DEMONSTRATIVES IN TURKISH: ON THE PRAGMATIC USE OF DEMONSTRATIVES IN THE CONTEXT OF A COLLABORATIVE PROBLEM SOLVING TASK

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

FARUK BÜYÜKTEKİN

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

AUGUST 2018

Approval of the thesis:

DEMONSTRATIVES IN TURKISH: ON THE PRAGMATIC USE OF DEMONSTRATIVES IN THE CONTEXT OF A COLLABORATIVE PROBLEM SOLVING TASK

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

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

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

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

Dr. Ceyhan Temürcü
Co-supervisor, **Cognitive Science, METU** _____

Examining Committee Members:

Prof. Dr. Hüseyin Cem Bozşahin
Cognitive Science Department, METU _____

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

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

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

Assist. Prof. Dr. Özkan Kılıç
Computer Engineering Department, Yıldırım Beyazıt University _____

Date: _____

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

Name, Last Name: FARUK BÜYÜKTEKİN

Signature :

ABSTRACT

DEMONSTRATIVES IN TURKISH: ON THE PRAGMATIC USE OF DEMONSTRATIVES IN THE CONTEXT OF A COLLABORATIVE PROBLEM SOLVING TASK

Büyüktekin, Faruk

M.S., Department of Cognitive Science

Supervisor : Assist. Prof. Dr. Murat Perit Çakır

Co-Supervisor : Dr. Ceyhan Temürcü

August 2018, 55 pages

Multimodal user interfaces have become increasingly important. They are being used in various applications such as computer games, online education, simulations, and assistive technology. Such interfaces now can make use of inputs coming from different modalities such as speech and eye gaze. This multimodality enables situated dialogue between users and artificial agents, encouraging collaborative language use. Naturally, people need to refer to entities in the surrounding environment using demonstratives during the conversation. Multimodal interfaces must resolve these referring expressions for successful communication. However, demonstrative reference is still a challenging problem due to the complex nature of the phenomenon. Traditional accounts of demonstratives are individual and speaker-centric, suggesting a distance contrast. On the contrary, recent findings reveal it is a joint and multimodal action. Consequently, there is a need to study demonstratives in a situated and distributed framework to better understand the cognitive processes behind reference generation and resolution. To this end, this study investigated the relationship between demonstrative use and eye gaze of the participants collaborating in a situated environment. Our findings suggest demonstrative reference does not occur randomly, rather follows certain cognitive principles, which questions the prevailing speaker-centric view. They clearly show it is a joint activity in which both the speaker and the listener play active roles. They also indicate eye gaze proves to be a significant visual cue, which is temporally linked with demonstrative use.

Keywords: Demonstratives, referring expressions, eye gaze, joint attention, situated dialogue

ÖZ

TÜRKÇE'DE İŞARET İFADELERİ: İŞBİRLİKÇİ BİR PROBLEM ÇÖZME GÖREVİ BAĞLAMINDA İŞARET İFADELERİNİN PRAGMATİK KULLANIMI

Büyüktekin, Faruk

Yüksek Lisans, Bilişsel Bilimler Programı

Tez Yöneticisi : Doktor Öğretim Üyesi Murat Perit Çakır

Ortak Tez Yöneticisi : Dr. Ceyhan Temürcü

Ağustos 2018 , 55 sayfa

Çok kipli kullanıcı arayüzleri giderek daha önemli hale gelmektedir. Halihazırda bilgisayar oyunları, çevrimiçi eğitimler, simülasyonlar ve yardımcı teknolojiler gibi çeşitli uygulamalarda kullanılmaktalar. Bu tür arayüzler artık konuşma ve göz gibi farklı kiplerden gelen girdileri kullanma imkanına sahip. Bu çok kiplilik, kullanıcı ve yapay aktörler arasında işbirlikçi dil kullanımını teşvik ederek yerleşik diyalog kurulmasını sağlamakta. Doğal olarak, insanlar konuşma sırasında işaret ifadelerini kullanarak çevredeki varlıklara göndermede bulunma ihtiyacı duymaktadır. Çok kipli arayüzler, başarılı iletişim için bu gönderme ifadelerini çözümlemelidir. Ancak, işaret göndermesi, olgunun karmaşık doğası nedeniyle hala önemli bir sorun olarak durmaktadır. İşaret ifadelerine dair geleneksel görüşler bireysel ve konuşmacı merkezli olup, uzaklık karşıtlığına dayanmaktadır. Ancak, yeni bulgular bunun işbirlikçi ve çok kipli bir eylem olduğunu ortaya koymaktadır. Sonuç olarak, referans üretimi ve çözümlemesinin ardındaki bilişsel süreçleri daha iyi anlamak için, yerleşik ve dağıtılmış bir çerçevede işaret ifadelerinin incelenmesi gerekmektedir. Bu amaçla, bu çalışma, yerleşik bir ortamda iş birliği yapan katılımcıların göz izleri ve işaret ifadesi kullanımları arasındaki ilişkiyi araştırmıştır. Bulgularımız, işaret göndermesinin rastgele ortaya çıkmadığını, aksine konuşmacı merkezli bakış açısını sorgulayan belirli bilişsel ilkeleri takip ettiğini ileri sürmektedir. Dahası, hem konuşmacı hem de dinleyicinin aktif rol oynadığı ortak bir faaliyet olduğunu ortaya koymuştur. Ayrıca, göz izinin, işaret ifadesi ile zamansal ilişkisinden dolayı önemli bir görsel yardımcı olabileceğini göstermiştir.

Anahtar Kelimeler: İşaret ifadeleri, gönderme ifadeleri, göz izi, ortak dikkat, yerleşik diyalog

To my family

ACKNOWLEDGMENTS

Like demonstrative reference, this thesis is a joint effort of many people. They have contributed in so many ways that I could not have written it without them.

I would like to express my deepest gratitude to my adviser Assist. Prof. Dr. Murat Perit Çakır for his guidance, understanding, and patience throughout this work. He always encouraged me to preserve and keep going on the way. I learnt so many things from him besides scientific study.

I am also indebted to Prof. Dr. Deniz Zeyrek Bozşahin, who was always welcoming and supportive. She introduced discourse to me and gave me the chance to work in the METU TDB project. It was very valuable.

I am also grateful to Prof. Dr. Cem Bozşahin for introducing me cognitive science as a field. He also introduced the computational theory and changed the way I look at language and grammar.

I am also thankful to Assoc. Prof. Dr. Cengiz Acartürk for introducing me vision science and how machines learn. I saw how I learnt complicated concepts so easily thanks to him.

I would like to thank Dr. Ceyhan Temürcü for his philosophical discussions about mind, language, and meaning. They were very enlightening.

I greatly appreciate Assoc. Prof. Dr. Annette Hohenberger for her effort to show us how to carry out an experiment from scratch.

I am also thankful to Hakan Güler and the rest of the Informatics family for always being helpful and friendly.

I would like to express my sincere gratitude to Murathan Kurfalı and Evren Aykaç. They were always there whenever needed since the beginning. Their company is priceless.

I am also indebted to Korkut and Ferda Koçak, Erkan and Açelya Tepe, and Mehmet Kara. Without them, everything would have been much more difficult. I owe it all to them.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

Suppose there is a couple at a theater who would like to see a movie, but cannot decide on which one. Gesturing toward one of the many posters, the man suggests the woman

"How about that?"

How is it possible for the woman to disambiguate the movie the man is referring to among many others? This might seem trivially easy to do. However, to be able to infer the referent, the movie in this case, the woman needs to attend the man's pointing gesture along with the utterance "that". Only then can she choose the most salient one among the competing candidates with the help of the common ground, mutual contextual information, between them.

Demonstratives are cross-linguistic phenomena. These expressions are among the most frequently used words in a language and among the first words learnt during language acquisition. Infants generally acquire content words earlier, but demonstratives are among the first function words babies learn (Clark, 2003). They are so fundamental and primordial in language that Diessel (2006) even argues the distinction between content and function words are not sufficient to characterize them.

In conversation, speakers often use words like *this* and *that*, and interlocutors try to disambiguate these expressions instantaneously. Occasionally they might fail, but generally they are able to identify the referent even though referring is an intricate endeavor. During the process, they cooperate for successful communication to occur, as Grice (1975) suggests. Speakers usually use a deictic pointing gesture such as eye gaze, head nod, or body orientation which interlocutors use as visual cues.

1.1 Thesis

This thesis aims to investigate the Turkish demonstratives *bu*, *şu*, and *o* within a social and multimodal perspective. The traditional accounts of demonstratives are individual and speaker-centric, suggesting a distance contrast. On the contrary, recent findings reveal it is a joint and multimodal action.

Consequently, there is a need to study demonstratives in a situated and distributed framework to better understand the cognitive processes behind reference generation and resolution. To

this end, we specifically looked at the relationship between the demonstrative use and eye gaze of the participants collaborating in a situated environment in order to answer the following questions:

- Do demonstrative expressions differ in terms of their frequency distribution and functional use in a situated dialogue?
- Do demonstrative expressions differ in terms of the role of the participants?
- Do demonstrative expressions differ in terms of cue conditions?
- Can eye gaze be used as a visual cue to resolve demonstrative reference?
- Is there a temporal relation between eye gaze and demonstrative use?
- Do demonstrative expressions differ in terms of the presence of joint attention?

1.2 Motivation

There is a considerable increase in the number of multimodal user interfaces used in various applications such as computer games, online education, simulations, and assistive technology. Such interfaces allow users to interact with computers in a natural way, making use of inputs coming from different modalities such as speech and eye gaze. This multimodality enables situated dialogue in virtual environments where users can communicate with artificial agents.

Unlike traditional dialogue systems, situated dialogue encourages language use as a collaborative activity. Naturally, people need to make exophoric references to entities in the surrounding environment using demonstratives during the conversation. Multimodal interfaces must resolve these referring expressions to facilitate the ongoing communication. However, demonstrative reference still remains a challenging problem due to the complex nature of the phenomenon among other things.

Previous studies show referring behavior follow certain cognitive principles. However, the egocentric view of demonstrative reference fails to explain them thoroughly. With an empirically supported social and multimodal perspective, we aim to explore the principles governing exophoric demonstrative use in Turkish and suggest a basis for further research.

