' in Farsi are described. Finally, the most frequent words among referring expressions, used in dialogues are stated.

4.1. Dual Eye-Tracking Analysis Results

Gaze scarf plots and gaze overlap distribution diagrams were produced for each tangram trial to summarize which AOI each participant was looking at and the overlaps between gaze sequences of both participants. For example, in Figure 25 and Figure 26, the gaze overlap analysis results of the most successful group during a trial in the Color condition are presented. The first graph indicates 71 seconds of gaze overlap (i.e. total duration in which both participants were looking at the same AOI). The red line in the second picture demonstrates the distribution of gaze overlap at different time lags (i.e. ±4 seconds). The blue line in the second graph shows the percent of gaze overlap when one of the participants' gaze sequence was randomly re-ordered, which can be used as a baseline to contrast against the red line. The red line deviates from the blue line especially for lags of ±2 seconds, which suggest that both peers' gaze sequences overlapped within a lag of 2 seconds. Since the graph is symmetric with respect to y=0, both peers followed each other in a balanced way, suggesting a high level of mutual gaze coordination between peers. All results for other groups are attached in Appendix B: GAZE ANALYSIS OUTPUTS.

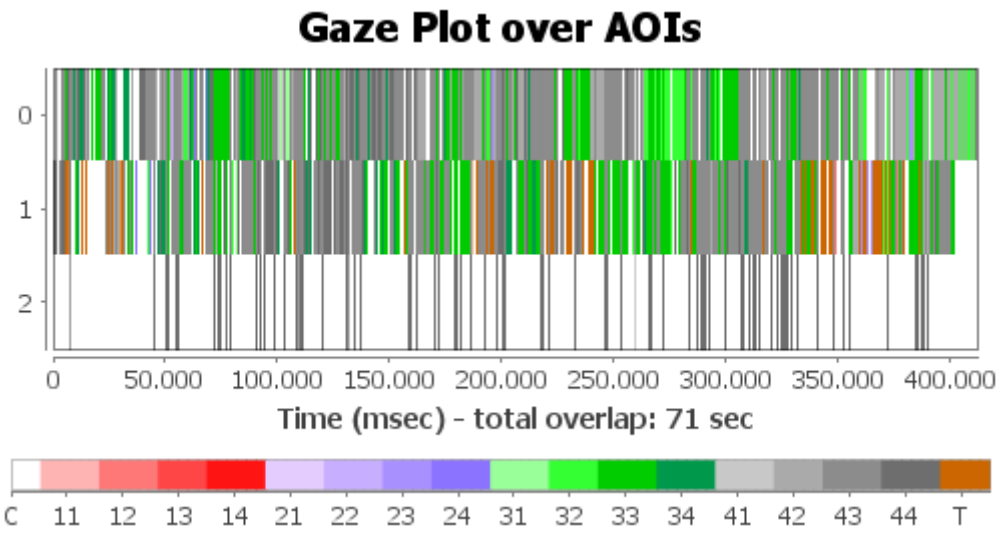


Figure 25 Gaze Plot over AOIs for the best gaze match

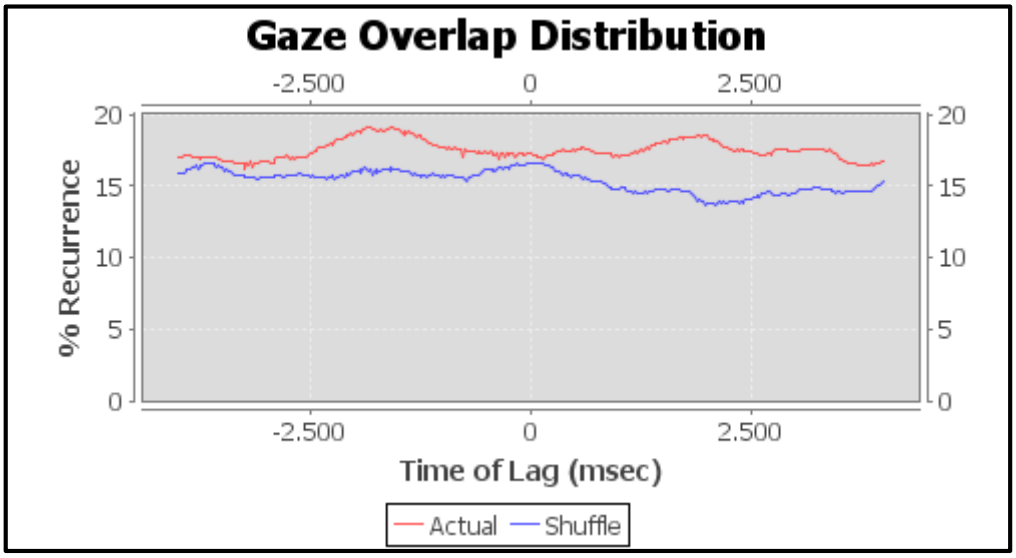


Figure 26 Gaze Overlap Distribution for the best gaze match

4.2. Gaze Recurrence Analysis

A 3x2 repeated measures ANOVA was conducted on percent gaze recurrence levels obtained for each tangram trial where condition (normal, color and gaze cueing) and trial (first and second) were within subjects independent variables. Gaze recurrence levels are computed as described in Dale & Richardson (2005), where overlaps in gaze locations of two participants within a duration of plus/minus 2 seconds were defined as recurrent cases (i.e. if participant A's gaze falls on to where participant B's gaze has fallen within plus/minus 2 seconds were assumed as instances of coordinated gaze activity).

Mauchly's test indicated that the assumption of sphericity was tenable ($W(2)=.49$, $p>.05$ for condition and $W(2)=.624$, $p>.05$ for the interaction term), so no sphericity correction was made. The repeated measures ANOVA test indicated that there was no significant effect due to condition ($F(2,8)=.35$, $p>.05$) or trial ($F(1,4)=3.57$, $p>.05$), but a significant interaction effect was found between condition and trial, $F(2,8)=9.76$, $p<.01$, $\eta^2=.71$. The line graph shown in Figure 27 below suggests that the trend in gaze percentage change across conditions differ for the color and the other two conditions. In the color condition, the gaze recurrence percentage tends to increase when participants did their second trial in the same condition. This may suggest that the availability of the colored tangram pieces elicits a larger increase in gaze recurrence as subjects get accustomed to using color terms to refer to individual pieces. Moreover, the differences between the first and last trials were smaller in the gaze cueing condition, as compared to the normal condition.

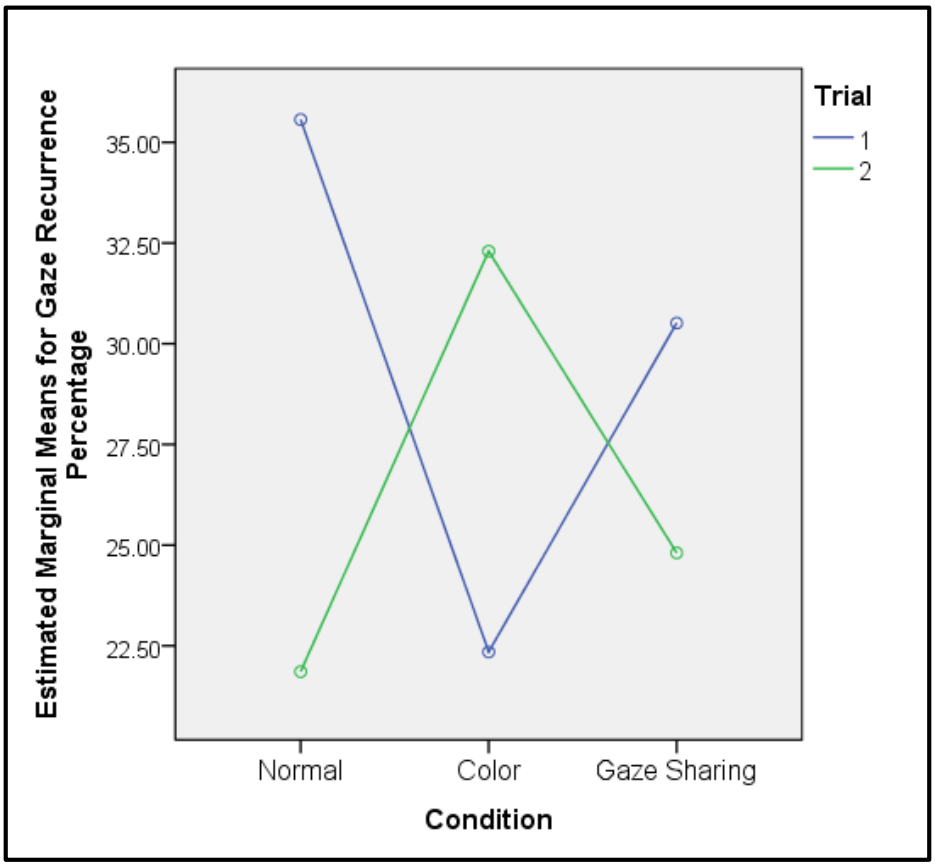


Figure 27 Gaze percentage across conditions regarding trials

In addition to this, a one-way ANOVA conducted on mean gaze recurrence percentages for each group found a significant difference among the 5 pairs who participated in the experiment, $F(4,30)=7.61$, $p<.01$, $\eta^2=.55$. The bar-chart below shows the average gaze recurrence percentage of each pair in all of their 6 trials. Pairwise LSD comparisons listed in Table 4 found a significant difference between pair 5 and all other pairs, as well as pairs 1 & 3 and 2 & 3. This result seems to be related to the performance of each group during the experiment, where the first group had an accuracy of 100%, whereas the second and third groups had accuracy percentages of 66.67%, and finally the remaining two groups had an accuracy of 33.33%. Although we had a small sample of pairs, this suggests that pairs that performed well together tend to have a higher degree of gaze recurrence. No significant differences were observed for the group's gaze recurrence percentages across conditions and trials, so the pattern of relationship among the mean gaze percentages observed for each group was consistent among levels of condition and trial.

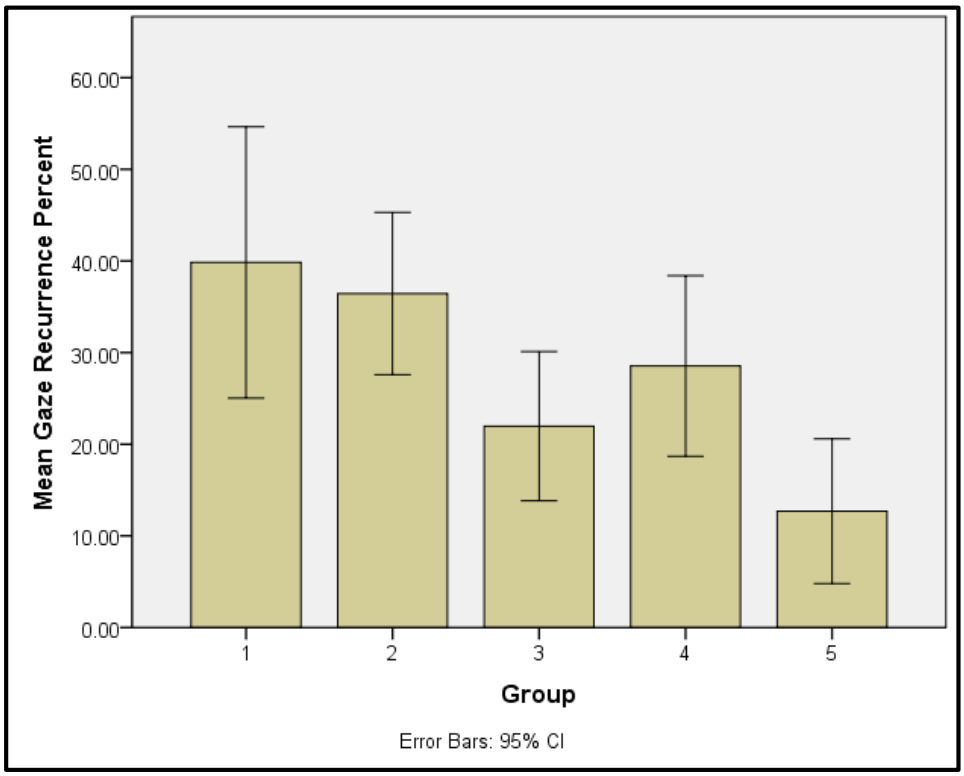


Figure 28 Mean gaze recurrence percentage for whole groups

Table 4 Pairwise LSD comparisons

Multiple Comparisons

Dependent Variable: Gaze Recurrence Percent

LSD

(I) Group	(J) Group	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
1	2	3.3925	5.62913	.552	-8.2009	14.9859
	3	17.8591*	5.62913	.004	6.2657	29.4525
	4	11.2891	5.62913	.056	-.3044	22.8825
	5	27.1463*	5.62913	.000	15.5529	38.7397
2	1	-3.3925	5.62913	.552	-14.9859	8.2009
	3	14.4666*	5.62913	.017	2.8732	26.0600
	4	7.8965	5.62913	.173	-3.6969	19.4900
	5	23.7538*	5.62913	.000	12.1604	35.3472
3	1	-17.8591*	5.62913	.004	-29.4525	-6.2657
	2	-14.4666*	5.62913	.017	-26.0600	-2.8732
	4	-6.5701	5.62913	.254	-18.1635	5.0234
	5	9.2872	5.62913	.111	-2.3062	20.8806
4	1	-11.2891	5.62913	.056	-22.8825	.3044
	2	-7.8965	5.62913	.173	-19.4900	3.6969
	3	6.5701	5.62913	.254	-5.0234	18.1635
	5	15.8573*	5.62913	.009	4.2639	27.4507
5	1	-27.1463*	5.62913	.000	-38.7397	-15.5529
	2	-23.7538*	5.62913	.000	-35.3472	-12.1604
	3	-9.2872	5.62913	.111	-20.8806	2.3062
	4	-15.8573*	5.62913	.009	-27.4507	-4.2639

Based on observed means.

The error term is Mean Square (Error) = 95.061.

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

Finally, we tested if the gaze recurrence percentage is affected during the course of the entire experiment. As the participants attempt new puzzles repeatedly one might expect that they would develop a common referential framework that may help them achieve higher levels of referential understanding and hence higher levels of gaze coordination. A one-way ANOVA test suggested that there was no significant difference among gaze coordination levels across the 6 puzzles attempted by each pair, $F(5,30)=0.28$, $p>.05$.

4.3. Linguistic Analysis Results

4.3.1. The most frequently occurring words in referring expressions in the Farsi corpus

Table 5 lists the most frequently used words among REs in the Farsi corpus in descending order. “Mosallas” (Triangle) with 14.91% appeared as the most frequent word, followed by “That” with 11.47% and “This” with 11.38%.

Table 5 The most frequent words in referring expression of Farsi corpus

Order	Farsi	Meaning (English)	Percentage
1	mosallas	Triangle	%14,91
2	un	That	%11,47
3	in	This	%11,38
4	yeki	One of	%6,38
5	bozorg	Big	%6,034
6	uno	That one	%5,26
7	morabba	Square	%5,17
8	ino	This one	%5,09
9	kochik	Small	%3,97
10	motavaziolazla	Parallelogram	%3,362
11	hamun	Same (far)	%1,81
12	ye	One	%1,21
13	motavasset	Average	%1,03
14	hamin	Same (near)	%1,03
15	payinio	The one at the Bottom	%1,03
16	chapio	The one on the Left	%0,86
17	rastio	The one on the right	%0,69
18	balayio	The on the Top	%0,69

4.3.2. The Influence of Color and Gaze Cues on the Length of Farsi Referring Expressions

The length of the referring expressions used by the participants is reported to change as the participants establish a common ground for the relevant objects for their joint task (Clark and Wilkes-Gibbs, 1986). Longer referring expressions provide more contextual details for the listener to disambiguate the intended co-referent. As participants establish a common ground for the objects in the shared scene, the referring expressions they use may be expected to get smaller as well. Therefore, the need for longer expressions and elaborations can be considered as indicators of referential difficulty.

We compared the three conditions in terms of the average length of the referring expressions used by the participants, to see if they have any influence in the way they help interlocutors reduce the effort for establishing a common ground (Figure 29). During the experiments participants completed two trials in each condition by switching roles. Since getting accustomed to the new cues present or not present in each condition may take a while, we also included trial number in the analysis to test the effect of this temporal factor.

Figure 30 below shows the percentage of referring expressions of length 1 to 6 in the Farsi corpus observed in each condition. In our corpus the longest referring expression we observed included 6 tokens.