1.3 Outline

This study is composed of five main chapters: Literature Review, Methodology, Results, Discussion, and Conclusion.

1. The first chapter firstly gives a brief account of what reference is and then focuses on demonstratives, their definition, types, and pragmatic uses. Secondly, it analyzes the types of Turkish demonstratives and summarizes the previous studies related to them. Thirdly, it explores the main approaches to demonstrative analysis and the essential aspects of demonstrative research. Finally, it provides a general overview of two central frameworks on reference with previous corpora on referring expressions.

2. The second chapter is mainly composed of two sections. The first section describes the experimental setup employed to build the corpus analyzed in this study, who the participants were, what apparatus was used, and the procedure followed. The second section outlines the transcription and annotation process of the data obtained from the participants.

3. The third chapter firstly outlines the results of the inter-rater reliability analysis regarding the annotation process and then gives a general overview of the data and reports the results of the statistical tests which investigate the relationship between demonstratives and eye gaze.

4. The fourth chapter chapter elaborates on the results presented in the previous chapter and discusses them in detail in relation with the research questions of this study.

5. The last chapter summarizes the findings of the study and suggests directions for future research.

CHAPTER 2

LITERATURE REVIEW

This chapter firstly gives a brief account of what reference is and then focuses on the definition of demonstratives, their types, and pragmatic uses. Secondly, it analyzes the types of Turkish demonstratives and summarizes the previous studies related to them. Thirdly, it explores the main approaches to demonstrative analysis and the essential aspects of demonstrative research. Finally, it provides a general overview of two central frameworks on reference and related studies on referring expressions.

2.1 Reference

Reference generation and resolution is a fundamental quality of human cognition. It involves the ability to represent objects, indicate to others what object(s) we are talking about and understand what object(s) others are talking about (Gundel & Hedberg, 2008). It sets a prime example of the interdisciplinary nature of cognitive science since the subject of reference necessitates almost all the areas in the field such as philosophy, linguistics and cognitive psychology to work together to explain the phenomenon.

Reference is an act where the speaker employs linguistic expressions to enable the audience to identify what is intended. It is strongly tied with the goals and intentions of the addresser and the addressee. Also, inference plays a significant role. For successful reference to occur, the audience needs to infer the entity implied by the speaker (Yule, 1996). Speaker reference is a four-place relation (Bach, 2008). It comprises a speaker, an audience, a referent, and an expression. The speaker uses the expression to refer the audience to the referent. The expressions employed in this process are called referring expressions. They mainly fall under three categories, *proper nouns* (John, London), *definite descriptions* (the king, this man), and *pronouns* (he, this).

2.2 Demonstratives

2.2.1 Definiton

While referring, people use demonstratives like *this* and *that* . Their use is a cross-linguistic phenomenon. Although all languages appear to have one or more demonstratives, coming up with a definition is a challenging task because their form, function, and use may differ a lot

from one language to another.

To define these function words, Diessel (1999) proposes three criteria:

- Firstly, they are deictic/pointing expressions serving syntactic functions.
- Secondly, they serve certain pragmatic functions by focusing the hearer's attention on objects and their locations in the speech situation and organizing information flow in a conversation.
- Finally, demonstratives serve a semantic function by encoding spatial distance and proximity.

Similarly, Dixon (2003) proposes eight properties which includes having deictic reference, spatial reference, making up a whole noun phrase, occurring with a noun phrase, substitution anaphora, substitution cataphora, textual anaphora, textual cataphora, and offers to use them to compare demonstratives of a given language with the 1st, 2nd, 3rd personal pronouns and definite article of the language due to their historical and etymological relationship, as shown in Figure 2.1 .

	nominal demonstratives		1st and 2nd person pronouns	3rd person pronouns		definite article
	this/ these	*that/ those*	*I, you, we*	*he, she, they*	*it*	*the*
1. Can have deictic function	✓	✓	✓	–	–	–
2. Has spatial reference	✓[x]	✓[x]	–	–	–	–
3. Can make up whole NP	✓[x]	✓[x]	✓	✓	✓	–
4. Can occur in NP with noun	✓	✓	✓[1]	–[2]	–	✓
5. Substitution anaphora	✓[x]	✓[x]	–	✓	✓	n/a[3]
6. Substitution cataphora	–	–	–	✓	✓	
7. Textual anaphora	✓[x]	✓[x]	–	–	✓	
8. Textual cataphora	✓[x]	✓[x]	–	–	✓	

[x] Although this is a property of nominal demonstratives in English, it is not shown by demonstratives in all languages.
1. This covers NPs such as *you women*.
2. It is possible to have sentences such as *They, the evil spirits, roamed around in the night*, but this is regarded as involving two NPs in apposition (*they* and *the evil spirits*) rather than a single NP.
3. Not applicable; only items which make up a whole NP can have anaphoric or cataphoric function.

Figure 2.1: Features identified by Dixon (2003)

2.2.2 Types

In the same typological study, Dixon (2003) divides this closed-class category of words into three main types: *nominal*, *local adverbial*, and *verbal demonstratives*. *Nominal demonstratives* may appear in a noun phrase with a noun (Ex. 1) or can make a noun phrase on their own as a pronoun (Ex. 2).

(1) [This stone] is hot.

(2) [This] is hot.

Local adverbials may occur either alone (Ex. 3) or with a noun taking local marking (Ex. 4).

(3) Put it [here].

(4) Put it (on the table) [there].

Verbal demonstratives may occur as the only verb in a predicate, or together with a lexical verb (Ex. 5), usually with an accompanying mimicking action.

(5) Do it like [this].

2.2.3 Pragmatic Uses

Halliday & Hasan (1976) group the pragmatic uses of demonstratives under two main categories: *exophoric* and *endophoric*. They use the notion *exophoric* for demonstratives that are used with reference to entities in the surrounding situation, the term *endophoric* for all other uses. Following this classification, Diessel (1999) further divides *endophoric use* into (i) *anaphoric*, (ii) *discourse deictic*, and (iii) *recognitional* subcategories as in Figure 2.2

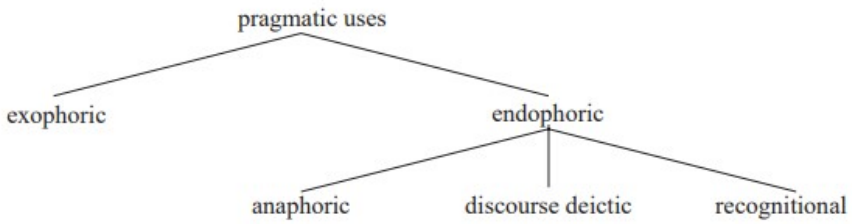


Figure 2.2: Pragmatic uses of demonstratives (Diessel, 1999)

With reference to the outside world, *exophoric* use of demonstratives focus the hearer's attention on people, objects or locations in the speech situation. *Anaphoric* and *discourse deictic* demonstratives are related to the ongoing discourse. While *anaphoric* demonstratives are coreferential with a noun phrase in the preceding discourse, *discourse deictic* uses refer to a non-NP chunk of the surrounding discourse such as sentences, string of sentences and verb phrases. *Recognitional* demonstratives are used to indicate that speaker and hearer are familiar with the referent due to shared experience.

2.3 Demonstratives in Turkish

Turkish has a three-term demonstrative system. Dixon (2003)'s all three types exist in the language. Turkish *nominal demonstratives* are *bu* (this), *şu* (that), and *o* (it). Syntactically, they can occur with a noun phrase as a determiner (Ex. 6) or make up a whole noun phrase alone as a pronoun (Ex. 9). When they are in the form of a pronoun in a sentence, they can inflect for number and/or case.

(6) [Bu ev] bizim.
THIS house ours
"This house is ours."

(7) [Şunları] sevdim.
THOSE-acc like-past-1sg
"I liked those."

Local adverbials are *bura*, *şura*, and *ora* . They usually occur alone, but sometimes with a noun phrase. Morphologically, they can inflect for number and/or case.

(8) [Buraya] gel.
HERE-dat come
"Come here."

Verbal demonstratives are *böyle*, *şöyle*, and *öyle*. They occur with a lexical verb, generally with a mimicking action.

(9) [Böyle] yapma.
Like THIS do-neg
"Don't do like this."

2.4 Previous Studies on Turkish Demonstratives

Early studies on the Turkish demonstrative system follows the traditional account and therefore indicate demonstratives *bu*, *şu*, and *o* encode the relative distance of the referent from either the speaker or the listener (Banguoğlu, 1974; Kornfilt, 1997). Turan (1996) focuses on the cognitive status of *bu* and *şu* in the attention structure in a text. The study reveals that *bu* and *şu* refer to different linguistics elements in a discourse. While *bu* refers to an entity on the right to signal continuation, *şu* refers to the forthcoming entity.

Unlike Turan (1996), Özyürek & Kita (2000) investigate Turkish demonstratives in conversational data (as cited in Kuntay & Ozyurek, 2002). They suggest the presence of joint attention between interlocutors is a powerful determinant of the speaker's choice of demonstrative. They argue *bu* and *o* are used when joint attention is achieved between the addresser and addressee. The difference between them lies in the relative distance of the referent to the interlocutors. On the other hand, *şu* is used when the addressee does not visually attend the intended referent. Their analysis is shown in Figure 2.3

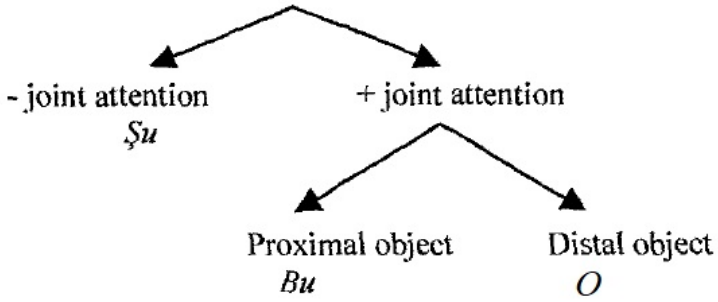


Figure 2.3: Analysis of Turkish demonstrative pronoun system by Özyürek & Kita (2000)

Kuntay & Ozyurek (2002) investigate the developmental pattern of learning to use *şu* in relation to *bu* and *o*. The results show adults' use of demonstratives differ from children's. Adults use more demonstratives than children per utterance. They use *şu* most ferequently, followed by *bu*. The study also shows there is a distance contrast between *bu* and *o*.

2.5 Analysis of Demonstratives

There are two main approaches to investigate demonstrative reference. The first one is the traditional egocentric view based on distance contrast. The second one is a relatively recent effort which is based on a social and multimodal understanding.

2.5.1 The Speaker-centric Account

The prevailing view regarding reference states it is an ego-centric act done by a speaker without worrying about the listener's belief and attention towards the referent (Olson, 1970; Clark & Bangerter, 2004). Naturally, the traditional views on demonstratives follow this speaker-centered understanding. Such views claim that the speaker places his ego at the center and sees everything from there while using a demonstrative reference (Lyons, 1977). Similarly, Diessel (2014) suggests the speaker employs a coordinate system based on the speaker's body while performing a demonstrative utterance and the demonstrative meaning heavily depends on it.

What is critical here is the semantics of the demonstratives. The speaker-oriented account claims they encode *physical distance* (Coventry et al., 2008). For instance, English, which accommodates a two-term demonstrative system, encodes physical proximity to the speaker with *this* and distance with *that*. Languages which have a three-term system like Turkish employs a medial demonstrative like *şu*, in addition to the proximal *bu* and the distal *o*. It is used for referents which are close to the addressee or at the middle distance from the speaker. This speaker-centric view is still very influential.

2.5.2 A Social and Multimodal Alternative

In contrast, relatively recent accounts of reference suggest it is a collaborative initiative in which the addresser and the addressee need to act together to establish a common ground about a particular entity (Clark & Wilkes-Gibbs, 1990; Brennan & Clark, 1996). The interaction between the speaker and listener is essential to this agreement and the listener contributes to the process not less than the speaker. While referring, speakers tend to utilize demonstratives very often to establish joint attention for a visible entity in the shared scene (Levinson, 1983).

Analyses of everyday multimodal and face-to-face spoken corpora reveal speakers do not rely on an egocentric coordinate system during demonstrative use; on the contrary they take listeners' cognitive status into account. For instance, Küntay & Özyürek (2006) argue Turkish demonstrative *şu* is used when the referent is not visually accessible to the addressee while the demonstrative *o* is used when the entity is in the addressee's visual focus of attention. After working on the two-term demonstrative system of Lao, Enfield (2003) comes to the conclusion that physical distance is not sufficient to discriminate between the terms and suggests demonstrative reference depends on how interlocutors perceive and interpret the physical world surrounding them during their conversation. Moreover, Piwek et al. (2008) argue Dutch speakers' choice of demonstratives is shaped by the cognitive status of the referent in the minds of interlocutors. Such studies challenge the traditional speaker-centered accounts of demonstrative reference and calls for situated and multimodal approach to demonstratives (Peeters & Özyürek, 2016).

2.5.3 Joint Attention

Understanding joint attention is crucial for reference research, especially demonstrative reference, since referring behavior is inextricably linked with it. Joint attention is certainly a complex phenomen, including an addresser, an addressee, and an object of reference (Diessel, 2006). For successful communication, both the speaker and the listener must focus their attention on the same entity. To achieve this common ground, the speaker directs the listener's focus of attention to a particular object in the surrounding situation. This may occur thanks to eye gaze, gestures, or body orientation in addition to language use. When both the speaker and the listener come to an agreement on the target object, they inform each other that they attend the same entity using eye gaze or gestures (Eilan et al., 2005).

As mentioned above, eye gaze and language use work together in the establishment of joint attention. The recent developments in eye tracking technology has made it possible to detect the context of ambiguous utterances (Henderson, 2003). The location where a speaker is looking at provides substantial information regarding his attention and intention in situated dialogue. Humans are capable of noticing their partners' face-directed gaze at a conversational distance (Pusch & Loomis, 2001). During everyday interaction, interlocutors generally follow each other's gaze and attend the entity under consideration. Moreover, Richardson & Dale (2005) find out communication turns out to be more successful if interlocutors achieve gaze alignment over the intended object. All these findings imply there is a strong relationship between eye gaze and reference generation, and eye gaze can be an effective cue during reference resolution.

2.5.4 Dual Eye Tracking

Cognition appears to be distributed between people and their environment (Hutchins, 1995a). The theory of distributed cognition goes beyond the individual and investigates the interaction between individuals and artifacts around them (Hollan et al., 2000). What distinguishes it from traditional theories is that it extends the understanding of the cognitive beyond a single individual and does not limit cognitive processes within a brain or a skull of an individual. People might work so closely for some time that a single cognitive system can arise (Hutchins, 1995b).

On the other hand, most eye tracking studies have been carried out with single participants performing certain tasks. The findings of such studies naturally do not reveal much about the real-life situations in which people collaborate to solve a problem. To overcome this limitation, researchers have come up with the technique of dual eye tracking to explore the socio-cognitive processes which characterize joint attention looking at eye gaze patterns of collaborators.

Dual eye tracking paradigm have enabled researchers to analyze the gaze patterns of participants while they are working on a problem together. Richardson & Dale (2005) have employed dual eye tracking to explore the gaze recurrence of pairs working on a shared scene. They aim to measure the degree and time course of gaze coupling during a joint attention task. They find out that on average it takes about 2 seconds for the addressee to decode the intended referent and shift his/her gaze over the associated location. Nüssli (2011) investigates the relationship between collaboration processes and eye movements and concludes cross recurrence analysis can prove to be a fruitful tool in joint attention studies.

2.6 Referring and Theories on Referring Expressions

Referring behavior has long been subject to scrutiny of many fields. Traditionally, philosophers, linguists, and psychologists have considered referring as if it were an addressee-blind act. Clark & Wilkes-Gibbs (1990) liken this model to writing to distant readers. That is, just uttering a referring expression is sufficient for the speaker to direct the listener's attention towards the referent. Upon hearing the expression, the listener can infer the intended entity successfully. They claim there are four assumptions of such a model:

1. The referring expression appears in one of three forms: as a proper noun, definite description, or a pronoun.
2. The speaker uses the expression to enable the listener to infer the referent.
3. Just uttering the expression is sufficient for the act to successfully occur.
4. The whole act is controlled by only the speaker.

These assumptions, however, are far from explaining the intricate nature of the act in a situated dialogue due to several reasons. Firstly, there is a natural interdependence between the speaker and the listener. They work together to establish joint attention in between. The listener needs to attend, hear, and understand the reference made by the speaker. If the process

fails somewhere, they cooperate to repair it. Moreover, reference is definitely multimodal. When observed, it is clearly seen that reference is generally accompanied by a gesture. Eye gaze, hand movements, and body orientations are cues listeners make use of to resolve the expressions.

There are many studies trying to explain the nature of referring expressions. However, two frameworks are prominent with their effort to propose a model that accounts for this phenomenon: *centering theory* and *givenness hierarchy*.

2.6.1 Centering Theory

Originally Grosz et al. (1983) claim that certain entities in an utterance are more central than others and this imposes certain constraints on the speaker's choice of using different types of referring expressions. They propose centering as a theory to explain the phenomenon. They develop a theory that relates focus of attention, choice of referring expression, and perceived coherence of utterances within a discourse segment. They come up with the notion of centers of an utterance to refer to entities linking a certain utterance to other utterances in a discourse segment. They are discourse constructs and semantic objects, not syntactic forms.

In their model, each utterance (U) in a discourse segment (DS) is assigned a single *backward-looking center* (Cb) and a set of *forward-looking centers* (Cf). The *backward-looking center* of the following utterance connects with one of the *forward-looking centers* of the previous utterance. The forward-looking centers are ranked according to their salience. The most salient element in the set of *forward-looking centers* is more likely to become *backward-looking center* of the next utterance. Gordon et al. (1993) suggests that this ranking can be determined by the syntactic role. They also define three types of transition relations across pairs of utterances: *center continuation*, *center retaining*, and *center shifting*. These transitions were applied by Friedrich & Palmer (2014) as in Figure 2.4.

	COHERENCE $CB(U_i) = CB(U_{i-1})$	¬ COHERENCE $CB(U_i) \neq CB(U_{i-1})$
SALIENCE $CB(U_i) = CP(U_i)$	CONTINUE	SMOOTH-SHIFT
¬ SALIENCE $CB(U_i) \neq CP(U_i)$	RETAIN	ROUGH-SHIFT

$CB(U_i) = undef.$	NOCB
$CB(U_{i-1}) = undef$ and $CB(U_i) = def.$	ESTABLISH

Figure 2.4: Definitions of transitions employed byFriedrich & Palmer (2014)

2.6.2 Givenness Hierarchy

In an attempt to address the question of how people understand referring expressions, although they are temporarily ambiguous most of the time. Gundel et al. (1993) propose a theoretical framework called Givenness hierarchy. Their main claim is languages have pronouns and determiners which encode memory and attention status of the intended referent along with their conventional semantics. Therefore, a speaker chooses the form of the referring expression he would use depending on this cognitive status in the listener's mind. There are six cognitive statuses regarding the forms of referring expressions used in discourse. Gundel et al. (1993) organize them in a scale as in Figure 2.5.

in focus >	activated >	familiar >	uniquely identifiable >	referential >	type identifiable
it	*IT/this/that/ this NP*	*that NP*	*the NP*	*indefinite this NP*	*a NP*

Figure 2.5: Givenness Hierarchy with related English referring expressions (Gundel et al., 1993)

2.6.3 Previous Corpora on Referring Expressions

As suggested earlier, situated dialogue has become increasingly important. This has led researchers to study referring expressions to help create more natural and efficient virtual interactions. To this end, there is a recent attempt to create corpora of referring expressions. For instance, Di Eugenio et al. (2000) constructed a corpus called COCONUT including referring expressions generated during a 2D interior design task. With the same purpose, the corpora QUAKE (Byron & Fosler-Lussier, 2006) and SCARE (Stoia et al., 2008) were created based on expressions used during a collaborative 3D task. Similarly, Spanger et al. (2009) designed a collaborative problem-solving task and built the REX J corpus including Japanese referring expressions (Spanger et al., 2012).