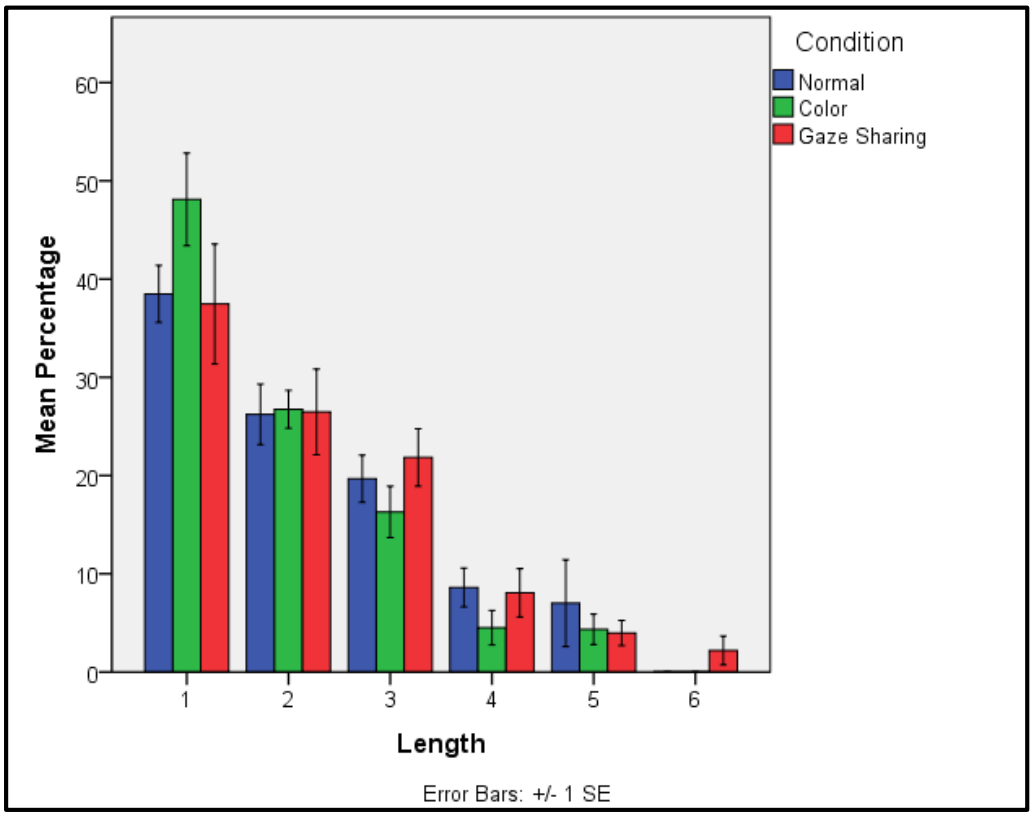


Figure 29 Percentage of length distribution over referring expressions

The bar chart suggests that referring expressions of length 1, 2 and 3 are the most frequently occurring expressions. The color condition has the largest percentage of single word referring expressions, which seems to be due to color coding that allowed participants to uniquely refer to each piece with a color term.

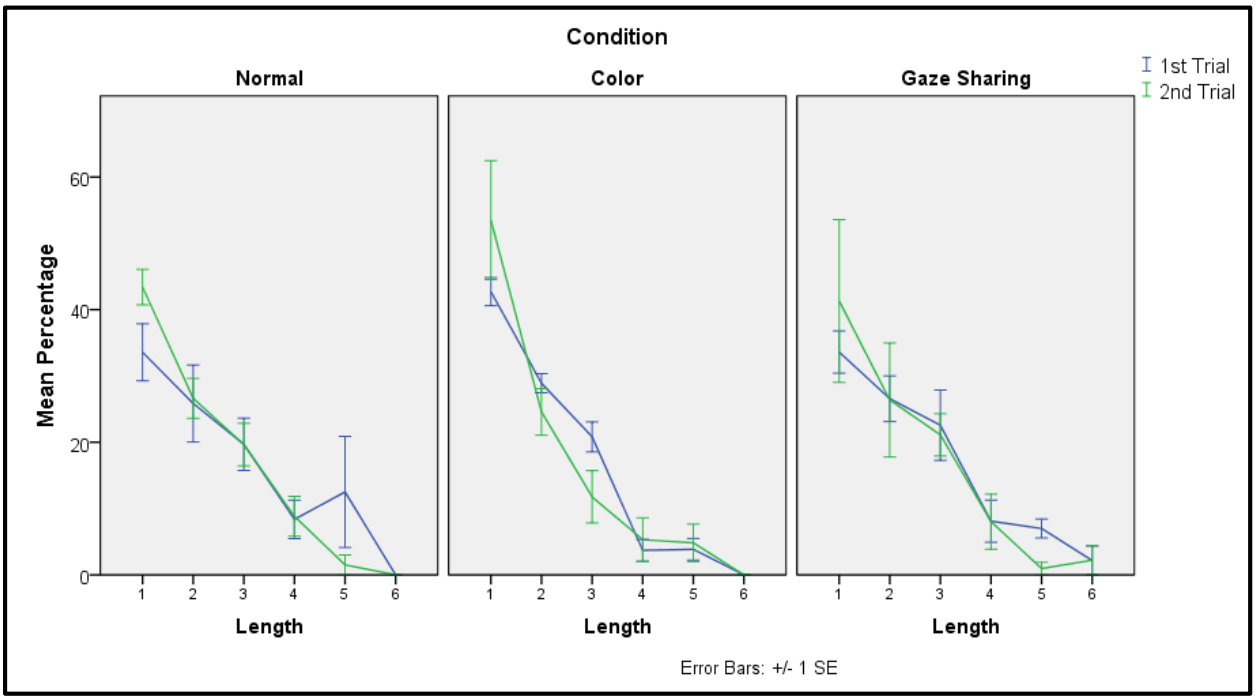


Figure 30 Repeated measure ANOVA for length distribution

Repeated measures ANOVA results summarized above showed that there is an interaction between Condition and Trial, i.e. there is a stronger switch to shorter expressions in the color condition as compared to the other two conditions.

4.3.3. The Influence of Color Cues and Gaze Cueing on Categorical Distribution of Farsi Referring Expressions

The bar chart in Figure 31 below shows the cumulative sum percentage of each referring expression category over each experimental condition in the Farsi corpus.

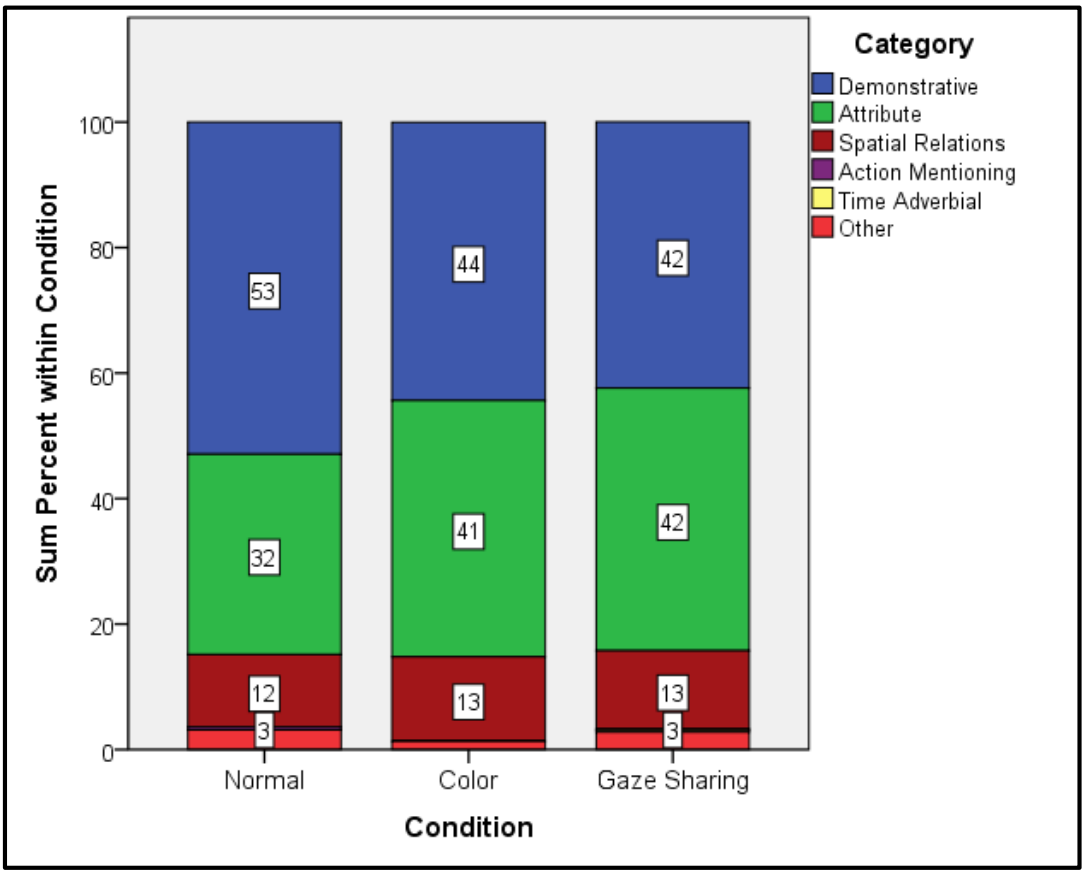


Figure 31 Accumulative distribution of referring expressions categories

The bar chart suggests that Demonstratives are more frequently used in the normal puzzle condition. When either color-coded puzzle pieces or gaze cueings are enabled, participants tend to use more Attribute type referring expressions. Referring expressions annotated as Spatial Relations have roughly the same percentage across all three conditions.

Demonstratives, attributes and spatial relations were further analyzed in terms of more specialized categories. The bar chart in Figure 32 shows the cumulative percentage of more-fine grained referring expression categories.

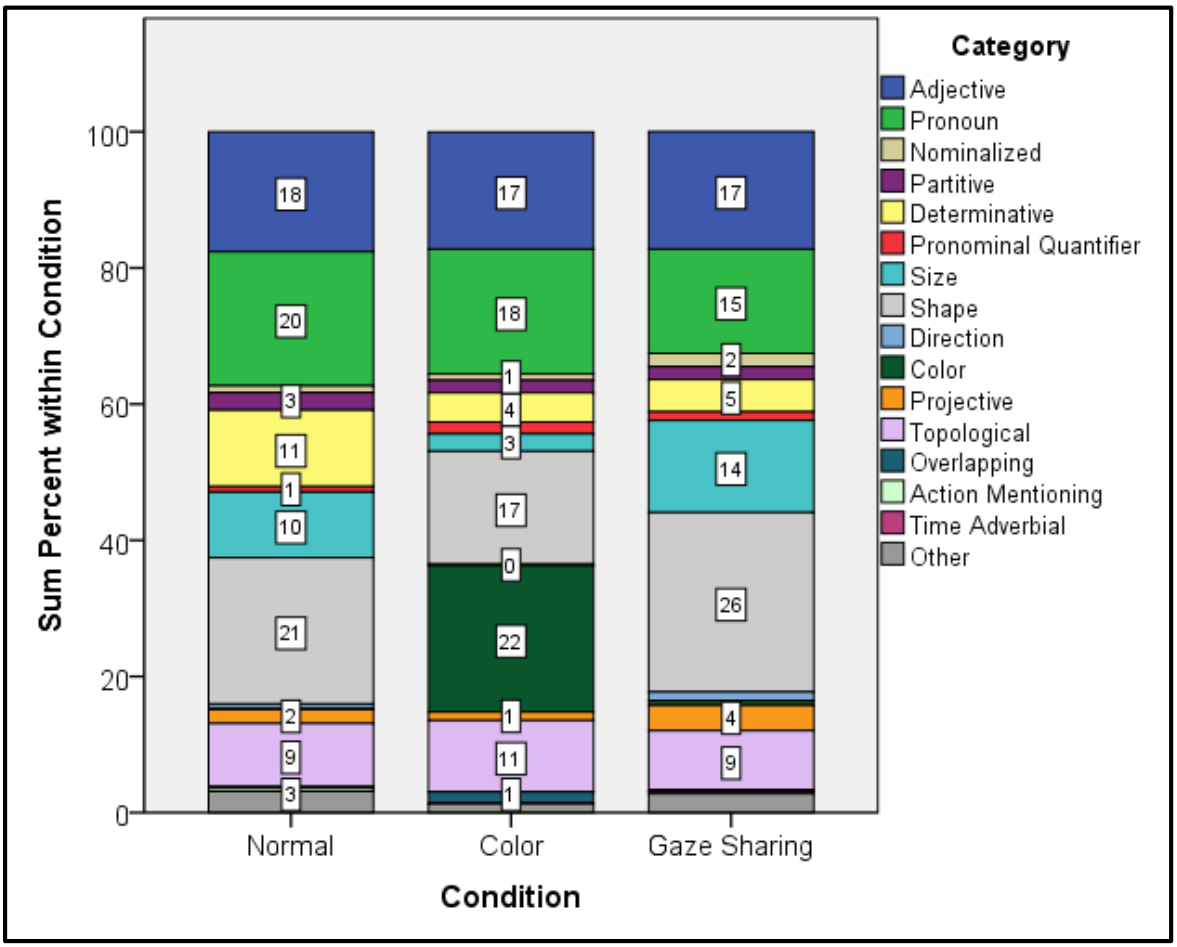


Figure 32 Distribution of referring expressions' categories in detail

Figure 32 shows the percentage of referring expression categories taken over the entire condition, without discriminating among groups. The most dramatic difference among the three conditions is the obvious case of color references which are enabled in the Color condition, which take about 22% of all the referring expressions in this condition. There is also a decrease in the percentage of shape, size and determinative references in the color condition as compared to other two conditions, possibly due to the availability of color as a cue. However, the percentages for adjectives, pronouns and topological references remain unaffected by the experimental conditions. Finally, the use of determinatives tend to decrease when either color or gaze cues are made available to the participants.

In order to test the statistical significance of these patterns, we computed the same percentages for each category within each group, and ran separate repeated measures ANOVAs to observe if the patterns are influenced by variability due to different pairs of speakers.

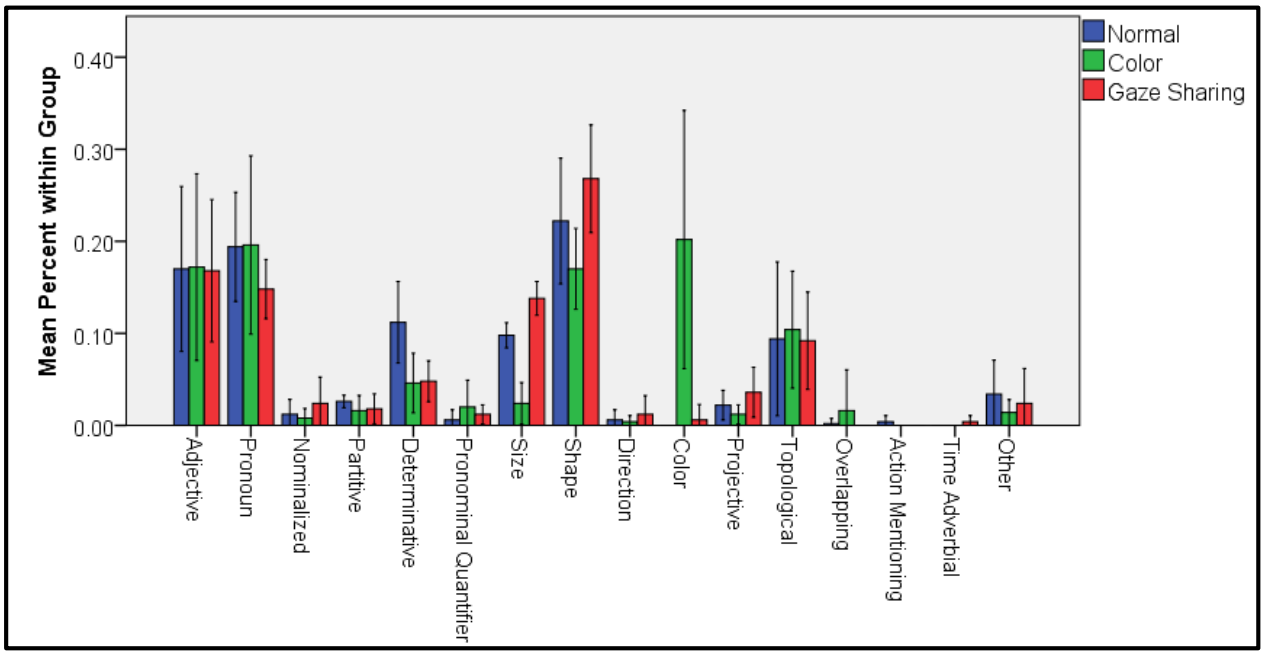


Figure 33 Mean percentage of each referring expression category observed in five teams across 3 conditions. Error bars indicate %95 confidence intervals.

Figure 33 shows the mean percentage of each referring expression type used by each pair during three different situated dialog conditions. Separate repeated measures ANOVAs conducted over each category found significant difference among experimental conditions for the Determinative ($F(2,8)=23.75$, $p<.01$, $\eta 2=.86$), Size ($F(2,8)=116.70$, $p<.001$, $\eta 2=.96$), Shape ($F(2,8)=5.30$, $p<.05$, $\eta 2=.57$) and Color ($F(1.002, 4.007)=13.77$, $p<.05$, $\eta 2=.78$ Greenhouse-Geisser corrected) categories.