The related work on Turkish referring expressions mainly focuses on anaphora resolution in text. Say et al. (2002) created the METU corpus composed of 2 million words, with a morphologically and syntactically tagged subcorpus of 65,000words. Zeyrek et al. (2010) contributed to the METU corpus with a discourse subcorpus of 500,000 words. The relevant previous work on Turkish referring expressions involves pronoun resolution (Kılıçaslan et al., 2009) and reference generation (Yüksel & Bozsahin, 2002) in text using natural processing techniques. However, there is a recent attempt to build a corpus of Turkish referring expressions by Acartürk & Çakır (2012). They employed Tokunaga et al. (2010)'s experimental setup to collect the data and provided a critical analysis of referring expressions. The data analyzed in this thesis is obtained from Acartürk & Çakır (2012)'s work.

2.7 Summary of the Literature Review

This chapter outlines the current literature on demonstrative research. Firstly, it explains what demonstrative reference is regarding types of demonstratives and their pragmatic uses. There are basically three types of demonstratives: *nominal*, *local adverbial* and *verbal demonstratives*. Pragmatically, their use can be classified under two categories: *exophoric* and *endophoric*.

Secondly, the chapter gives a brief analysis of the Turkish demonstrative system along with the relevant previous work. These studies are mostly based on the traditional egocentric view of demonstrative reference which suggests a distance contrast. However, recent findings show demonstrative use is a joint activity involving different modalities. Moreover, they claim distance contrast is not physical, rather psychological. Obviously, there is a need to analyze demonstrative reference within a situated distributed framework.

Hence, this chapter also gives a brief account of two approaches to demonstrative analysis,

namely speaker-centric and social and multimodal, and then discusses why we should explore demonstrative reference from a collaborative perspective. After elaborating on the essential concepts of a situated dialogue study, the chapter introduces two important frameworks on referring expressions with the relevant previous studies on referring expressions.

The aforementioned literature clearly shows that the prevailing view based on physical distance fails to account for the intricate nature of demonstrative reference. It assumes an abstract coordinate system which is firmly grounded in the speaker's body and places the demonstrative terms according to the speaker's perspective. Moreover, the explanation neglects or completely ignores the addressee. The dominant view on the Turkish demonstrative system is based on this very idea. However, the recent studies indicate demonstrative reference is a joint activity in which interlocutors actively collaborate to establish a common ground. They also argue other modalities such as eye gaze body gestures are effective in shaping demonstrative use. Considering such findings, there is an obvious need to reexamine the Turkish demonstrative system within a social and multimodal understanding. Therefore, this study aims to investigate the use of Turkish demonstrative terms *bu*, *şu*, and *o* in the context of a situated dialogue, following the methodology described in the next chapter.

CHAPTER 3

METHODOLOGY

This chapter is mainly composed of two sections. The first section describes the experimental setup employed by Acartürk & Çakır (2012) to produce a corpus of Turkish referring expressions including who the participants were, what apparatus was used, and the procedure followed. The second section outlines the transcription and annotation process developed as part of this thesis work.

3.1 Experimental Setup

In an attempt to create a corpus of Turkish referring expressions, Acartürk & Çakır (2012) carried out an experiment that requires joint action of pairs to solve a geometrical puzzle. The basic design was the same as Spanger et al. (2009), which was later used by Tokunaga et al. (2010) to build a bilingual corpus. Kuriyama et al. (2011) made a minor modification to this setting by recording eye gaze of pairs in addition to their utterances simultaneously. This modification was also included in the experiment since eye gaze is at the heart of the study along with demonstrative use. They used tangram puzzles as the problem under consideration.

The tangram is believed to be the first puzzle in the world. It originated in China. There are many stories and even myths explaining the origin of the game. It is known as *Ch'i Chi'ao t'u* in Chinese, which literally translates to "seven boards of skill". It is a two dimensional rearrangement game. There are seven pieces formed by cutting a square into seven geometrical shapes, which are originally called tans. The pieces include two large triangles, a medium-size triangle, two small triangles, a parallelogram, and a square, as in Figure 3.1. The goal of the puzzle is to achieve a target shape generally presented as a silhouette problem by organizing pieces. The pieces are organized to form a rich variety of shapes ranging from people in motion to objects and animals (see Figure 3.4 for examples). The tangram still attracts many people all over the world because of its capacity to transform seven simple geometrical pieces into sophisticated and elegant figures.

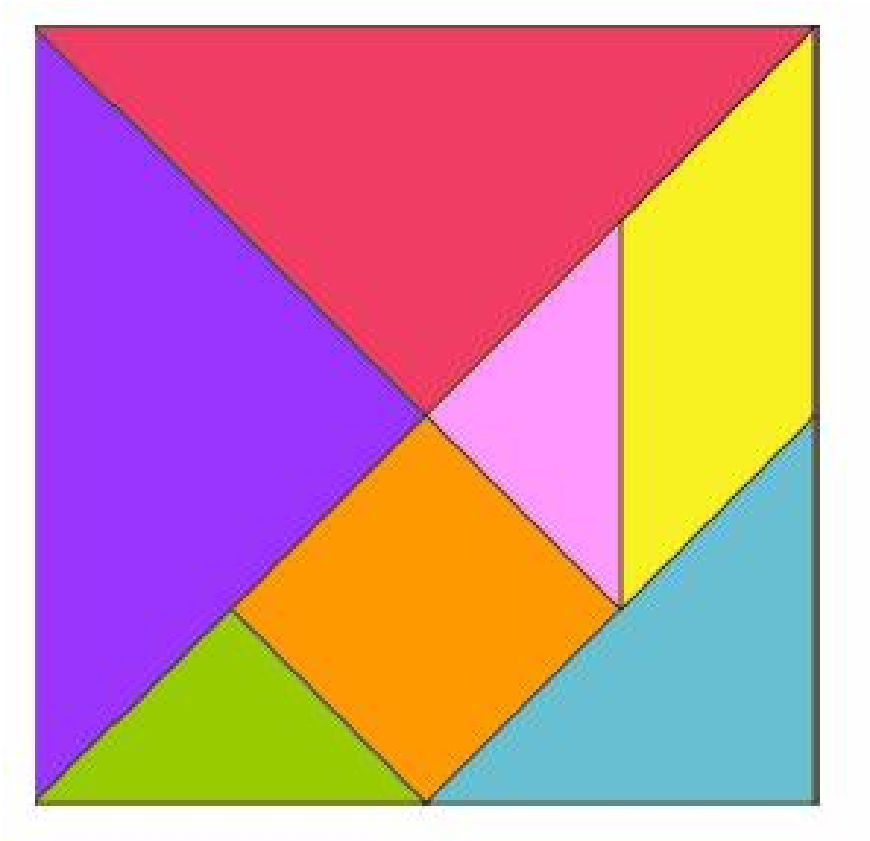


Figure 3.1: Tangram Pieces (Tans)

3.1.1 Participants

The participants were expected to know each other earlier and be the same gender. Therefore, four graduate students were recruited from the Informatics Institute, METU to form pairs. They were female and between 20 and 30 years of age. They had no problems with their vision and hearing. They had a training session to familiarize them with the software before the actual tasks and were assigned specific roles.

3.1.2 Apparatus

The pairs were exposed to the problem through a tangram simulator similar to the one used by Spanger et al. (2009) and Tokunaga et al. (2010). The simulator displays two distinct areas: a part for the target shape and another part for the working space, as in Figure 3.2. The working space enables the operator to move, flip, and rotate the pieces shown on the computer screen.

To record eye gaze of the pairs, two Eye Tribe eye-trackers with a sampling rate of 60 Hz which were mounted on two laptops with Intel Core i7-4510U processors. Each pair was provided with a microphone and a headset to converse during the sessions. A screen sharing software which also enables voice communication called Team Weaver (www.teamweaver.com)

was installed to laptops in order to coordinate the work of pairs. A desktop computer was used to record eye gaze and speech of the participants.

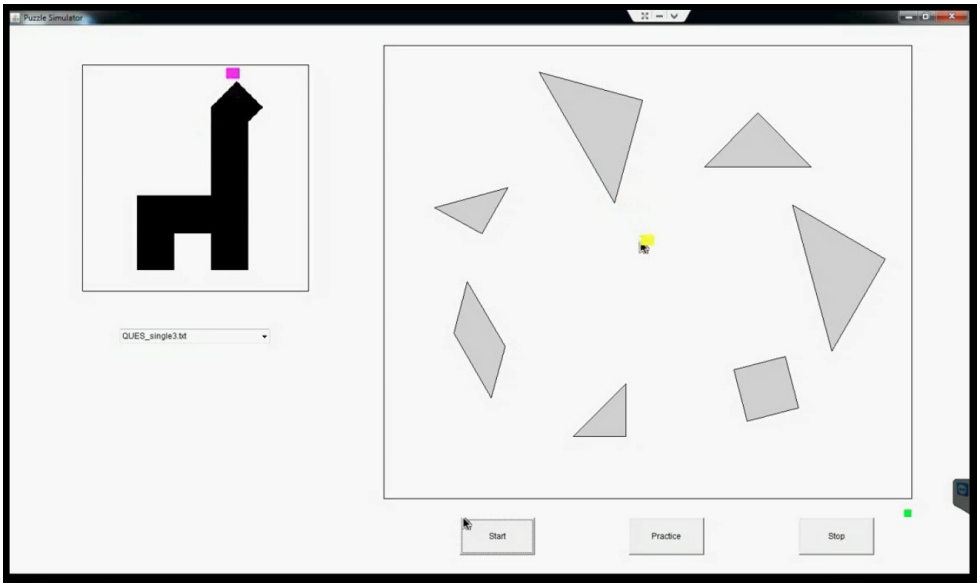


Figure 3.2: Tangram simulator

3.1.3 Procedure

The goal of the task:

Each pair was asked to solve puzzles on the tangram simulator. Their goal was to build the shape they were provided with arranging the seven pieces provided on the working space.

Role assignment:

Each participant of a pair was assigned a different role: presenter or operator. The presenter has access to both areas on the screen (the target shape and the working space), but does not have a mouse. The operator, on the other hand, can see only the working space and has a mouse to manipulate the pieces. This asymmetry is expected to encourage them to utter referring expressions.

Positioning the pair:

The presenter and the operator were situated back to back in a way that prevented them from seeing each other. These constraints of the experiment setting were designed to urge a natural need for communication between the presenter and the operator. The presenter had to provide the operator with necessary instructions and the operator ought to follow them to be able to manipulate pieces with the purpose of achieving the target shape (see Figure 3.3 for details).

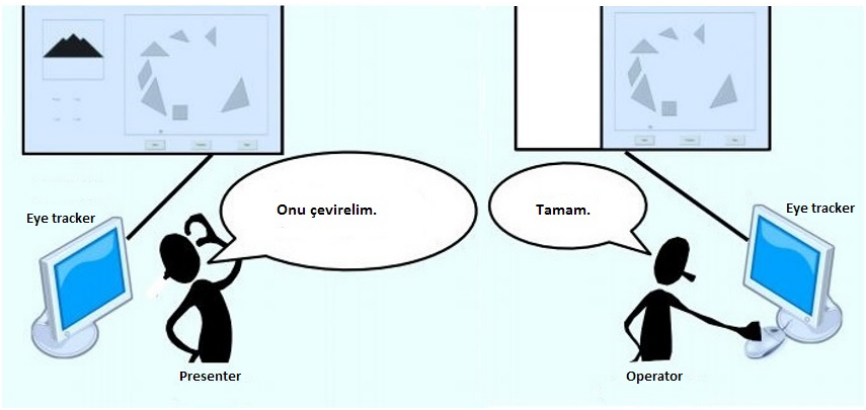


Figure 3.3: The design of the experiment

Target shapes:

Each pair was expected to complete 6 target shapes. The shapes resemble a swan (a), a chair (b), a fish (c), a mountain (d), a seal (e), and a vase (f) as in Figure 3.4.

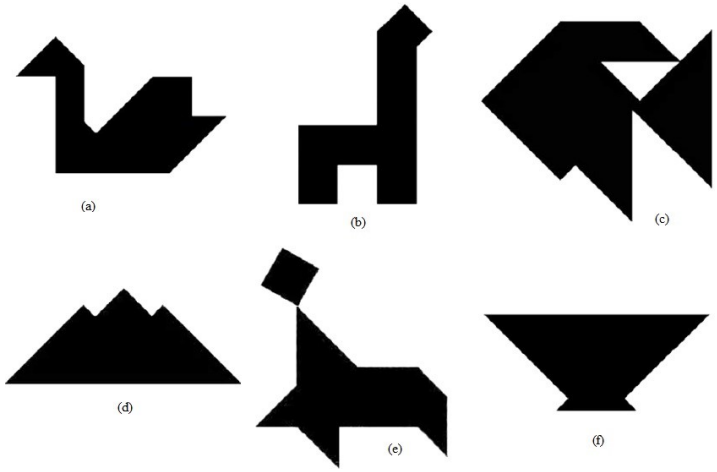


Figure 3.4: Tangram shapes

Conditions:

There were three conditions of the experiment in which the pairs were asked to achieve two of the shapes above. While one of the conditions provides the participants with no extra information, the other two offer extra cues, namely color and cue. The purpose of adding these cues to the task environment is to see whether they would have an effect on the demonstrative use of the pairs.

1. No-cue condition where the pieces were displayed in grey.

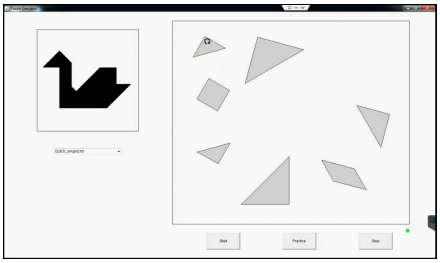


Figure 3.5: No-cue condition

2. Color-cue condition where each piece was displayed in a different color.

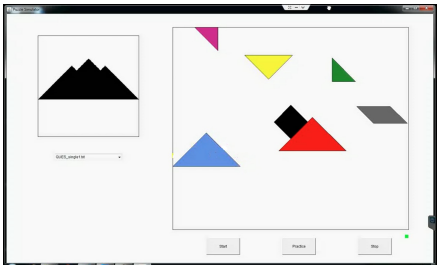


Figure 3.6: Color-cue condition

3. Gaze-cue condition where the participants were able to see where their partner's eye gaze falls.

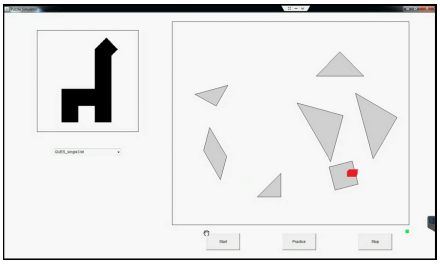


Figure 3.7: Gaze-cue condition

Role switching:

The participants changed their roles after two puzzles and the one who was to be the operator had some training to learn how to use the software and manipulate the mouse before starting the experiment session.

Location of the pieces:

The pieces were randomly located at the beginning of each session by the simulator. They could be moved, rotated and flipped by the operator.

Duration of the sessions:

The pairs were given 15 minutes to complete each puzzle per session and a hint was dropped every first and second five minutes so that the participants would not get stuck. The hint displays the correct location of a piece on the target shape, as in Figure 3.8. The session ended when the pairs achieved the task or the time was over. Eye gaze of the pairs were recorded through synchronized eye trackers and utterances through headset microphones.

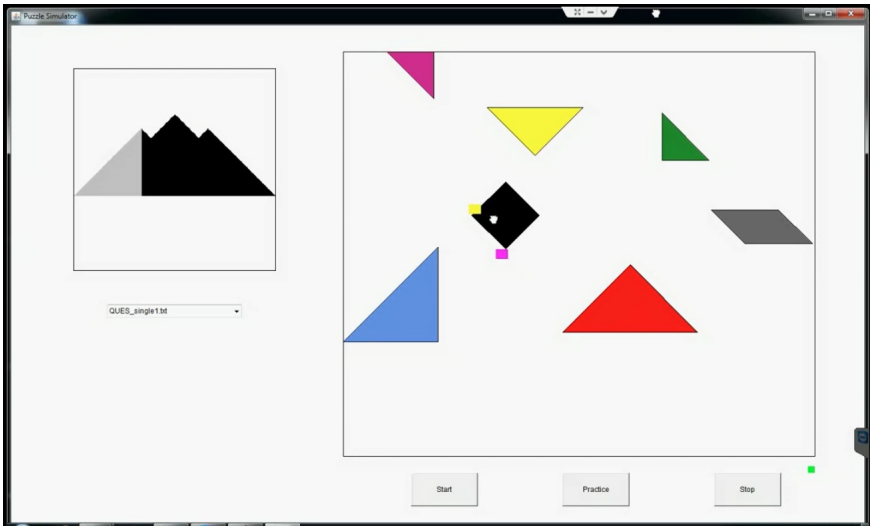


Figure 3.8: The simulator shows the correct location of a piece in the target shape.

The experiment is specifically designed to enable pairs to use as many referring expressions as possible to create a common ground to solve the puzzle. The roles presenter and operator they assume create a natural dependence on each other. Locating participants back to back eliminate the chance of making use of cues like pointing gestures and body orientations.

3.2 Transcription and Annotation Process

The data obtained from the participants were transcribed and then annotated according to the guideline.

3.2.1 Transcription

Out of the videos recorded for building the corpus of Turkish referring expressions, 12 of the videos (30% of the corpus) were selected in order to analyze the relationship between demonstratives and eye gaze. They were transcribed so that each demonstrative use could be identified. The videos were transcribed by two native speakers of Turkish who were familiar with the task. They carefully listened to the videos and write down the utterances simultaneously, paying attention to:

- The roles of the participants (presenter and operator)
- Turn-taking
- Pauses and backchannels

3.2.2 Annotation Scheme

Due to the multimodal and dynamic nature of the task, we needed to develop a novel annotation scheme. Our aim was to capture the relationship between demonstratives and eye gaze. Therefore, we determined a certain set of features and a guideline to guide us during the annotation process:

Feature Set:

- **The role of the speaker:**

  The participants were assigned either the role of *presenter* or *operator* during the task. The presenter was to give instructions to the operator since the target shape was only visible to this role. The operator was to manipulate the pieces following the instructions with the mouse. The asymmetry here led them to use demonstratives frequently.

  Recent findings suggest demonstrative reference is a joint enterprise, therefore we think that specific roles assigned to the participants might reveal a distribution which suggests a difference between roles. Therefore, we identified who uttered the demonstrative to whom during the task. Henceforth, the one who uttered the demonstrative is the *speaker* and to whom it was uttered is the *listener*. We looked at the role of the speaker to explore which role uttered which demonstrative and is it has a relation with the listener.

- **The type of demonstratives:**

  We firstly went over the speech data to isolate the demonstratives in the utterances. We spotted which one of the demonstratives (i.e. *bu*, *şu*, or *o*) was used in the utterances. The point to consider here is that which uses should be included in the data. As analyzed before, there are two main pragmatic uses of demonstratives: *exophoric* and *endophoric*. We are concerned only with the *exophoric* ones, with reference to outside the world:

  1. *Bu* doğru mu? (with a reference to one of the small triangles)
     "Is [this] correct?"
  2. *Şunu* mu döndürüyorum? (with a reference to one of the large triangles)
     "Am I rotating [that]?"
  3. *Onu* bir flip etsene. (with a reference to the parallelogram)
     "Flip [it] once."

  Such demonstrative uses in the utterances were all included in the data. However, *anaphoric*, *discourse deictic*, and *recognitional* uses were removed from the data during the annotation:

1. Presenter: Sehpa üzerinde *geniş kase gibi bir şey olur*.
   "There is something like a large bowl on the coffee table."
   Operator: Hı.
   "Ok."
   Presenter: *Onun* gibi böyle bir şey. (with a reference to the noun phrase)
   "Something like [it]."
2. Operator: *Ne yapalım*?
   "What shall we do?"
   Presenter: *Bunu* kestiremedim. (with a reference to the sentence)
   "I could not figure [this] out."