One of the dramatic changes obtained here is the reduction of Determinatives in both color and gaze conditions it reveals that participants rationally use color or gaze factors to reduce uncertainties and comprehension effort due to identifying objects; for instance, they use color or they look at the pieces instead of using terms such as same one -one of-the other one. Another striking difference is about Shape and Size decrement in Color condition and increment in Gaze condition. It can be deducted that contributors marked the objects with colors as the dominant factor or to some extent highly recognizable property when there is just one piece with same color, and they do not find it necessary to give extra information by uttering shape or size even if they are used as null-referring in somehow. This research highlighted the outstanding role of color in object identification which can also influence the rate of usage of attributes such as shape and size which are known to be constant whenever other features evaporate (Braje et al., 1999).

Considering Gaze case in which the rate of use of shape and size had a significance upward trend, it can be understood that regardless of the outstanding effect of color in comprehension, Shape and size are respectively important factors for realization alignments, particularly shape became a most prevailing factor in Gaze situation as it was in Normal condition. In comparison with normal condition, beside the growth in use of shape and size terms there is also a slight decrease in usage of pronoun which appeared to be substituted by shape and color's addition. In Gaze cue state, despite the existence of eye gaze cue, participants still rely on and also develop the amount of uttering shape and size characteristics to increase collaboration.

4.4. Cross-linguistic Analysis of Referring Expressions

We annotated single word referring expressions and each word in multi-word referring expressions according to their syntactic and semantic features similar to studies by Takenobu et al.(2010), Acarturk and Cakir (2012). In Table 6 the results obtained for Persian from a corpus of 5 situated dialogs are contrasted to the results reported for the Turkish (Acarturk & Cakir, 2012), Japanese and English (Takenobu et al., 2010) corpora that use the same tangram based situated dialog task. In order to ensure compatibility of the results, only annotations for the normal condition (e.g. no color/gaze cues) were used for the Persian data.

Table 6 Cross-linguistic Comparison

Attribute	Persian	Turkish	Japanese	English
Demonstrative				
Adjective	17.6	10.5	8.2	6.0
Pronoun	19.7	26.0	31.2	46.5
Nominalized	1.0	0.09	NA	NA
Partitive	2.6	1.56	NA	NA
Determinative	11.2	4.69	1.5	3.5
Pronominal	0.8	1.02	NA	NA
Attribute				
Size	9.6	14.4	13.2	14.0
Shape	21.5	34.4	30.1	21.2
Direction	0.7	0.04	0.3	0.0
Spatial				
Projective	2.0	2.65	6.5	2.1
Topological	9.3	0.09	0.4	0.8
Overlapping	0.3	0	0.1	0.0
Action mentioning	0.4	2.04	4.7	1.0
Time	0.0	0.04	NA	NA
Other	3.1	3.5	3.7	4.9

Table 6 compares the distribution of each referring expression attribute across four different languages. A dissimilarity measure based on Euclidean distance is computed over these distributions to observe the relationships among these languages in terms of the syntactic/semantic features presented in their respective referring expression corpora. Table 7 shows the obtained dissimilarity matrix, where smaller distance values suggest similarity.

Table 7 Euclidean distance between attribute distributions among the four languages

	Persian	Turkish	Japanese	English
Persian	0			
Turkish	20.31	0		
Japanese	22.99	9.38	0	
English	31.88	24.97	18.9	0

The dissimilarity matrix suggests that the distribution of referring expressions seems to be nearer to Turkish language rather than Japanese and English. The next similar language

to Farsi is Japanese and the less similar one seems to be English. Along similar lines, for Turkish, Japanese results are more resemble than the others and For English the most homogenous language reported to be Japanese.

4.5. Functional Study of 'In' (This) and 'An/ Un' (That) for constructing further researches

To functionally observe the use of This and That in Farsi Ten first pronoun (In (this) / An/Un (that)) from Normal trials are selected and assessed to clarify if they are used by director or operator and what is the feature of the word after them.

From Table 8 qualitatively it can be deducted that shapes are the dominant words which are used after 'This' and 'That' in Farsi (in half of situations 10 over 20(5 for 'this' and 5 for 'that'). Besides it shows that mostly 'this' and 'that' are used from directors side (15 over 20(7 for 'in' and 8 for 'an/un)) while for operator it is (5 over 20(3'in' and 2 'an/un')) This analysis can be extendedly done over all the pronouns in further works to clarify This and That pronouns' role in referring to objects.

Table 8 Functional usage of ‘this’ and ‘that’ in Farsi corpus

Functional Analysis of this and that		Normal 1	Function,	Normal 2	Function
group 1	İn (This)	dota (Two of)	Director	yeki	Operator
	un That	mosallas (Triangle)	Director	mosallas (Triangle)	Director
group 2	İn (This)	dota Two of)	Director	mosallas (Triangle)	Operator
	un That	Verb	Operator	yeki	Director
group 3	İn (This)	mosallas (Triangle)	Director	noke	Operator
	un That	yeki (One)	Director	VERB	Director
group 4	İn (This)	mosallas (Triangle)	Director	yeki (One)	Director
	un That	Motavaziolazla (Parallelogram)	Director	Morabba (Square)	Operator
group 5	İn (This)	Morabba (Square)	Director	mosallas (Triangle)	Director
	un That	EOS*	Director	mosallas (Triangle)	Director

Related categorization protocol based on color is demonstrated here:

End of Sentence(EOS)
Determinative
Shape
Topological

4.6. TT (Turn Taking) Results

As the conversation fluctuated between listener and speaker, they take turns for speaking. Based on Clark and Wilkes-Gibbs (1986) toward trials number of turn takings decreases. Hereunder the role of different cues and trials over the number of turn takings is evaluated.

4.6.1. Total Average of TT

The average number of turn takings in normal, color and gaze cue conditions is reported as below (Figure 34). The figure that indicates more detailed results for all groups attached in APPENDIX C.

4.6.2. TT Based on Six Trials

Each group had six trials and the total amount of turn takings were recorded. Figure 35 shows the total average of TT of all groups (regardless of their condition) and the separated charts for all groups in detail.

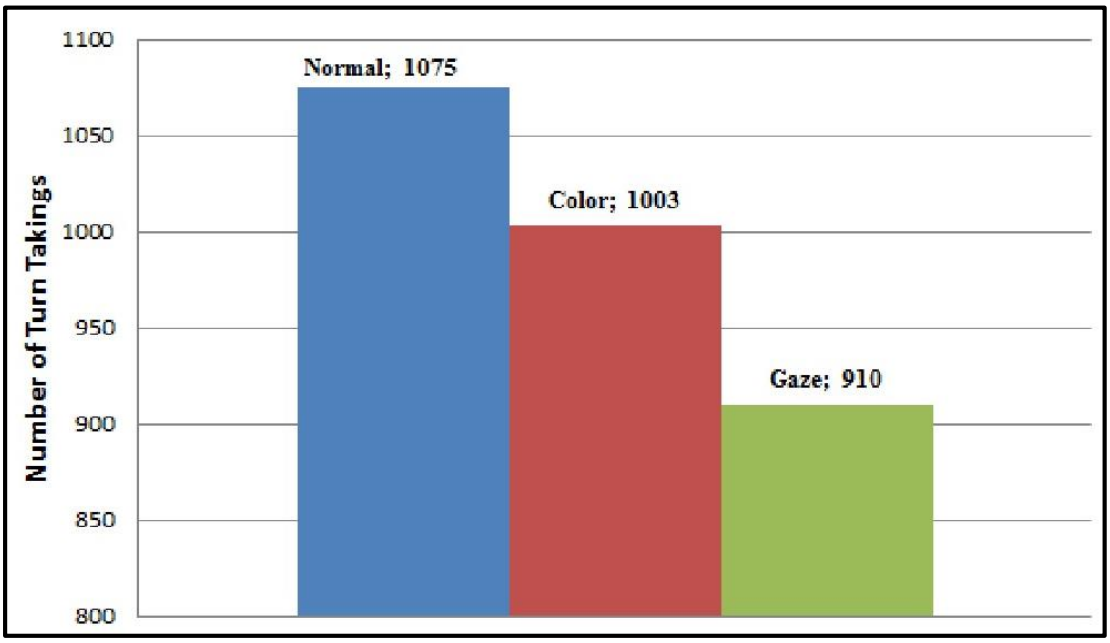


Figure 34 Average of all turn takings.

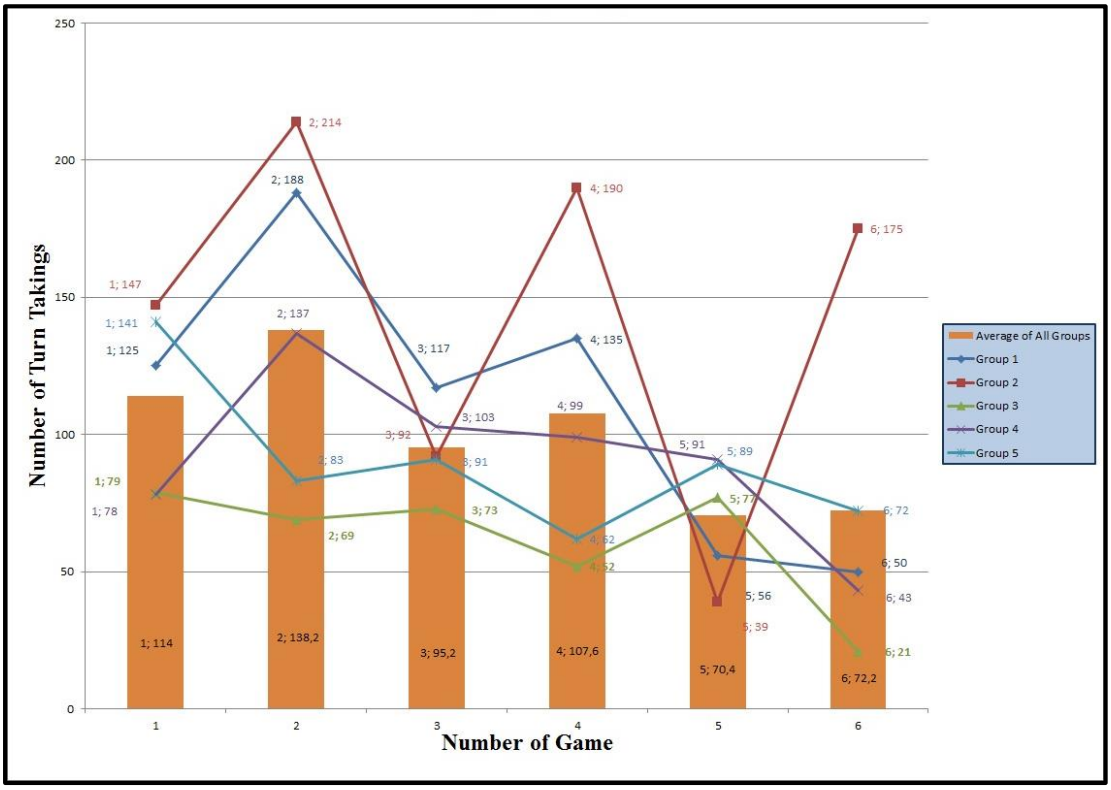


Figure 35 TT regarding to six groups

4.7. Summary of Results

This dissertation constructed the Farsi corpus referring expressions in the flow of situated domain of Tangram problem solving and attempted to converge the data with dual eye tracking data in the purpose of evaluating the increment in shared comprehension. Also, the Farsi corpus referring results are entered into the cross linguistic analysis with Turkish, English and Japanese data to find resemblances.

In the path of reaching joint comprehension while assigning and updating common ground, participants' length of referring expressions in the real conversation are noted. The most frequent referring expressions are composed of one, two or three words. Significantly, there was an upward trend in the rate of usage of one word REs in Color condition which provides firm evidence for the usage of color terms for identifying objects.

Furthermore, categorical assessments revealed the dominance of demonstratives in normal condition which was followed by attributes and spatial relations respectively. The existence of color or gaze cueing improved the quantity of Attributes, while the Spatial Relation terms amount remained constant along different task conditions. A closer look at the final outcomes of categories with more specifications, clarified the most striking appearance of color terms among Attributes in Color condition which grows the support for the previous claim that pairs used color terms for recognizing objects, another underlying argument in this respect is the reduction in amount of shape, size and determinative terms in referring expressions in color condition. Besides, addition of color or Gaze cueing to the normal condition does not influence the rate of adjectives, pronouns and topological referring terms. It is also emphasized that the rate of determinatives are diminished in Gaze cueing condition, in contrast, the quantity of shape and size bounced. These comments convincingly convey that color, shape and size are extremely identifiable factors for objects.

The percentages of categorized referring expressions' results in Farsi corpus compared with other languages and it brought out that Turkish is similar than the others to Farsi in this case. The next place allotted for Japanese and the last one for English.

Eye gaze analysis over the gathered data indicated that despite the significant differences in the usage of referring expressions across experimental conditions, in dual eye tracking analyze conditions did not create dramatic changes during passing through the status. But findings helped to uncover a significant interaction effect between condition and trial. Particularly, gaze percentage change for color condition had an upward trend in second trial which might clarify that, participants use advantageous of color condition when they became acquainted with the colored Tangram statues. In addition, the difference of gaze recurrence percentage between first and second tasks was less than other conditions under existence of Gaze condition. Yielded results fostered debate on, whether there are differentiations over groups' eye performances. Further searches, confirmed significant distinction for two of the groups' efficiency which is also hand in hand with their

successfulness in trials; for example, the most successful pair had the most aligned gaze match. Moreover, mulling over the results refuted the idea that passing through 6 trials, participants reached to the higher gaze match.

CHAPTER 5

DISCUSSION

This chapter discusses the results obtained to answer the main research questions. According to the cross linguistic analysis, although Farsi is distant from other languages in the usage of RE's, the resemblance between Farsi and Turkish language over RE's dispersion is more than English and Japanese languages. This might be due to geographical proximity or the existence of many common words among Farsi and Turkish.

The discourse analysis findings suggest that as participants struggle and negotiate to decrease joint effort by forming and extending a common ground, the length of RE's used tended to decrease as it was claimed by Clark and Wilkes-Gibbs (1986). It is clear that by using longer referring expressions peers tend to convey more details about pieces and their configurations. Along the current research, the frequency of RE's under the scrutiny of using color and gaze cueing is evaluated. Due to the results, it is highlighted that the role of Color, size and shape terms for recognizing objects is prominent. In the color condition there is a significant increase in the frequency of one word referring expressions. That is coupled with an increase in the number of attributes and a decrease in the usage of shape, size and determinative terms. These results support the claim that participants tend to use color terms instead of longer expression to refer to the pieces. For example, participants tend to produce 'rotate blue' instead of 'rotate one of those big triangles'. Besides, in the gaze cueing condition it is observed that the amount of shape and size references increased, whereas similar to the color condition the determinatives were less frequently used. Probing over outcomes, after color, shape and size it is worth to talk about adjective and pronouns which are appeared with almost 20% and they remained stable through cases.

Mulling over the details of RE's specifications (which also contain the 6 cognitive status stages of GH) the frequency of Pronoun terms in Farsi referring expressions are found to be less than other languages. Although the amount of usage of 'un' (that) [second position in word frequency table], In (This) [third position in word frequency table], ino (this one) [eighth position] and uno (that one) [sixth position] are prominent, overall the mentioned

words' existence have not created improvement in the rate of pronouns in Farsi in comparison with other languages investigated. Shokouhi (1996) asserted that unlike the English language, which starts with NP and continues with pronominal, Farsi starts with NP and continues with Null referring expressions in a dialogue. This can be one of the reasons underlying the observed differences between Farsi and other languages. Meantime, it should be considered that Farsi does not contain a word like 'It' in English or 'Şu' in Turkish. It puts forward the claim that instead of 'it', most probably, participants tend to use more Determinatives or Adjectives, as both have the upward quantity in Farsi rather than other languages. In Farsi for the term 'same' there are two words including 'hamin'(same this) and hamun/ haman (same that), their usage is very common in discourse, also these words sometimes can be used semantically in the place of 'This' and 'That'. Their usage may be another effective factor of increment in Determinative's frequency in Farsi.. Another striking difference between Farsi and the rest of the languages studied is the frequency of topological terms in Farsi which are enormously more than other languages, it address the issue that in this study the referents for segments of the pieces are also considered while in other studies just the pieces are regarded.

Apparently, Givenness Hierarchy and Centering frameworks restrict the usage of referring expressions in conversation. The constructed Farsi corpus seems to agree and cover GH, which is proposed by Gundel et al. (1993). Although, there is no corresponding word for "it" to put an entity in 'in-focus' cognitive status in Farsi, there are other words used instead like hamin (same this), hamun (same that) and so on as explained above. Considering Grosz et al. (1995), the coherence of dialogue is related to the compatibility between the usage of referring expressions and centering attitudes. As centering frame work cares about local coherence (Yoshida, 2008), it appears to be consonant with the current research because selected referring expressions are anaphoric or deictic expressions and exophoric ones are not included. However, incorporation of centering with global focus for covering anaphoric relations seems to be needed to create links between referent of nouns and pronouns beyond segment boundaries or refer to prior discourse context.