The demonstrative uses above were removed from the data since they did not have an exophoric reference. There are also other non-referential uses of demonstratives which function as a connective or an adverb. We removed them from the data, such as:

1. *şu* an (now)
2. *o* zaman (then), *bu*/*o* yüzden (therefore)
3. *bu* sefer/kez (this time)

- **Referents:**

  There are 7 geometrical pieces the speaker was expected to refer to. They include 2 large triangles, 1 medium-size triangle, 2 small triangles, 1 square and 1 parallelogram in the game. They were the target referents the speaker was expected to indicate to the listener, using a demonstrative. Also, the participants can refer to the target shape or the shape they have built using the pieces.

- **Attended by the speaker:**

  There appears a link between speech and eye gaze. The eye might look at the object being cognitively processed (Just & Carpenter, 1976). Hanna & Brennan (2007) suggest such fixations might accompany reference production. Therefore, we looked at whether the speaker visually attends the geometrical shape being referred to at the time of demonstrative use. Our criterion to decide on this is to look at whether the eye gaze falls on the referent or 100 pixels around it. When raw eye gaze data falls over or within 100 pixels around an object and stays there consistently, then we treat it as attention towards that object. In the eye tracking literature, this is a rather simplifying assumption since there is a more complex relationship between gaze location and attention. However, in this situated dialogue context this assumption is reasonable because the scene is a visual world composed of limited geometrical pieces located in a two dimensional environment, not a real setting with complex scenes.

  We employed 100-pixel criterion because the participants were using 17-inch flat screens with resolution 1024x768, and they were seated at a distance of 60-65 cm away from the screen. At this distance the screen would cover about 20 degrees of the visual field. The highest concentration of light sensitive cells in the human retina are clustered around a region called fovea and parafovea, which roughly covers 2 degrees of visual angle (Holmqvist et al., 2011). In other words, the highest visual acuity can cover only a small portion of the visual field, which is compensated by quick saccade and fixation movements by the visual system. Therefore, 2 degrees of visual angle corresponds to

about 1/10th of the screen, which can be approximated with a circle with radius 1024/10 ~100 pixels.

- **Recently attended by the pair:**

  The speaker's choice of demonstrative use differs during the conversation. Gundel et al. (1993) suggests the speaker depends on the memory and attention status of the referent in the listener's mind. The speaker selects the referring expression to be used based on this status. The presence or absence of joint attention seems to play an important role here. If there is an established joint attention towards a referent between the speaker and the listener, it constitutes a different cognitive status and therefore affects the speaker's choice of demonstrative.

  In line with this, we looked at whether the participants both visually attend the referent earlier or not just before a demonstrative utterance. Our aim is to explore whether an object is recently attended by the pair or not influences the demonstrative use. It might be a strong determinant of the demonstrative type that the speaker would use for the referent next time.

- **Gaze Overlap at the time of utterance:**

  The traditional account of demonstratives adopts an ego-centric explanation. The demonstrative system is firmly grounded in a coordinate system which is based on the speaker body and demonstrative terms encode physical distance accordingly. To see if distance has an effect on the speaker's demonstrative choice, we looked at the distance between the eye gaze locations of the speaker and the listener at the time of utterance. We used the same 100 pixels criterion we employed for the feature of *attended by the speaker*. We measured the distance between eye gaze locations at the time of demonstrative utterance with the screen ruler. If the distance is less than or equal to 100 pixels, we assume there is a gaze overlap between the participants. However, this pixel based measurement might yield low accuracy in distances because eye movements are very fast and pausing the video a few frames earlier or later might result in a considerable difference in our measurement.

- **Gaze distance at the time of utterance:**

  We also recorded the exact distance in pixels we measured for the distance between eye gaze locations when the speaker uttered the demonstrative.

- **Demonstrative resolution within 2 seconds after utterance:**

  When people have an interactive dialogue, they also synchronize their eye gaze and start to look at the same locations. Richardson et al. (2007) refer to this alignment as gaze coordination. They claim it is a measure of joint activity of language use. Richardson et al. (2007) also claim listeners' eye gaze follows speaker's fixations closely. Listeners are more likely to look at the object being referred to in 2 seconds. Hence, we investigated whether the listener looked at the intended referent within 2 seconds. Motivated by the same study, we assume the listener resolves the reference if the listener's eye gaze falls over or 100 pixels around the object within 2 seconds after demonstrative use. The eye gaze is itself usually ambiguous. However, since our domain is limited to a well defined two dimensional environment, it is reasonable to conclude the listener resolves the temporal ambiguity.

- **Demonstrative resolution duration:** We also measured how long it would take for the listener to disambiguate the referent after the speaker uttered the demonstrative. We

divided the duration 2 seconds it into 4 half seconds and recorded the demonstrative resolution time accordingly.

3.2.3 Annotation Tools

- **QuickTime 7.7.9 for Windows**

  QuickTime is a free multimedia software for handling video, sound, animation, graphics, text, interactivity, and music. It was used to play the video recordings of the sessions

- **MPEG Streamclip 1.2 for Windows**

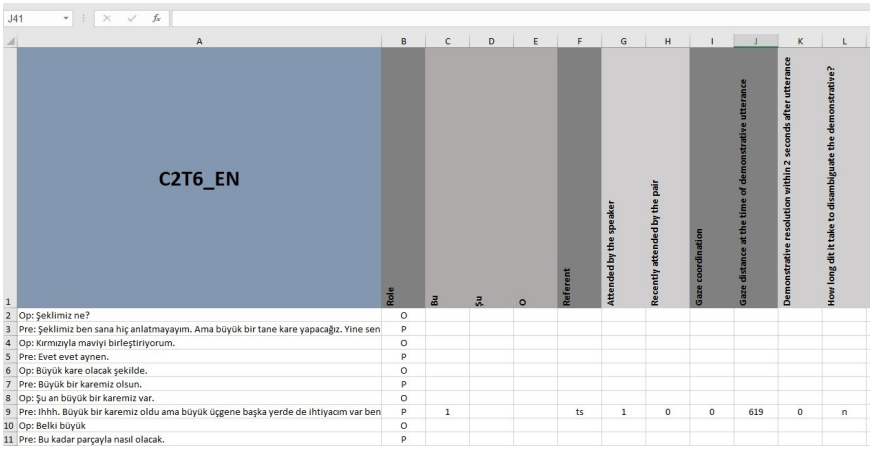
  MPEG Streamclip is a free video converter. It was used as an extension to Quicktime for measuring the time in milliseconds.

- **MB-Ruler - the triangular screen ruler 5.3 for Windows**

  MB-Ruler is a free screen ruler software for measuring distances and angles on the screen and distances on a map. It was used to measure the distance between the piece and the eye gaze of the speaker and the distance between the eye gazes of the participants

3.2.4 Annotation Environment

A laptop with an Intel Core i7-6500U 2.50 GHZ processor was used for annotation. The tools mentioned above were installed to the laptop. The transcript of each session was copied to a different Excel sheet. Each utterance in the transcript was copied to a different row and each feature to be annotated to a different column in the sheet, as in Figure3.9. The videos with the aligned eye gaze and speech data were transferred to the hard disk of the laptop. The video of the related transcript was opened with the Quicktime video player with the MPEG streamclip extension to display the time in milliseconds. The MB-ruler was ready to measure distance on the video in pixels when necessary.



	A	B	C	D	E	F	G	H	I	J	K	L
1	C2T6_EN	Role	Bu	Şu	O	Referent	Attended by the speaker	Recently attended by the pair	Gaze coordination	Gaze distance at the time of demonstrative utterance	Demonstrative resolution within 2 seconds after utterance	How long dit it take to disambiguate the demonstrative?
2	Op: Şeklimiz ne?	O										
3	Pre: Şeklimiz ben sana hiç anlatmayayım. Ama büyük bir tane kare yapacağız. Yine sen	P										
4	Op: Kırmızıyla maviyi birleştiriyorum.	O										
5	Pre: Evet evet aynen.	P										
6	Op: Büyük kare olacak şekilde.	O										
7	Pre: Büyük bir karemiz olsun.	P										
8	Op: Şu an büyük bir karemiz var.	O										
9	Pre: Ihhh. Büyük bir karemiz oldu ama büyük üçgene başka yerde de ihtiyacım var ben	P	1			ts	1	0	0	619	0	n
10	Op: Belki büyük	O										
11	Pre: Bu kadar parçayla nasıl olacak.	P										

Figure 3.9: Annotation sheet

3.2.5 Annotation Guideline

After starting the video, the annotator begins to carefully watch the video and listen to the participants. Whenever one of the pair utters a demonstrative, the annotator carries out the following steps in the same order:

1. The speaker is assigned either the role of *presenter* or *operator* during the session. The role of the speaker is coded as as "P" for the presenter or "O" for the operator.

2. The type of the demonstrative uttered by the speaker (i.e. *bu*, *şu*, or *o*) is entered into the related column as "1".

3. The referent is coded with *t1* and *t2* for the small triangles, *t3* for the medium-size triangle, *t4* and *t5* for the large triangles, *s* for the square, *p* for the parallelogram, *ts* for the target shape, and *rs* for the recent shape.

4. The distance between the eye gaze of the speaker and the referent is checked and if necessary measured with the MB-ruler. If the eye gaze falls over the referent or 100 pixels around it, it means the referent is attended by the speaker and coded with "1"; if not with "0" (see Figure3.10).

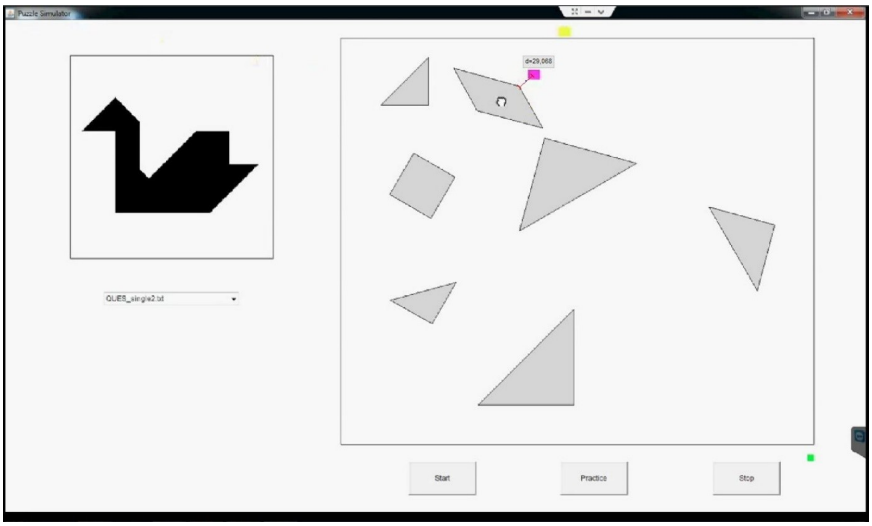


Figure 3.10: The distance is less than 100 pixels, therefore the object is attended by the speaker.

5. The piece being referred to with the demonstrative is recently attended by the pair if they look at it just before the demonstrative use and therefore it is still in their working memory and joint attention due to:

   - an early reference
   - operator's moving the piece with the mouse
   - the implicit agreement on the piece as a result of the common ground between the participants

and coded with "1"; if not with "0".

6. The distance between the eye gaze of the speaker and the listener at the time of the utterance is measured with the MB-ruler. If the distance is less than or equal to 100 pixels, it means there is gaze overlap between the participants and coded with "1"; if not with "0" (see Figure3.11).

7. The gaze distance is also recorded in pixel (see Figure3.11).

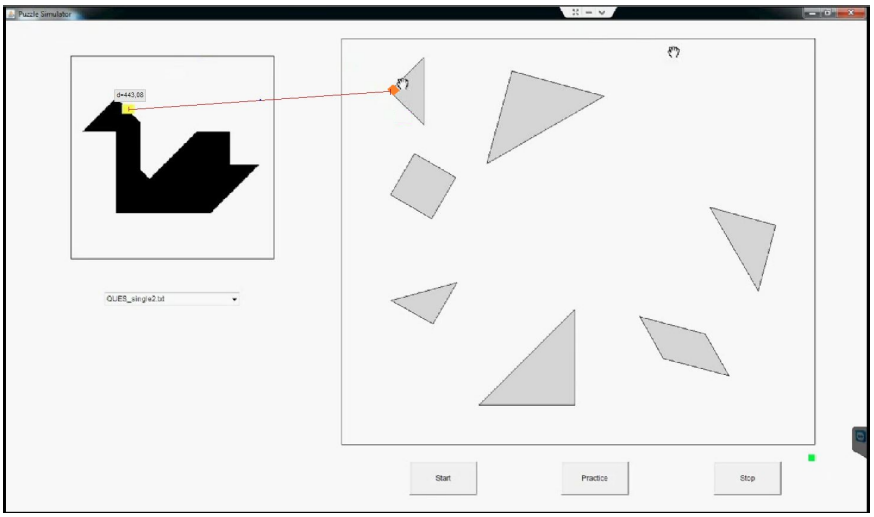


Figure 3.11: The eye gaze distance between participants

8. The listener's eye gaze is carefully observed with the help of the MPEG streamclip. If the listener's eye gaze follows the speaker's fixation and falls over the target referent within 2 seconds, it means the listener successfully disambiguates the demonstrative and coded with "1"; if not with "0" (see Figure3.12).

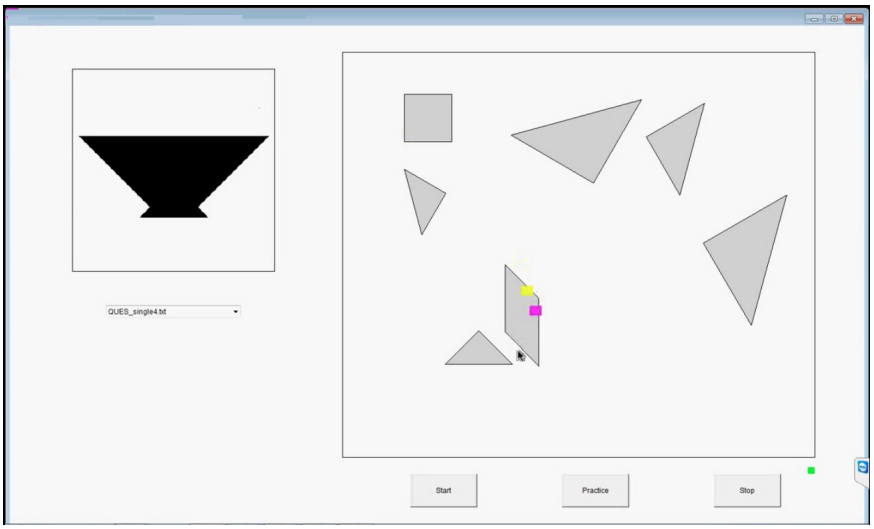


Figure 3.12: The yellow eye gaze indicates demonstrative resolution by the listener.

9. The resolution time is also measured in milliseconds with the help of the MPEG stream-clip. 2 seconds is divided into 4 half seconds. If the listener is able to disambiguate the reference, the time is coded as "0" (at the time of utterance), "0,5" (in half a second), "1" (in a second), "1,5" (in half one and half a second), "2" (in two seconds).

3.2.6 Inter-rater Reliability Analysis

To assess the reliability of the guideline, two independent raters annotated a video transcript for inter-rater reliability analysis. To this end, one of the twelve sessions were randomly selected. It included 150 utterances.

The annotators were both native speakers of Turkish and familiar with the task. They used the tools mentioned above for the annotation and independently annotated the utterances one by one, following the steps described in the guideline carefully.

CHAPTER 4

RESULTS

This chapter firstly outlines the results of the inter-rater reliability analysis regarding the annotation process and then gives a general overview of the data and reports the results of the statistical tests which investigate the relationship between demonstratives and eye gaze.

4.1 The Results of the Inter-rater Reliability Analysis

The reliability of the annotation procedure was tested by computing the Krippendorf α values for each dimension of the annotation scheme (Krippendorff, 2004). The KALPHA script was used in SPSS for the computations (Hayes & Krippendorff, 2007). Two annotators independently annotated a randomly selected video transcript that included 150 utterances.

Both annotators identified 8 *bu*, 10 *şu*, and 42 *o* instances. Only in one case there was disagreement, where one of the raters missed a *şu* instance. The coders were also in almost full agreement for the unannotated utterances (89 out of 90), where those utterances either do not include the tokens *bu*, *şu*, *o* or a deictic use of these demonstratives. The high degree of overlap naturally led to a high Krippendorf α of .98. Therefore, we conclude that the identification of the demonstratives can be consistently performed.

The two raters also tried to decode each demonstrative they identified by noting which puzzle piece was the intended referent. In this case, only those utterances that were identified as including a demonstrative were considered. When their referent assignments were compared a Krippendorf α of .82 was observed, which is higher than the recommended threshold of .70, indicating that the referents can be reliably identified.

Since other dimensions of the annotation scheme focused on different properties of the demonstratives, the rest of the dimensions were only compared for those utterances where a demonstrative was highlighted by the raters. The table 4.1 summarizes the Krippendorf α values obtained for each dimension.

Table 4.1: Krippendorf's α for each dimension in the annotation scheme

Dimension	**Krippendorf's α**
The Type of the Demonstrative (bu, şu, o)	.98
Intended Referent	.82
Attended by the Speaker	.75
Recently Attended by the Pair	.82
Gaze Overlap (at the time of demonstrative use)	.85
Demonstrative Resolution (within 2 seconds after demonstrative use)	.71
Gaze Distance (at the time of demonstrative use)	.63
Demonstrative Resolution Duration	.76

Satisfactory reliability was obtained over all dimensions except the gaze distance where the coders had to pause the video player at the time the demonstrative was uttered and measured the distance between the gaze location and the center of the intended object. Since most referring expressions were uttered by the presenters as they were moving their gaze back and forth between the workspace and the target, although raters often paused the video within a few frames, there were considerable differences in the estimated distance reported by both participants as they captured different portions of a saccade. This had a negative impact on the inter-rater reliability value obtained for this dimension. Using a pixel based, ratio level difference scale could be considered as an overly strict test, and one could consider computing the reliability after the gaze distance is converted into ordinal categories such as low, medium, high. We decided to report the more strict computation to highlight this potential difficulty.

4.2 Overview

A total of 1256 utterances obtained from the Turkish referring expressions corpus was annotated in this study. Among these 1256 utterances, 587 of them included the use of the demonstratives *bu*, *şu* and *o*. Because the task included a collaborative tangram puzzle solving task mediated by a shared task space (i.e. no face to face contact among the collaborators), the participants frequently relied on demonstratives to direct each others' attention on task relevant pieces and/or locations in the shared task space. Among these 587 instances, there were 140 *bu* cases (23.9%), 119 *şu* cases (20.3%), and 328 *o* cases (55.9%).

4.3 The Influence of Demonstrative Use and Speaker's Role

During the experiment, the participants were either in the role of the presenter or the operator. The presenters could see the entire workspace including the target shape, but could not control anything on the screen, whereas the operators could control the movement of the pieces but could only see the part of the screen where the pieces move around (i.e. operators cannot see the target shape). In order to see if the speaker's role had played any significant role on the distribution of *bu*, *şu*, and *o* cases, we conducted a cross tabulation of the demonstrative types with the role of the speaker who uttered the demonstrative (see Table 4.2).

Table 4.2: Cross tabulation of demonstrative types and the speaker's role

			Role		
			Operator	Presenter	TOTAL
Demonstrative	Bu	Count	52	88	140
		% of Total	8.9%	15%	23.9%
	O	Count	9	319	328
		% of Total	1.5%	54.3%	55.9%
	Şu	Count	74	45	119
		% of Total	12.6%	7.7%	20.3%
Total		Count	135	452	587
		% of Total	23%	77%	100 %

A loglinear analysis was conducted to test the main effects of role and demonstrative types as well as their interaction on the observed frequency distribution. The results indicate a strong main effect of role, $\chi^2(1)=180.7$, p<.001 and demonstrative type, $\chi^2(2)=126.8$, p<.001. In other words, presenters (77%) tended to use significantly more demonstratives than the operators (23%), and there is a significant difference in the frequency of use of *o* as compared to *bu* and *şu*. There is also significant interaction, $\chi^2(1)=208.1$, p<.001, which is due to the difference between presenters and speakers in terms of their frequency of using *o*. Presenters used *o* much more frequently (54%) as compared to operators (1.5%). Moreover, operators used *şu* (12.6%) more frequently than presenters (7.7%).