It should be regarded that the gaze alignments while participants passed from one condition to another did not show a dramatic change in the way we hypothesized. Likewise, gaze matching did not increase along the 6 trials. However, there was a prominent interplay between condition and trial. The significant increase in gaze-alignment in the second trials particularly in the Color condition suggests that the degree of gaze overlap increases as partners become accustomed to referring the color-coded pieces in the environment with color references. This observation does not conflict with the claim of Dale et al. (2011) that the synchronization of eye movements improves through trials, since we also observed an increase in gaze coordination when we compared the first and second trials of all conditions. Besides, second trials' adjacency express that experiencing new condition in game firstly develops dangling eye movements and uncertainty for the context; although, the existence of cues were helpful in improving their shared comprehension where as it was not beneficiary for Gaze match. Another

distinguished factor which is worth mentioning is about the success of pairs and the degree of alignment between their eye movements. The percentage of eye gaze adjacency of successful and unsuccessful pairs, regarding the percentages of hit in games, raises the possibility of interaction between conquest and gaze overlap (The pair which reached to the peak of 100% of success and solved all the puzzles had the highest gaze overlap with 40%) this agrees with the prior observations of Richardson and Dale, (2005) about the existence of relation between level of success in two person's collaborative environments and their eye movements coordination . The significance difference among groups' results for their gaze coupling, provides probable evidence that interpersonal abilities might affect successfulness of pairs in trials. As Cole, (1978) and Clark and Schaefer, (1989) claimed people enter in context with their background and form the common ground and update it.

In Tangram puzzle, different factors might affect participants performance, the usage of glass seems to be a negative factor as it can be seen for third group in which also they were capable in solving puzzles by 66.65 percent but their gaze overlap was weak almost 20% (for this group one of the participants was using glasses who felt comfortable without glasses but referring to recorded data, calibration deteriorated many times for this person). Meantime, geometrical and graphical capabilities of counterparts seem to affect their coupling. Even circumstances like how serious they take the game or encouragement of one of peers seems to be effective in the game procedure which can be assess in further works.

In essence, in the path of converging linguistic and dual eye tracking studies, there was a remarkable effect of color and Gaze in favor of RE terms related. albeit the experimental condition changes did not caused improvements in gaze matches and the hypothesis of the adjacency of Farsi RE's usage amount and participant's gaze overlap is refuted, but the second trial's gaze overlap was dramatic and coupling rate based on group results seemed to be hand in hand with participants' successful task quantities.

Beneath the limitations of the study and future works are reported:

5.1. Limitations of the Study

- As the test was done over pairs with same gender, managing time in order to access to both participants at the same time create difficulties.
- Eye trackers couldn't detect the eye movements of participants who were using thick glasses, so one pairs' data had to be excluded from analysis.
- Hand movements of participants during explaining create difficulties although they were asked not to use hands but sometimes it happened unconsciously.

5.2. Further Works

- Instead of cross- recurrence analysis, Levenshtein's Edit Distance (1996) might be use for finding the differentiations of coordination of two time series. This algorithm measures differences between two strings; it computes the number of characters required to change from one string (source) to the other one (target). For example for book and back produces the distance 2.
- The role of other cues such as numbering over the shapes or making the puzzle over the goal pattern in solving area in order to eliminate size misunderstanding can be further ground for research to identify which kind of hints create more adjacency between peers.
- The place of referring expressions can be studied under the cover of Centering framework or Givenness Hierarchy framework.
- The effect of using different referring expressions over the status of game and participants actions can be evaluated.
- As it shown in Functional analysis of 'This' and 'That' it can be extended to the whole data and it can be compare with other languages similar results.

.

REFERENCES

Acartürk, C., & Çakır, M. P. (2012). Towards Building a Corpus of Turkish Referring Expressions. In S. Demir, I. D. El-Kahlout & M. U. Dogan (Eds.), *Proceedings of the Workshop on Language Resources and Technologies for Turkic Languages. LREC 2012 Language and Resources Evaluation Conference.* (pp. 1-5) Istanbul, Turkey.

Amtrup, J. W., Rad, H. M., Megerdoomian, K., & Zajac, R. (2000). Persian-English machine translation: An overview of the Shiraz project. Memoranda in *Computer and Cognitive Science* MCCS-00-319, NMSU, CRL.

Braje, W., Legge, G., & Kersten, D. (1999). The influence of shadows on the recognition of natural objects. *Perception*.

Bruijn, J.T.P. (2015). Persian literature. In *Encyclopædia Britannica, Inc*. Retrieved from http://www.britannica.com/EBchecked/topic/452843/Persian-literature.

Bowmaker, J. K., & Dartnall, H. (1980). Visual pigments of rods and cones in a human retina. *The Journal of physiology*, *298*(1), 501-511.

Butterworth, G., Moore, C., & Dunham, P. (1995). *Joint Attention: Its origins and role in development. Origins of mind in perceptions and actions*. Hillsdale, NJ: Erlbaum

Byron, D. & Fosler-Lussier, E. (2006). The OSU Quake 2004 corpus of two party situated problem solving dialogs. In Proceedings of the Fifth Language Resources and Evaluation Conference (LREC 2006).

Chafe, W. (1987). Cognitive constraints on information flow. In R.S. Tomlin (ed.), *Coherence and grounding in discourse*, 21-51. Amsterdam: John Benjamins.

Clark, H. H., & Schaefer, E. F. (1989). Contributing to discourse. *Cognitive Science, 13*(2), 259-294

Clark, H. H., & Wilkes-Gibbs, D. (1986). Referring as a collaborative process. *Cognition, 22*(1), 1-39.

Clark, V. L. P., & Creswell, J. W. (2011). Designing and conducting mixed methods research. Sage Publications (pp. 104-106)

Clements, D. H., & Battista, M. T. (1992). Geometry and spatial reasoning. In: Douglas, A. G. (Ed.), *Handbook of Research on Mathematics Teaching and Learning* (pp.420-464). New York: Macmillan.

Coleman, J. L. (2008). The development of understanding of the concept of variable in grade seven beginning algebra students: the role of student interaction. Master Dissertation, Kingston, Canada: Queen's University.

Dale, R., Kirkham, N. Z., & Richardson, D. C. (2011). The dynamics of reference and shared visual attention. Frontiers in psychology, 2(pp. 1-11).

Dale, R., & Viethen, J. (2009). Referring expression generation through attribute-based heuristics. In Proceedings of the 12th European workshop on natural language generation (ENLG 2009) (pp. 58–65).

Di Eugenio, B., Jordan, P. W., Thomason, R. H. & Moore, J. D. (2000). The agreement process: An empirical investigation of human-human computer mediated collaborative dialogues. International Journal of Human-Computer Studies, 53(6): 1017-1076.

Duchowski, A. (2007). *Eye tracking methodology: Theory and practice*. New York, NY: Springer.

Eriksson, M. (2009). Referring as interaction: On the interplay between linguistic and bodily practices. *Journal of Pragmatics, 41*(2), 240-262.

Fillmore, C. J. (2014). An Alternative to Checklist Theories of Meaning. Annual Meeting of the Berkeley Linguistics Society, 1(0), 123–131. http://doi.org/10.3765/bls.v1i0.2315

Grosz, B. J., Weinstein, S., & Joshi, A. K. (1995). Centering: A framework for modeling the local coherence of discourse. *Computational Linguistics, 21*(2), 203-225.

Gundel, J. K., Bassene, M., Gordon, B., Humnick, L., & Khalfaoui, A. (2010). Testing predictions of the Givenness Hierarchy framework: A crosslinguistic investigation. *Journal of Pragmatics, 42*(7), 1770-1785.

Gundel, J. K. & Hedberg, N. (2008). *Reference: Interdisciplinary Perspectives*. New York, NY: Oxford University Press.

Gundel, J. K., Hedberg, N., & Zacharski, R. (1993). Cognitive status and the form of referring expressions in discourse. *Language*, 274-307.

Gundel, J. K., Hegarty, M., & Borthen, K. (2003). Cognitive status, information structure, and pronominal reference to clausally introduced entities. *Journal of Logic, Language and Information, 12*(3), 281-299.

Gustafsson, E., Brisson, J., Beaulieu, C., Mainville, M., Mailloux, D., & Sirois, S. (2015). How do infants recognize joint attention? *Infant Behavior and Development*, 40, 64-72.

Holmqvist, K., Nyström, M., Andersson, R., Dewhurst, R., Jarodzka, H., & Van de Weijer, J. (2011). *Eye tracking: A comprehensive guide to methods and measures*. New York, NY: Oxford University Press.

Hutchins, E. (1995). How a cockpit remembers its speeds. *Cognitive Science, 19*(3), 265-288.

Jermann, P., & Marc-Antoine, N. (2012). Effects of Sharing Text Selections on Gaze Cross-Recurrence and Interaction Quality in a Pair Programming Task. In Proceedings of the ACM 2012 Conference on Computer Supported Cooperative Work, 1125–34. CSCW '12. New York, NY, USA: ACM, 2012. doi:10.1145/2145204.2145371.

Karsh, R., & Breitenbach, F. W. (1983). Looking at looking: *The amorphous fixation measure. Eye movements and psychological functions:* International views, (pp. 53-64).

Lin, C. P., Shao, Y. J., Wong, L. H., Li, Y. J., & Niramitranon, J. (2011). The Impact of Using Synchronous Collaborative Virtual Tangram in Children's Geometric. Turkish Online Journal of Educational Technology-TOJET, 10(2), 250-258.

Nussli, M.A. (2011). Dual eye-tracking methods for the study of remote collaborative problem-solving. PhD thesis, École Polytechnique Fédérale De Lausanne

Thompson, R. F., & Madigan, S. A. (2013). Memory: The Key to Consciousness: *The Key to Consciousness*. New Jersey, Princeton University Press.

Prince, E. (1981). Toward a taxonomy of given/new information. In P. Cole (ed.), *Radical pragmatics*, 223-255. New York, NY: Academic Press.

Richardson, D. C., & Dale, R. (2005). Looking to understand: The coupling between speakers' and listeners' eye movements and its relationship to discourse comprehension. *Cognitive Science, 29*(6), 1045-1060.

Scarlatos, L., Landy, S., Breban, J., Horowitz, R., & Sandberg, C. (2002). On the effectiveness of tangible interfaces in collaborative learning environments. Proceedings of SIGGRAPH Conference 2002 (pp. 125-126), San Antonio, TX.

Sedighian K., & Klawe, M. (1996). Super Tangrams: A childcentered approach to designing a computer supported mathematics learning environment. International Conference on Learning Sciences 1996) pp. 490-495), Evanston, IL.

Sharma, K., Jermann, P., & Dillenbourg, P. (2015). Displaying Teacher's Gaze in a MOOC: Effects on Students' Video Navigation Patterns. In *Proceedings of the 10th European Conference on Technology Enhanced Learning* (No. EPFL-CONF-208029).

Sharma, K., Jermann, P., Nüssli, M. A., & Dillenbourg, P. (2013). Understanding collaborative program comprehension: Interlacing gaze and dialogues. In *Computer Supported Collaborative Learning* (CSCL 2013) (No. EPFL-CONF-184007).

Sharma, K., Jermann, P., Nüssli, M. A., & Dillenbourg, P. (2012). Gaze evidence for different activities in program understanding. In *24th Annual conference of Psychology of Programming Interest Group* (No. EPFL-CONF-184006).

Sternberg, R .J. (2004). *International Handbook of Intelligence*. Cambridge: Cambridge University Press. (pp. 330–331).

Shokouhi, H. (1996). Null referring expressions in a conversational context. *La Trobe Working Papers in Linguistics, 9*, 135–153.

Slocum, J. (2001). *The Tao of Tangram*. New York, NY: Barnes & Noble.Spanger, P., Yasuhara, M., Iida, R., Tokunaga, T., Terai, A., & Kuriyama, N. (2012). REX-J: Japanese referring expression corpus of situated dialogs. *Language Resources and Evaluation, 46*(3), 461-491.

Cole, P. (1978). Syntax and semantics, vol. 9: Pragmatics. *New York/San Francisco/London*, 113-127.Stoia, L., Shocley, D. M., Byron, D. K., & Fosler-Lussier, E. (2008). SCARE: A situated corpus with annotated referring expressions. In Proceedings of the Sixth International Conference on Language Resources and Evaluation (LREC 2008).

Takenobu, T., Masaaki, Y., Asuka, T., Morris, D., & Belz, A. (2010). Construction of bilingual multimodal corpora of referring expressions in collaborative problem solving. *In Proceedings of the Eighth Workshop on Asian Language Resources*. pp.38 – 46.

Van Deemter, K. (2007). TUNA: Towards a unified algorithm for the generation of referring expressions. Technical report, Aberdeen University. http://www.csd.abdn.ac.uk/research/tuna/pubs/TUNA-final-report.pdf

Vera-Díaz, F. A., & Doble, N. (2012). The Human Eye and Adaptive Optics. INTECH Open Access Publisher.

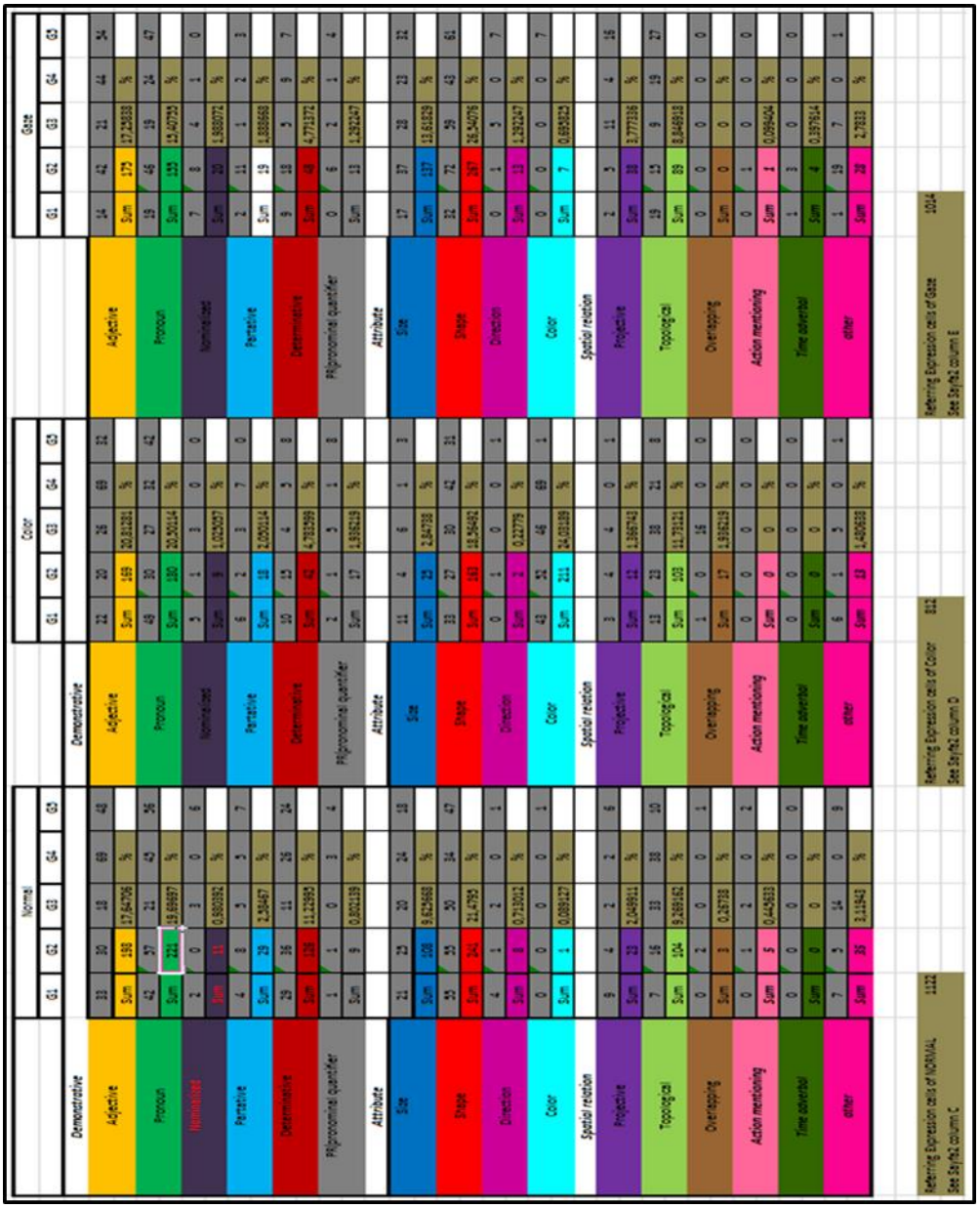
Yarbus, A. L. (1967). *Eye movements and vision*. New York: Plenum Press.

Yoshida, E. (2008). Patterns of use of referring expressions in English and Japanese dialogues. PhD thesis, University of Edinburgh. Retrieved from https://www.era.lib.ed.ac.uk/bitstream/handle/1842/4036/Yoshida2008.pdf?sequence=1

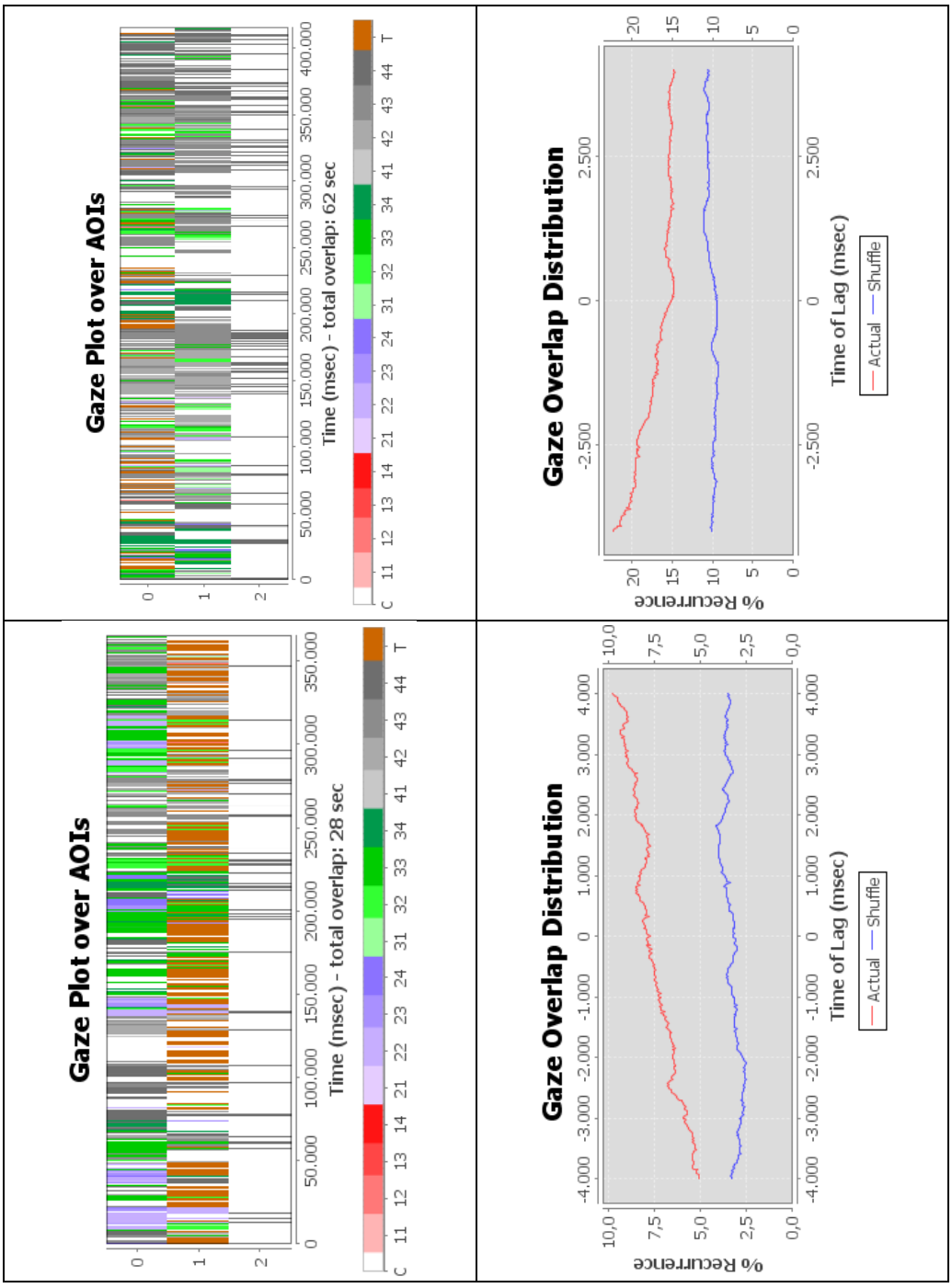
Yoshida, E. (2011). *Referring Expressions in English and Japanese: Patterns of use in dialogue processing*. (Vol. 208). Amsterdam/Philadelphia: John Benjamins Publishing.

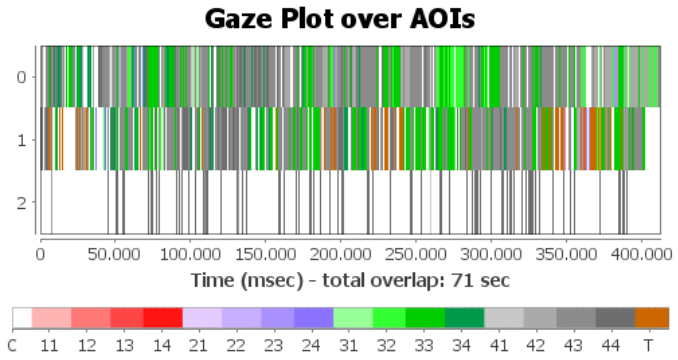
APPENDICES

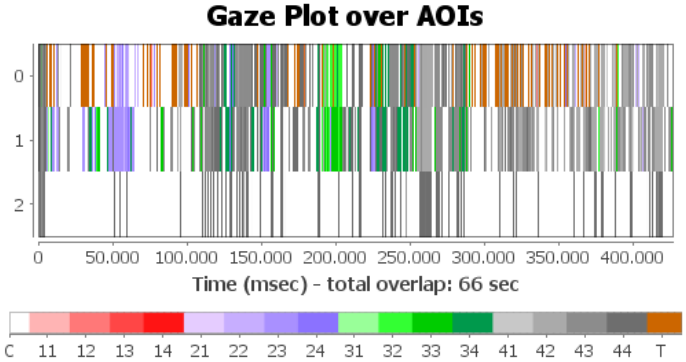
APPENDIX A: EXCEL RESULTS FOR DISTRIBUTION OF RE FEATURES FOR WHOLE GROUPS

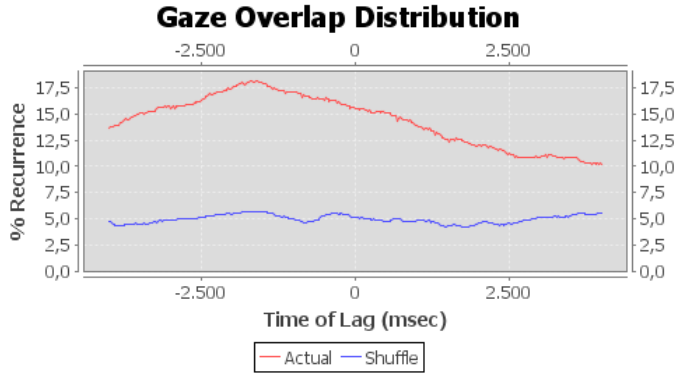


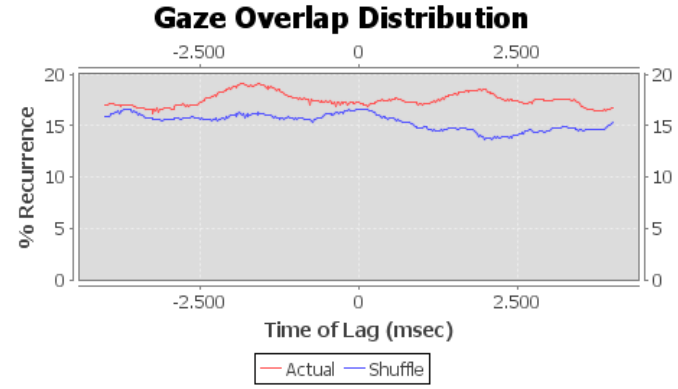
APPENDIX B: GAZE ANALYSIS OUTPUTS; SCARFPLOTS FOR AOI AND GAZE OVERLAP DISTRIBUTION

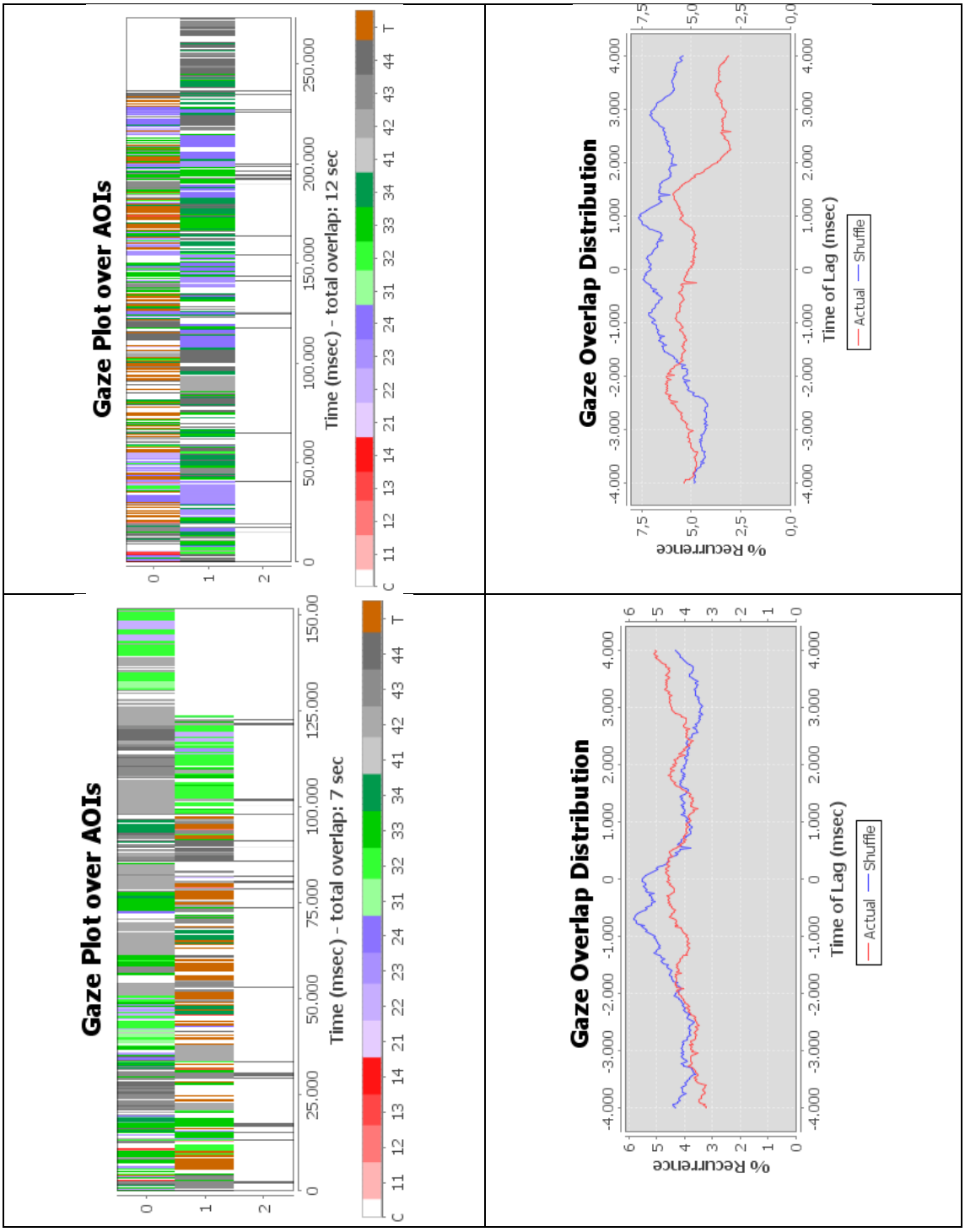


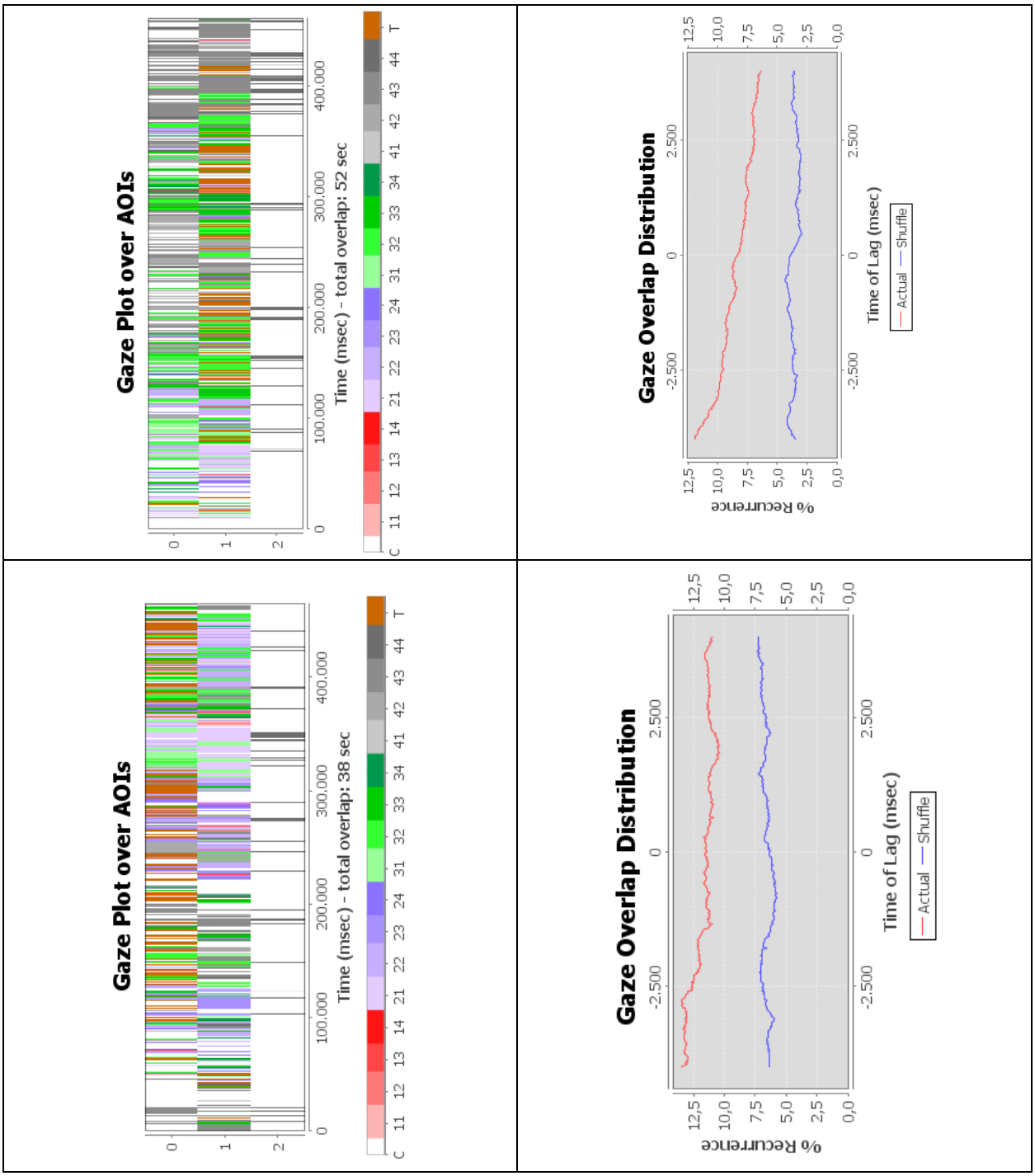

Gaze Plot over AOIs
0
1
2
0 50.000 100.000 150.000 200.000 250.000 300.000 350.000 400.000
Time (msec) - total overlap: 71 sec
C 11 12 13 14 21 22 23 24 31 32 33 34 41 42 43 44 T

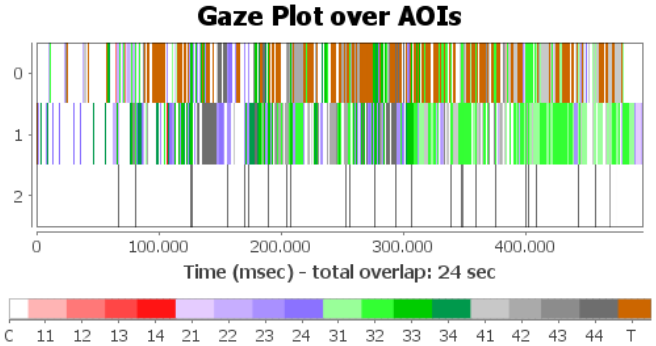

Gaze Plot over AOIs
0
1
2
0 50.000 100.000 150.000 200.000 250.000 300.000 350.000 400.000
Time (msec) - total overlap: 66 sec
C 11 12 13 14 21 22 23 24 31 32 33 34 41 42 43 44 T

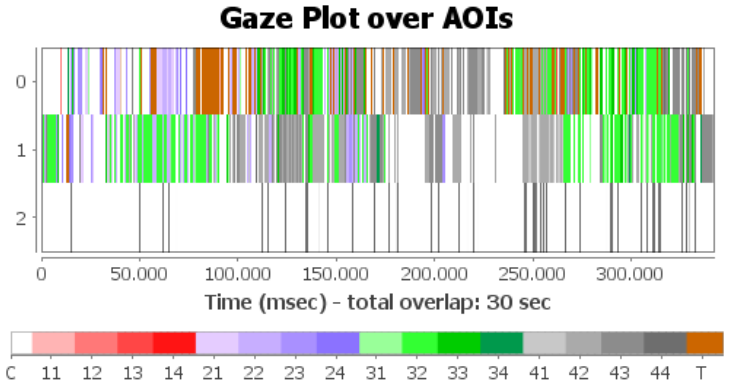

Gaze Overlap Distribution
-2.500 0 2.500
% Recurrence
17,5
15,0
12,5
10,0
7,5
5,0
2,5
0,0
Time of Lag (msec)
Actual Shuffle

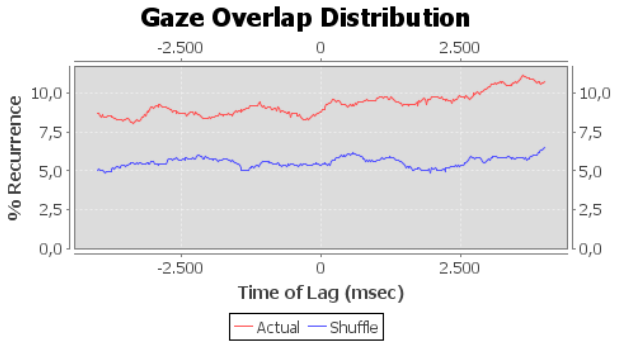

Gaze Overlap Distribution
-2.500 0 2.500
% Recurrence
20
15
10
5
0
Time of Lag (msec)
Actual Shuffle

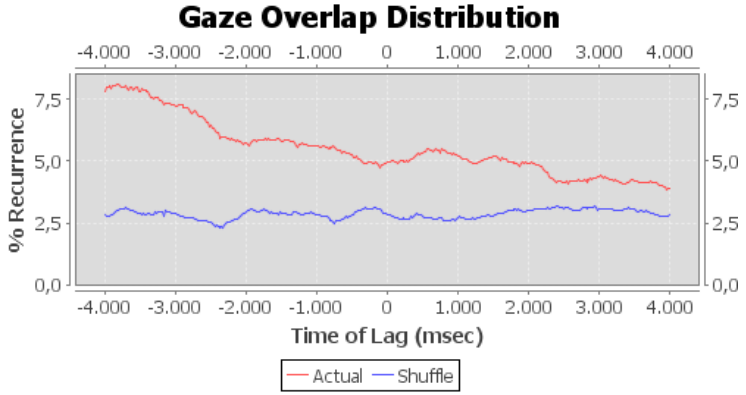

Gaze Plot over AOIs
Time (msec) - total overlap: 12 sec
Gaze Overlap Distribution
% Recurrence
Time of Lag (msec)
Actual
Shuffle
Gaze Plot over AOIs
Time (msec) - total overlap: 7 sec
Gaze Overlap Distribution
% Recurrence
Time of Lag (msec)
Actual
Shuffle

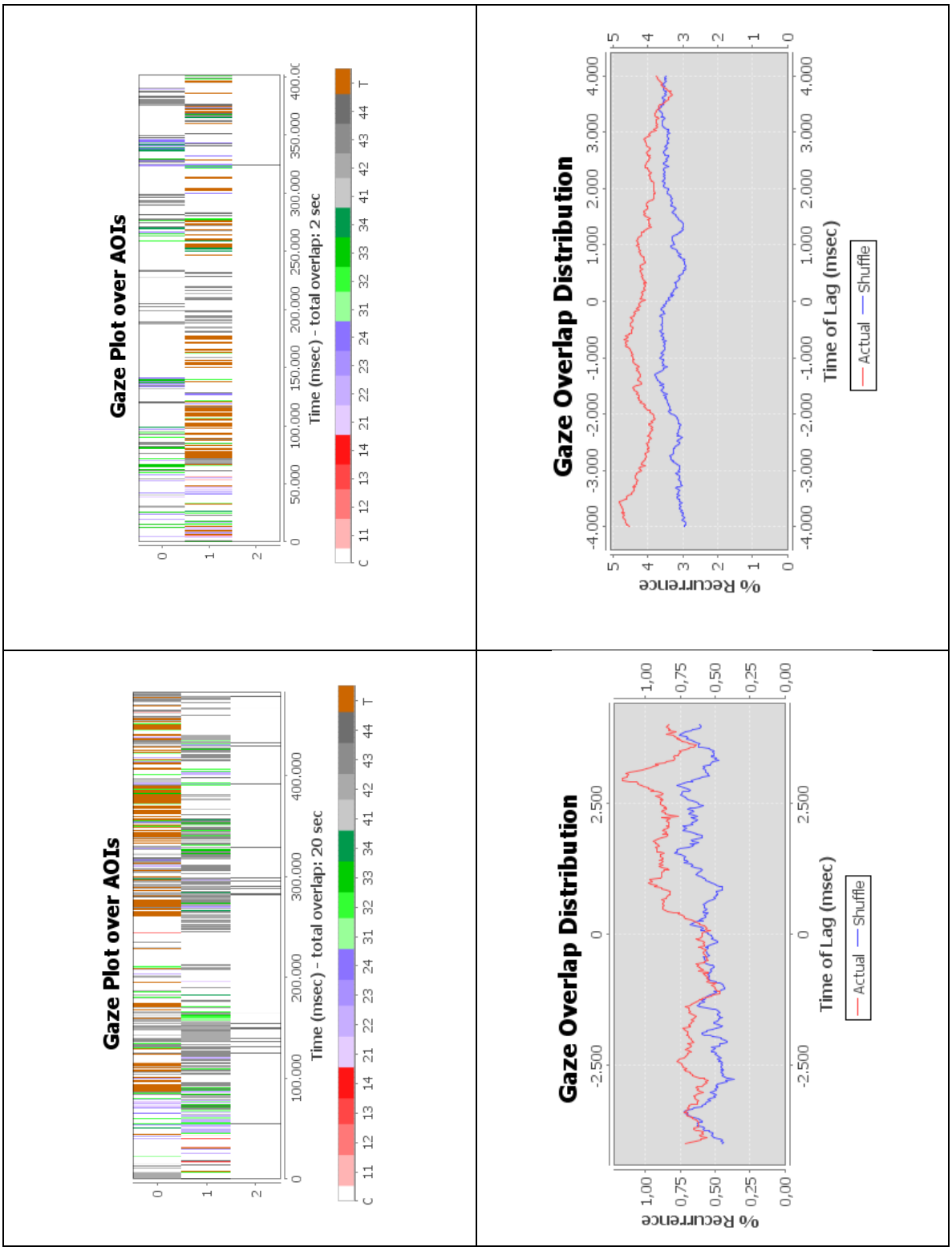

Gaze Plot over AOIs
Time (msec) - total overlap: 52 sec
Gaze Overlap Distribution
% Recurrence
Time of Lag (msec)
Actual
Shuffle
Gaze Plot over AOIs
Time (msec) - total overlap: 38 sec
Gaze Overlap Distribution
% Recurrence
Time of Lag (msec)
Actual
Shuffle

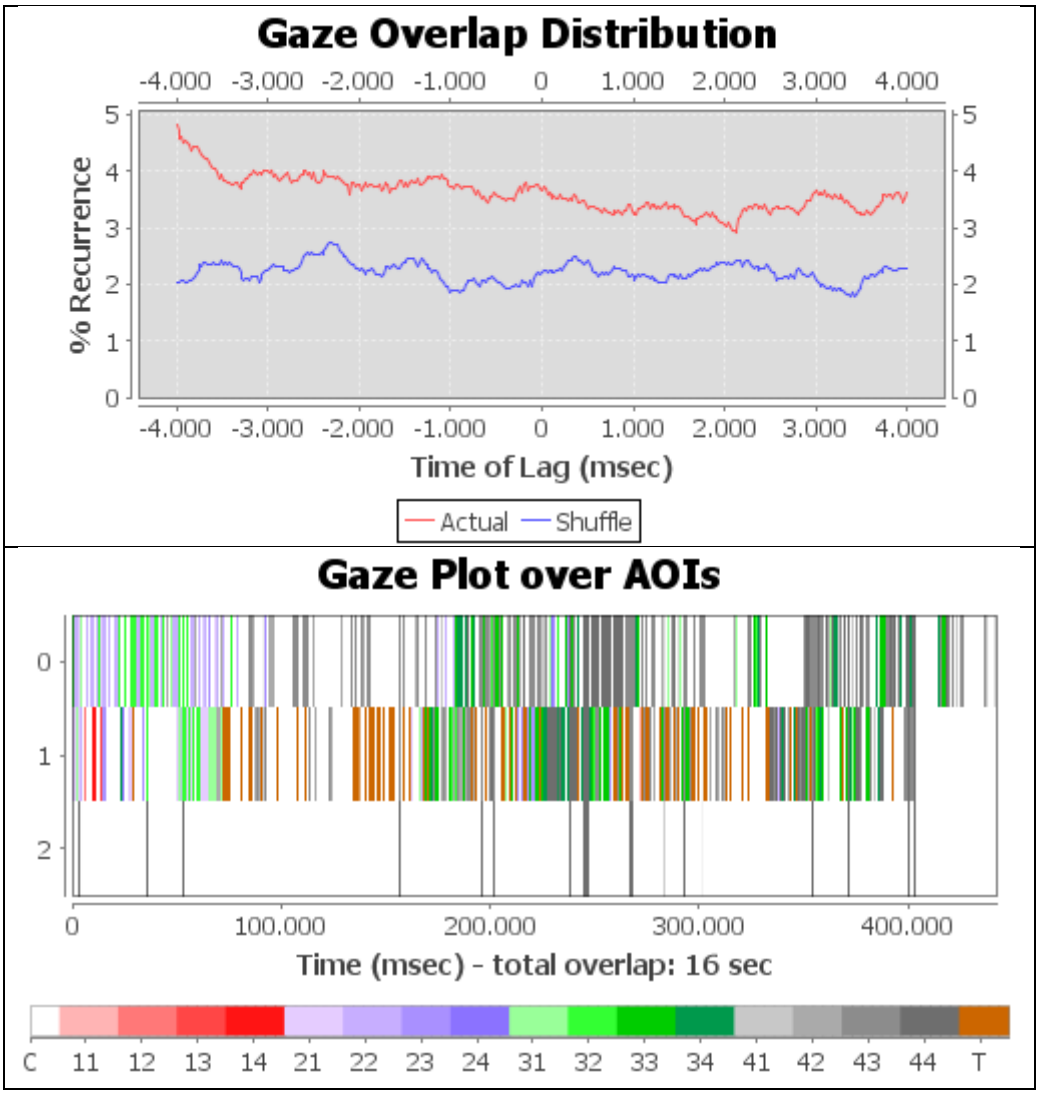

Gaze Plot over AOIs
0
1
2
0
100.000
200.000
300.000
400.000
Time (msec) - total overlap: 24 sec
C 11 12 13 14 21 22 23 24 31 32 33 34 41 42 43 44 T

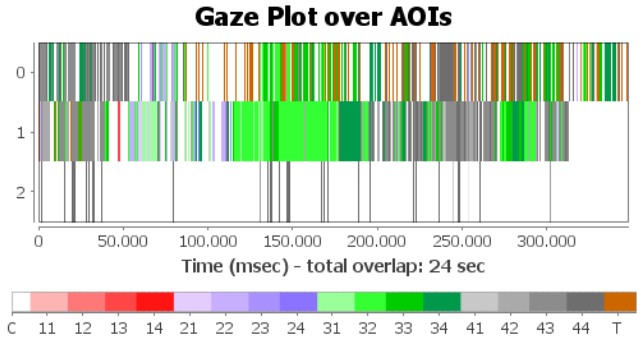

Gaze Plot over AOIs
0
1
2
0
50.000
100.000
150.000
200.000
250.000
300.000
Time (msec) - total overlap: 30 sec
C 11 12 13 14 21 22 23 24 31 32 33 34 41 42 43 44 T

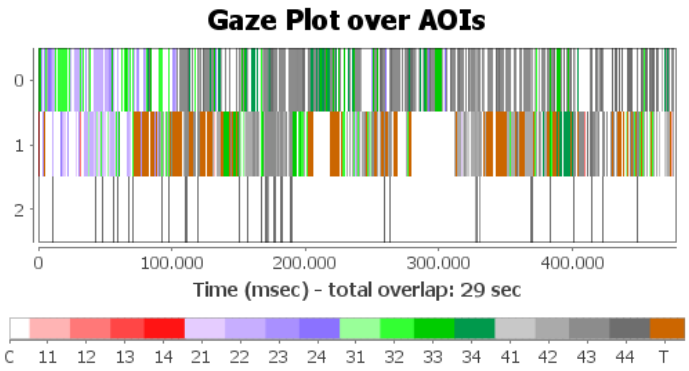

Gaze Overlap Distribution
-2.500
0
2.500
% Recurrence
10,0
7,5
5,0
2,5
0,0
Time of Lag (msec)
Actual
Shuffle

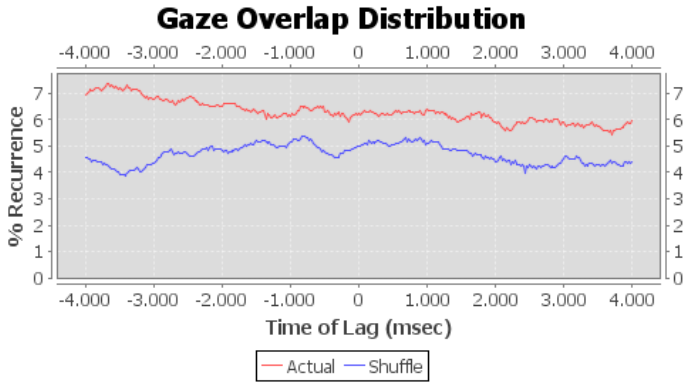

Gaze Overlap Distribution
-4.000
-3.000
-2.000
-1.000
0
1.000
2.000
3.000
4.000
% Recurrence
7,5
5,0
2,5
0,0
Time of Lag (msec)
Actual
Shuffle

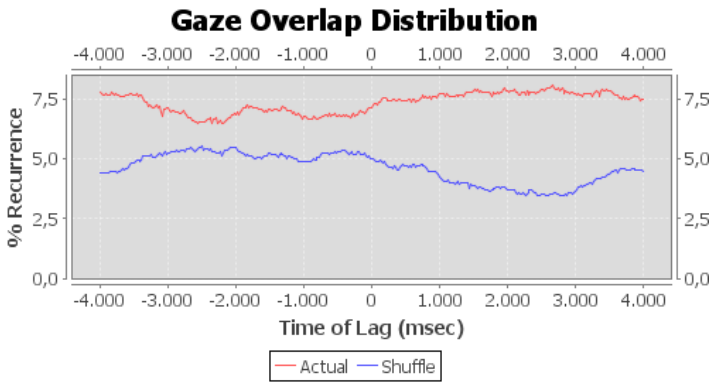

Gaze Plot over AOIs
Time (msec) - total overlap: 2 sec
Gaze Overlap Distribution
% Recurrence
Time of Lag (msec)
Actual
Shuffle
Gaze Plot over AOIs
Time (msec) - total overlap: 20 sec
Gaze Overlap Distribution
% Recurrence
Time of Lag (msec)
Actual
Shuffle

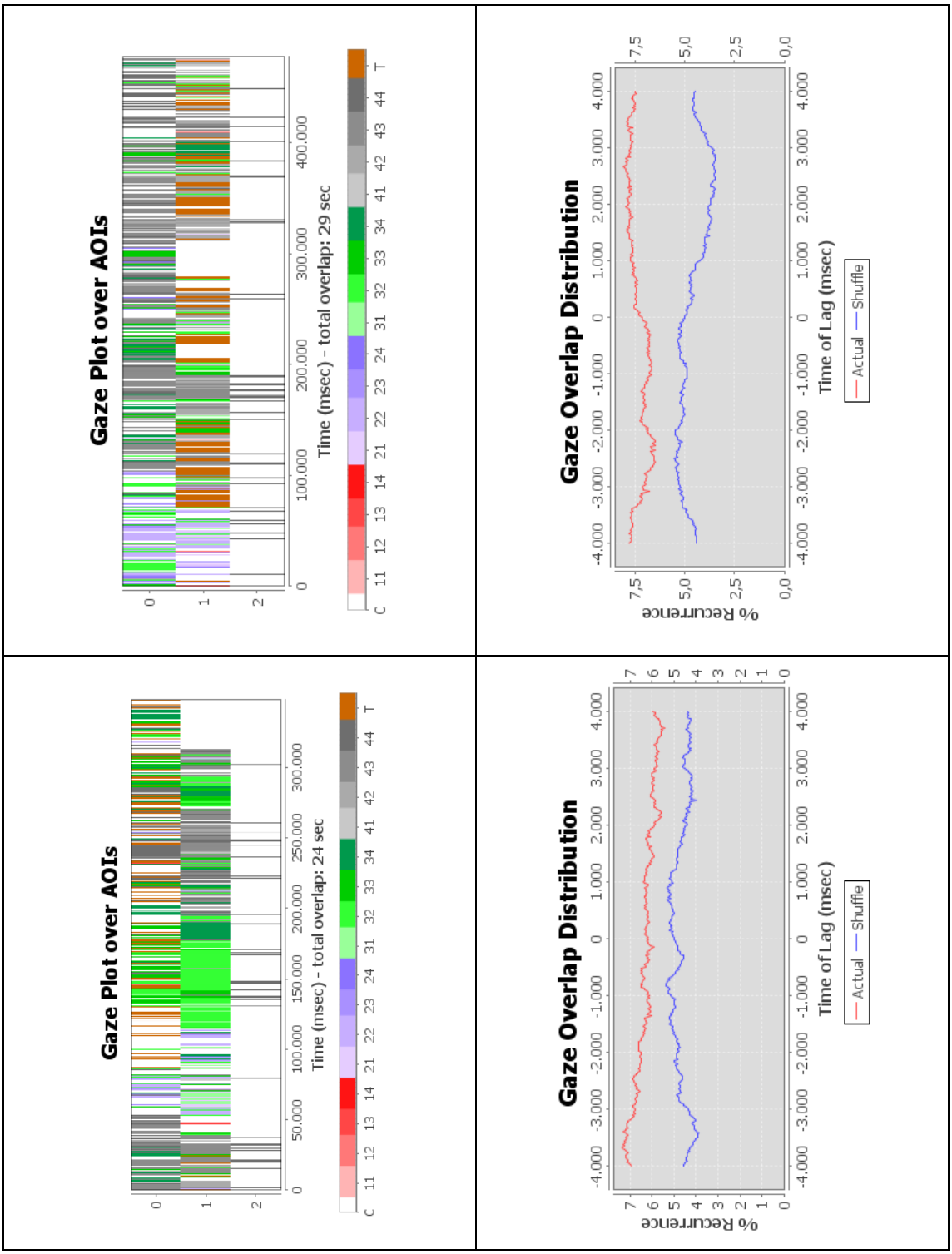

Gaze Overlap Distribution
-4.000 -3.000 -2.000 -1.000 0 1.000 2.000 3.000 4.000
% Recurrence
0 1 2 3 4 5
Time of Lag (msec)
Actual
Shuffle
Gaze Plot over AOIs
0 1 2
0 100.000 200.000 300.000 400.000
Time (msec) - total overlap: 16 sec
C 11 12 13 14 21 22 23 24 31 32 33 34 41 42 43 44 T

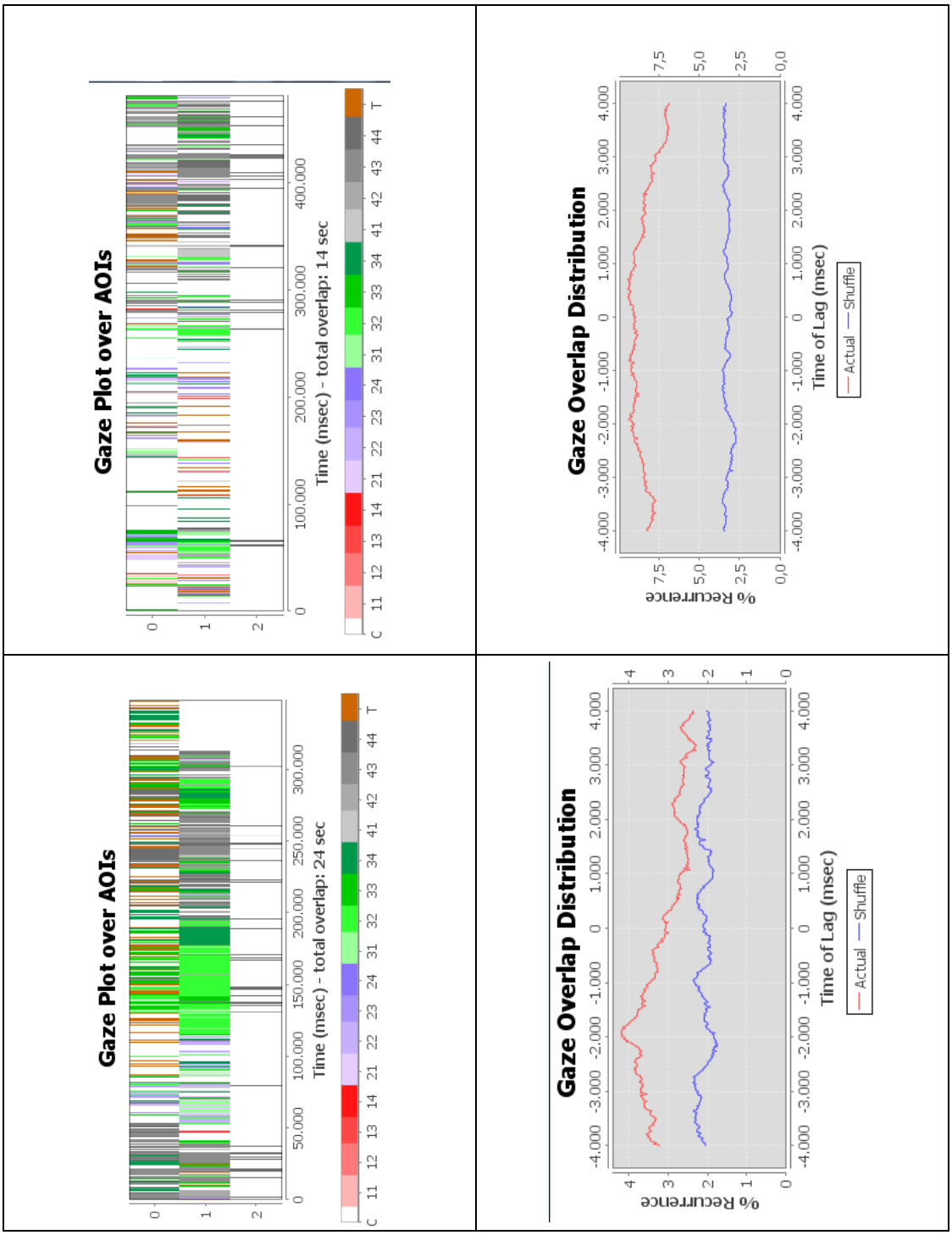

Gaze Plot over AOIs
0
1
2
0
50.000
100.000
150.000
200.000
250.000
300.000
Time (msec) - total overlap: 24 sec
C 11 12 13 14 21 22 23 24 31 32 33 34 41 42 43 44 T

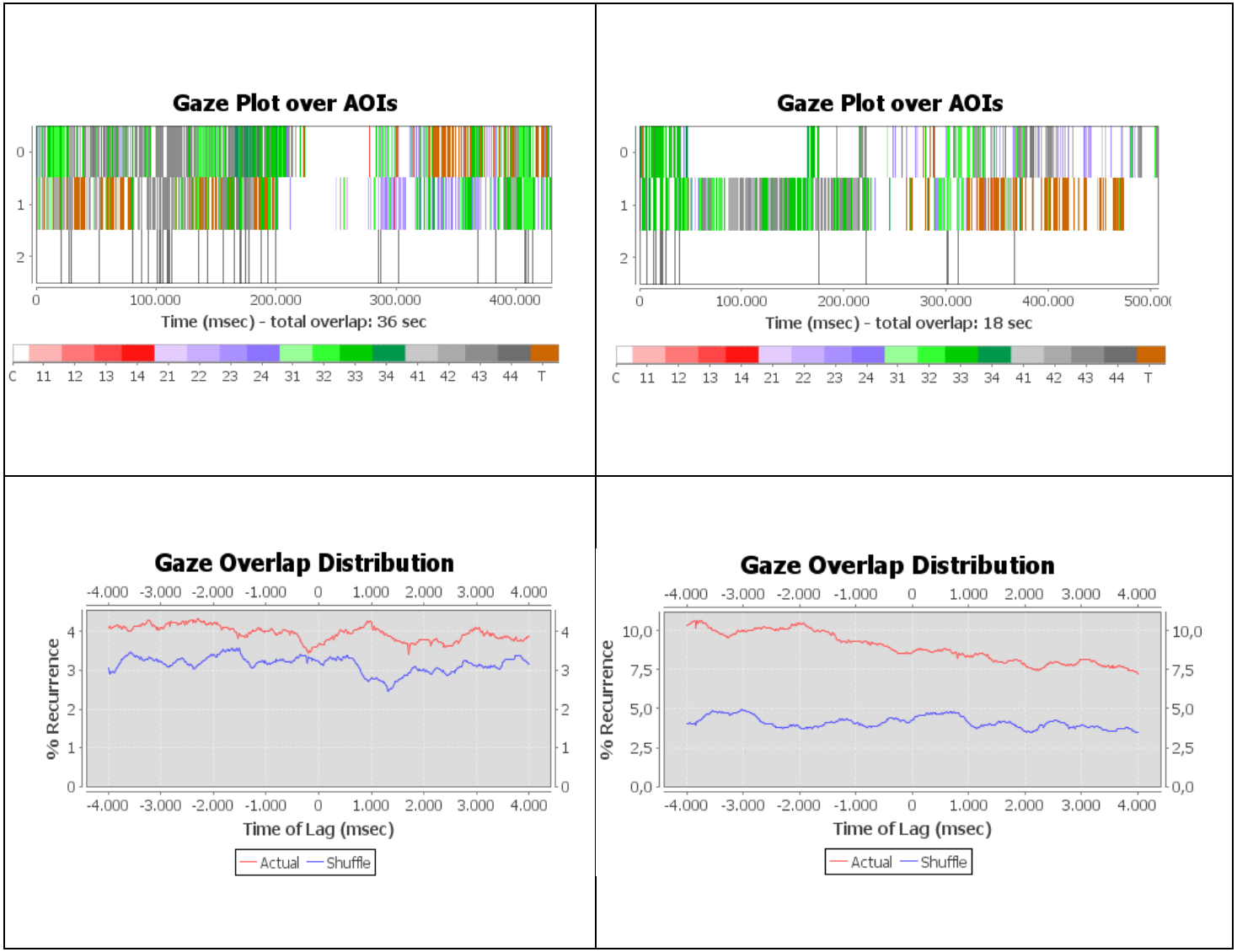

Gaze Plot over AOIs
0
1
2
0
100.000
200.000
300.000
400.000
Time (msec) - total overlap: 29 sec
C 11 12 13 14 21 22 23 24 31 32 33 34 41 42 43 44 T

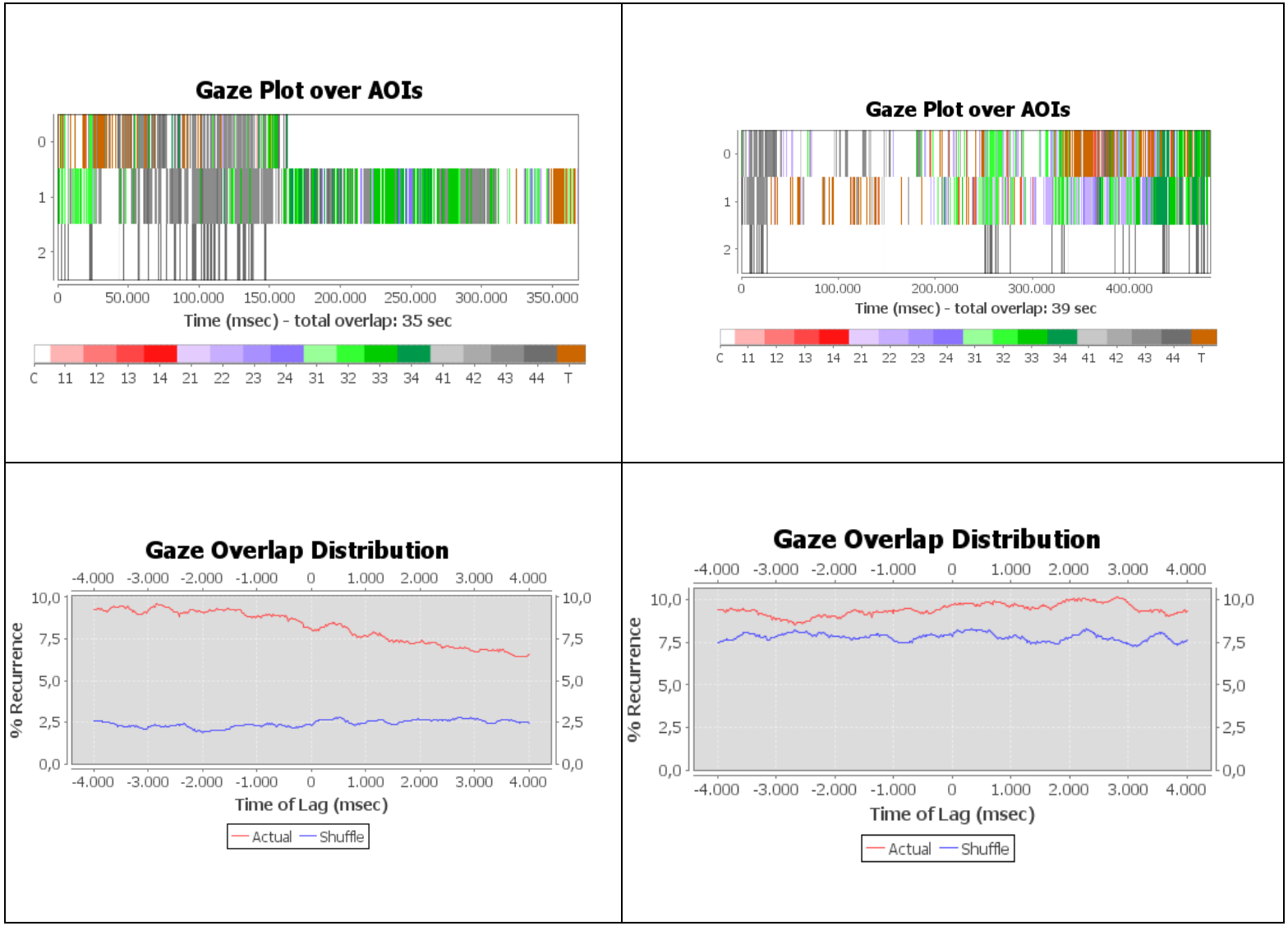

Gaze Overlap Distribution
-4.000 -3.000 -2.000 -1.000 0 1.000 2.000 3.000 4.000
% Recurrence
0 1 2 3 4 5 6 7
Time of Lag (msec)
Actual
Shuffle

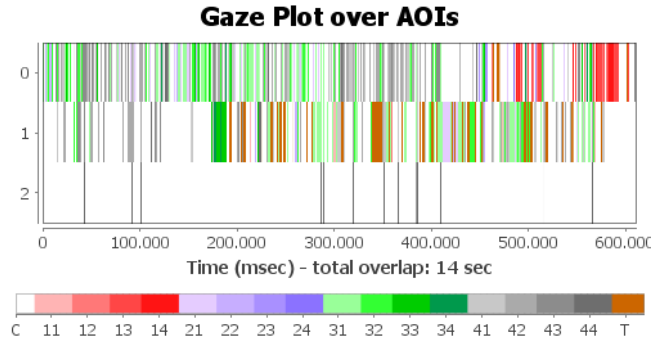

Gaze Overlap Distribution
-4.000 -3.000 -2.000 -1.000 0 1.000 2.000 3.000 4.000
% Recurrence
0,0 2,5 5,0 7,5
Time of Lag (msec)
Actual
Shuffle

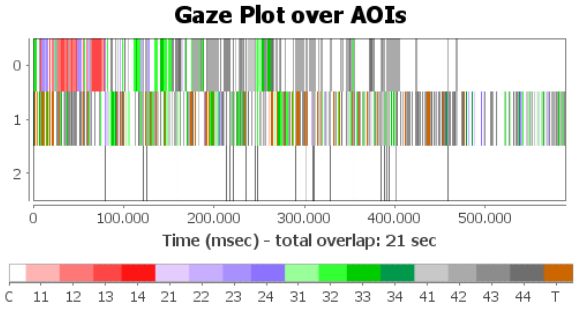

Gaze Plot over AOIs
Time (msec) - total overlap: 29 sec
Gaze Overlap Distribution
% Recurrence
Time of Lag (msec)
Actual
Shuffle
Gaze Plot over AOIs
Time (msec) - total overlap: 24 sec
Gaze Overlap Distribution
% Recurrence
Time of Lag (msec)
Actual
Shuffle

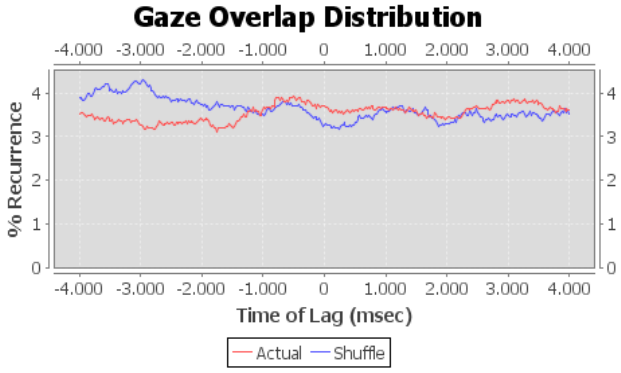

Gaze Plot over AOIs
Time (msec) - total overlap: 14 sec
Gaze Overlap Distribution
% Recurrence
Time of Lag (msec)
Actual
Shuffle
Gaze Plot over AOIs
Time (msec) - total overlap: 24 sec
Gaze Overlap Distribution
% Recurrence
Time of Lag (msec)
Actual
Shuffle

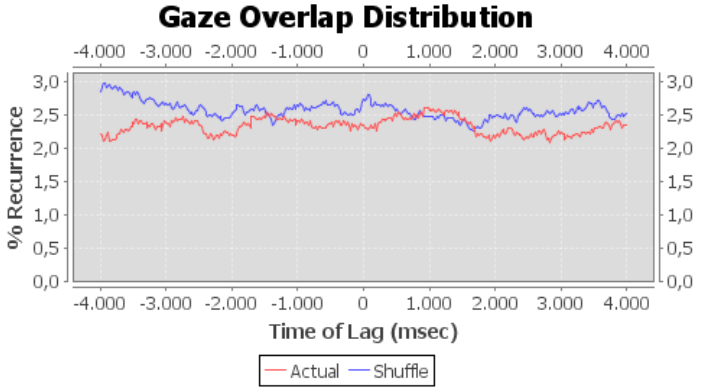

Gaze Plot over AOIs
0
1
2
0
100.000
200.000
300.000
400.000
Time (msec) - total overlap: 36 sec
C 11 12 13 14 21 22 23 24 31 32 33 34 41 42 43 44 T
Gaze Plot over AOIs
0
1
2
0
100.000
200.000
300.000
400.000
500.00
Time (msec) - total overlap: 18 sec
C 11 12 13 14 21 22 23 24 31 32 33 34 41 42 43 44 T
Gaze Overlap Distribution
-4.000 -3.000 -2.000 -1.000 0 1.000 2.000 3.000 4.000
% Recurrence
0 1 2 3 4
Time of Lag (msec)
Actual Shuffle
Gaze Overlap Distribution
-4.000 -3.000 -2.000 -1.000 0 1.000 2.000 3.000 4.000
% Recurrence
0,0 2,5 5,0 7,5 10,0
Time of Lag (msec)
Actual Shuffle

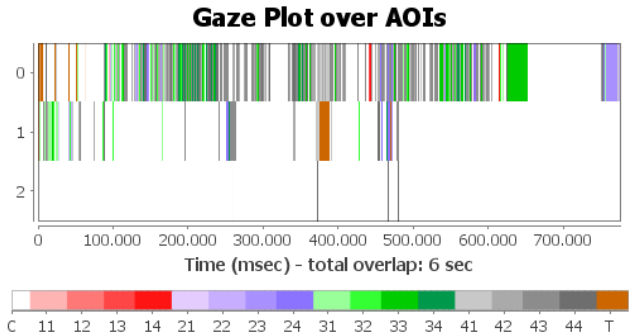

Gaze Plot over AOIs
0
1
2
0
50.000
100.000
150.000
200.000
250.000
300.000
350.000
Time (msec) - total overlap: 35 sec
C 11 12 13 14 21 22 23 24 31 32 33 34 41 42 43 44 T
Gaze Plot over AOIs
0
1
2
0
100.000
200.000
300.000
400.000
Time (msec) - total overlap: 39 sec
C 11 12 13 14 21 22 23 24 31 32 33 34 41 42 43 44 T
Gaze Overlap Distribution
-4.000 -3.000 -2.000 -1.000 0 1.000 2.000 3.000 4.000
10,0
7,5
5,0
2,5
0,0
% Recurrence
Time of Lag (msec)
Actual Shuffle
Gaze Overlap Distribution
-4.000 -3.000 -2.000 -1.000 0 1.000 2.000 3.000 4.000
10,0
7,5
5,0
2,5
0,0
% Recurrence
Time of Lag (msec)
Actual Shuffle

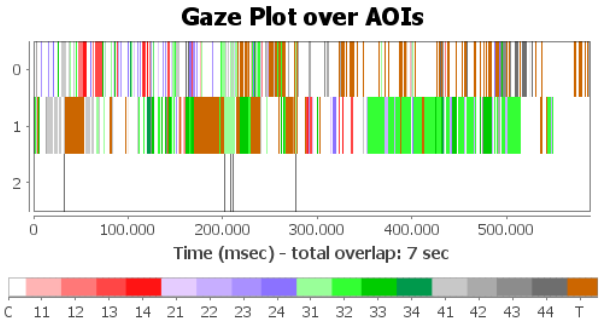

Gaze Plot over AOIs
0
1
2
0
100.000
200.000
300.000
400.000
500.000
600.000
Time (msec) - total overlap: 14 sec
C 11 12 13 14 21 22 23 24 31 32 33 34 41 42 43 44 T

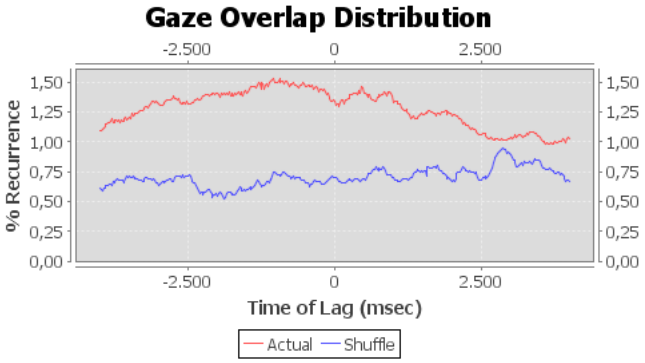

Gaze Plot over AOIs
0
1
2
0
100.000
200.000
300.000
400.000
500.000
Time (msec) - total overlap: 21 sec
C 11 12 13 14 21 22 23 24 31 32 33 34 41 42 43 44 T

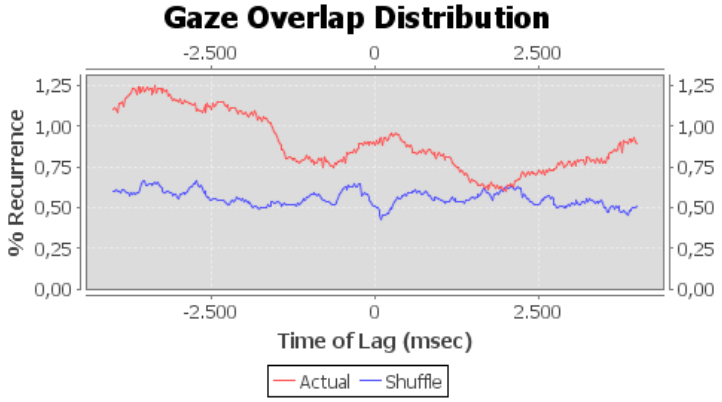

Gaze Overlap Distribution
-4.000 -3.000 -2.000 -1.000 0 1.000 2.000 3.000 4.000
% Recurrence
0
1
2
3
4
Time of Lag (msec)
Actual
Shuffle

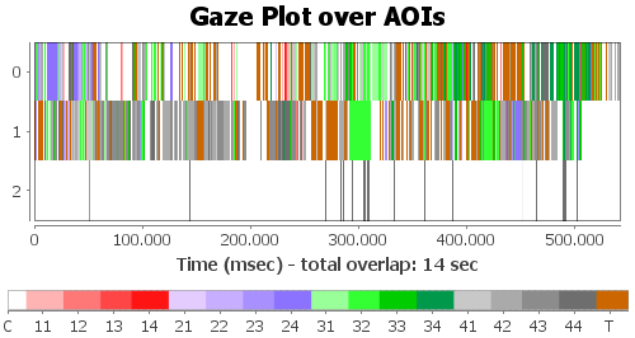

Gaze Overlap Distribution
-4.000 -3.000 -2.000 -1.000 0 1.000 2.000 3.000 4.000
% Recurrence
0,0
0,5
1,0
1,5
2,0
2,5
3,0
Time of Lag (msec)
Actual
Shuffle

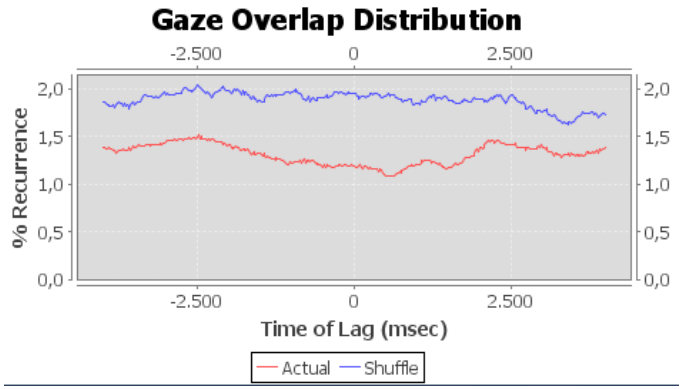

Gaze Plot over AOIs
0
1
2
0
100.000
200.000
300.000
400.000
500.000
600.000
700.000
Time (msec) - total overlap: 6 sec
C
11
12
13
14
21
22
23
24
31
32
33
34
41
42
43
44
T

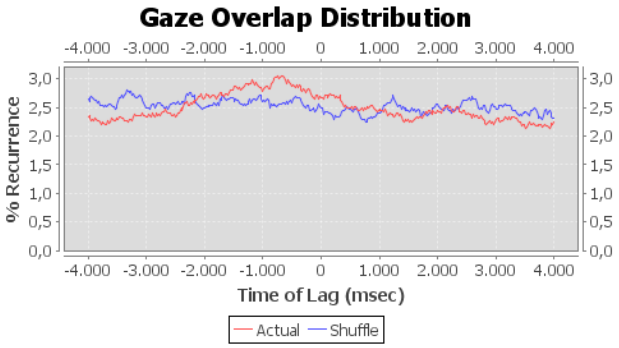

Gaze Plot over AOIs
0
1
2
0
100.000
200.000
300.000
400.000
500.000
Time (msec) - total overlap: 7 sec
C
11
12
13
14
21
22
23
24
31
32
33
34
41
42
43
44
T


Gaze Overlap Distribution
-2.500
0
2.500
% Recurrence
1,50
1,25
1,00
0,75
0,50
0,25
0,00
Time of Lag (msec)
Actual
Shuffle

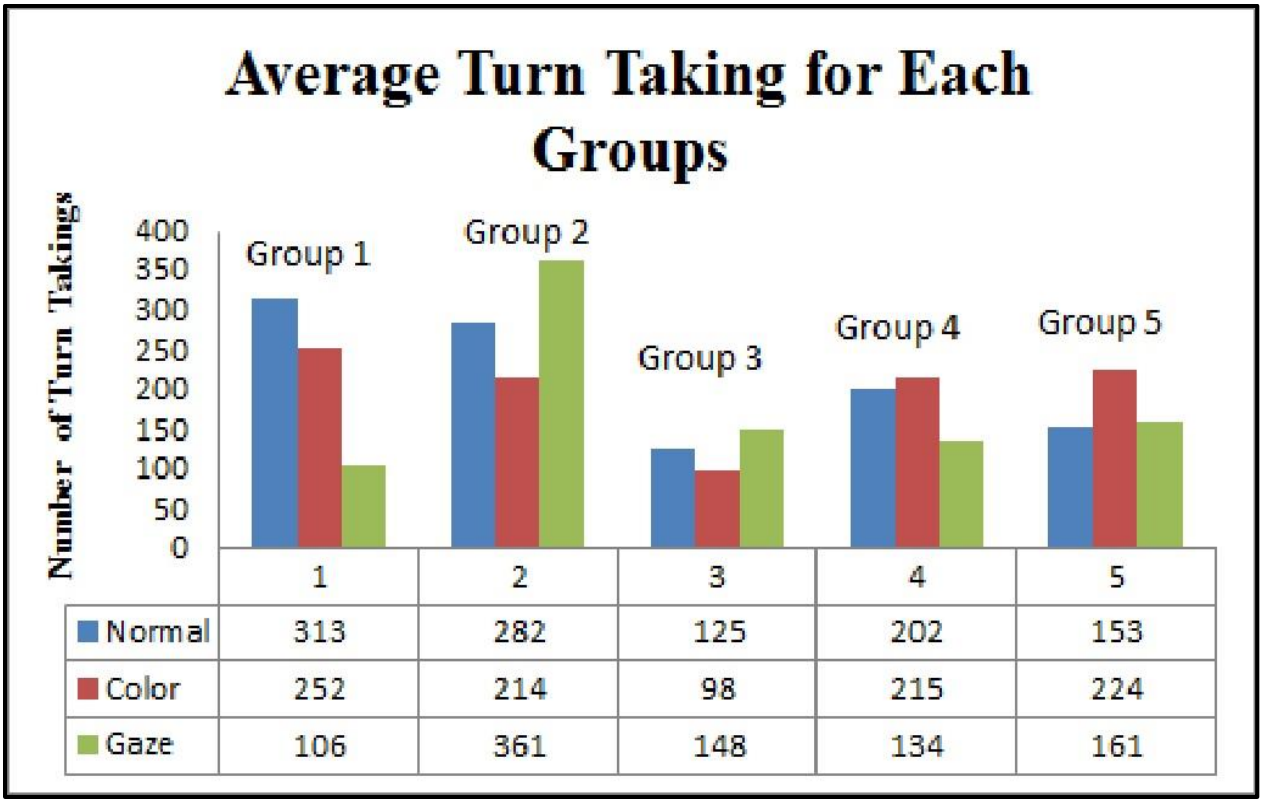

Gaze Overlap Distribution
-2.500
0
2.500
% Recurrence
1,25
1,00
0,75
0,50
0,25
0,00
Time of Lag (msec)
Actual
Shuffle


Gaze Plot over AOIs
0
1
2
0
100.000
200.000
300.000
400.000
500.000
Time (msec) - total overlap: 14 sec
C
11
12
13
14
21
22
23
24
31
32
33
34
41
42
43
44
T


Gaze Overlap Distribution
-2.500
0
2.500
% Recurrence
2,0
1,5
1,0
0,5
0,0
Time of Lag (msec)
Actual
Shuffle


Gaze Overlap Distribution
-4.000
-3.000
-2.000
-1.000
0
1.000
2.000
3.000
4.000
% Recurrence
3,0
2,5
2,0
1,5
1,0
0,5
0,0
Time of Lag (msec)
Actual
Shuffle

APPENDIX C: TURN TAKING RESULTS



	1	2	3	4	5
Normal	313	282	125	202	153
Color	252	214	98	215	224
Gaze	106	361	148	134	161

APPENDIX D: RELIABILITY RESULTS FOR TWO OF DIALOGUES

comparison	Original	Reliability	match	not match	
Adjective	16	18	16	2	
Pronoun	25	22	22	3	
Nominalized	3	1	1	2	
Partative	3	4	3	1	
Determinative	6	6	6	0	
PR(pronominal	3	4	3	1	
Attribute	0	0			
Size	17	19	17	2	
Shape	33	34	33	1	
Direction	0	0	0	0	
Color	0	0	0	0	
Spatial	0	0			
Projective	2	3	2	1	
Topological	7	7	7	0	
Overlapping	0	0	0	0	
Action mentioning	0	0	0	0	
Time adverbal	0	0	0	0	
other	4	3	3	1	PERCNTAGE OF RELIABILITY
	119	121	113	14	88.97638

COMPARISON	Original	Reliability	MATCH	NON-MATCH	
Adjective	11	12	11	1	
Pronoun	27	26	26	1	
Nominalized	2	3	2	1	
Partative	3	4	3	1	
Determinative	7	5	5	2	
PR(pronominal quantifier	1	1	1	0	
Attribute					
Size	16	16	16	0	
Shape	25	25	25	0	
Direction	1	1	1	0	
Color	0	0	0	0	
Spatial relation					
Projective	1	4	1	3	
Topological	6	5	5	1	
Overlapping	0	0	0	0	
Action mentioning	***0***	***0***	0	0	
Time adverbal	***0***	***0***	0	0	
other	***5***	***4***	4	1	PERCENTAGE OF RELIABILITY
			100	11	90.09009

TEZ FOTOKOPİSİ İZİN FORMU

ENSTİTÜ

Fen Bilimleri Enstitüsü ☐

Sosyal Bilimler Enstitüsü ☐

Uygulamalı Matematik Enstitüsü ☐

Enformatik Enstitüsü ☐

Deniz Bilimleri Enstitüsü ☐

YAZARIN

Soyadı : …….

Adı : …….

Bölümü : ……...

TEZİN ADI (İngilizce) : ……....

……
……
……
……...

TEZİN TÜRÜ : Yüksek Lisans ☐ Doktora ☐

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

TEZİN KÜTÜPHANEYE TESLİM TARİHİ : …….