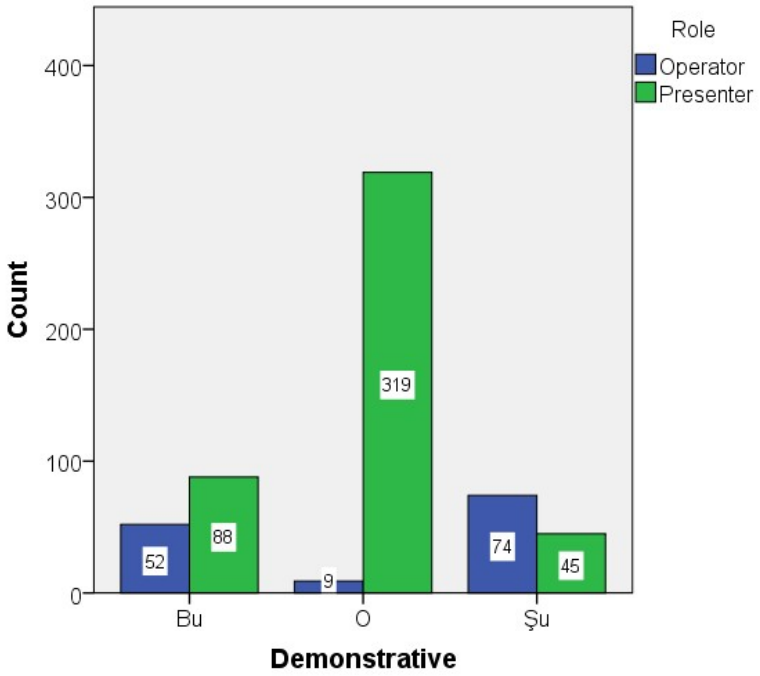


Figure 4.1: Frequency distribution of Turkish demonstratives *bu*, *şu*, and *o*

4.4 The Influence of Role, Demonstrative Use and Visual Condition

The utterances were collected during 3 different conditions. In the first condition all puzzle pieces were provided in the same gray color. In the second condition, each of the 7 Tangram pieces were displayed in a different color. Finally, in the third condition, the gaze of the partner was provided on the participant's screen continuously as a gaze cue.

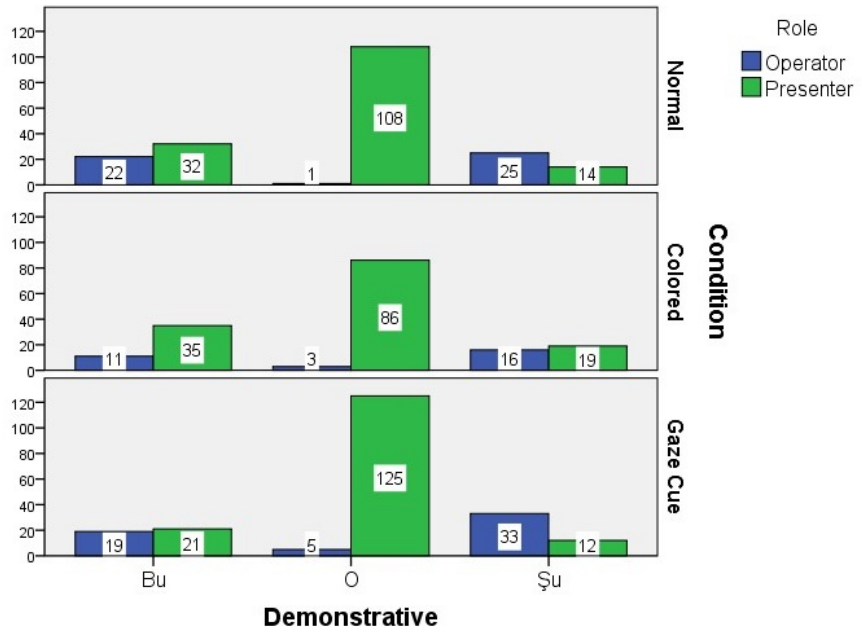


Figure 4.2: Frequency distribution of demonstratives *bu*, *şu*, *o* across visual cue conditions and role

A three-way loglinear analysis was conducted to explore the mutual effects of condition, role and the demonstrative type on the observed frequency distribution. The resulting model indicated significant main effect of demonstrative type, role and their interaction, as it was discussed in the previous analysis.

The main effect of condition was marginally significant, $\chi^2(2)=5.58$, p=.061, which is possibly due to the slight decrease in the total number of demonstratives *bu*, *şu*, and *o* in the color-cue condition (N=170) as opposed to no-cue (N=202) and gaze-cue conditions (N=215). The two-way interaction between demonstrative types and condition was also significant, $\chi^2(4)=12.12$, p<.05, which is due to the decrease in the use of *o* during the color condition as compared to the other two conditions.

The interaction of role and condition was also significant, $\chi^2(2)=11.04$, p<.01, which seems to be due to the increase in the operators' use of the term *şu* in the gaze-cue condition, and the increase in the use of *bu* by the presenters in the color version.

4.5 The Relationship between Role, Demonstrative Use and Gaze Allocation

We also annotated whether the speakers' gaze fell within a radius of 100 pixels around the intended object (i.e. the specific tangram piece that we inferred from the context) when they referred to it by using one of the demonstratives of interest. We assumed that the speaker visually attended to the referred object when this criterion was met.

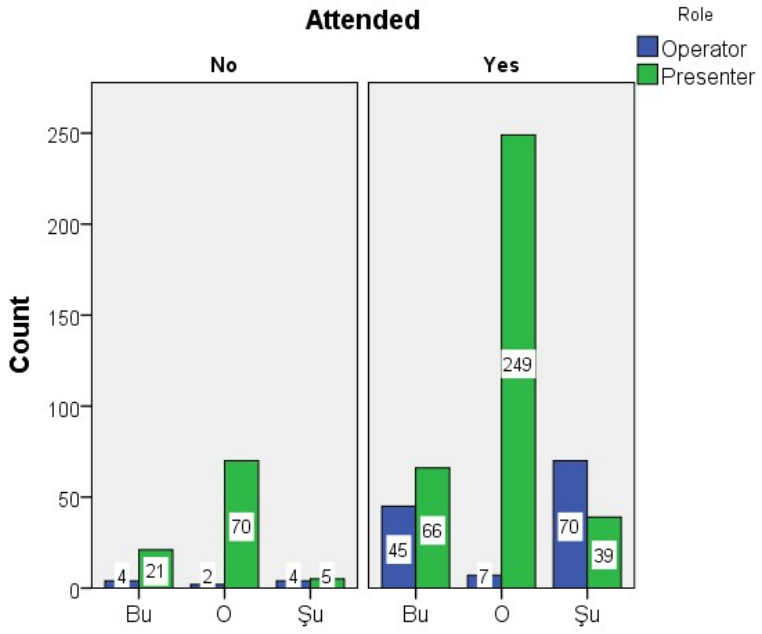


Figure 4.3: Frequency distribution of demonstratives *bu*, *şu*, *o* across role and attention categories

A three-way loglinear analysis was conducted to explore the mutual effects of attention, role and the demonstrative type on the observed frequency distribution. The resulting model indicated significant main effect of demonstrative type, role and their interaction, as it was discussed in the previous analysis.

The main effect of attention was significant, $\chi^2(2)=254.38$, $p<.001$, which is demonstrated in Figure 4.3, where the frequency distribution suggests that participants tended to attend to the object they were referring to for all demonstrative types. The interaction between attention and role was also significant, $\chi^2(1)=5.72$, $p<.05$, which is indicated by the 70 *o* and 21 *bu* cases which were uttered by the presenter without looking at the intended object as detected by the eye tracker. As for the operator, almost all instances of demonstrative use include a glance close to the referred object. The interaction between attention and demonstrative type, and the three-way interaction did not reach significance.

4.6 The Relationship between Role, Demonstrative Use and Recently Attended Objects

Another feature we annotated kept track of the list of recently attended objects by the pair in the unfolding dialogue. This feature was used to explore whether the demonstrative types and roles differ in terms of whether the intended referent is an already attended/discussed object or whether the expression brings forth a new referent into the discourse.

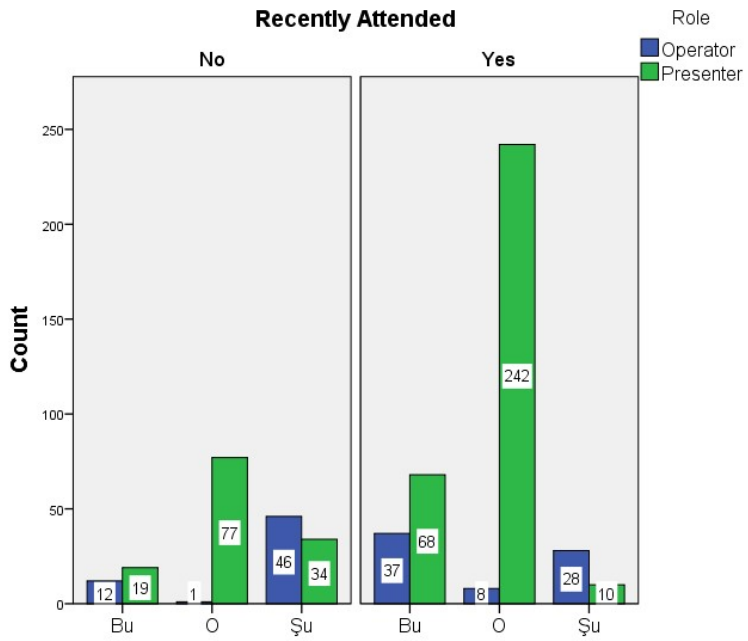


Figure 4.4: Frequency distribution of demonstratives *bu*, *şu*, *o* across role and recently attended categories

A three-way loglinear analysis was conducted to explore the mutual effects of role, the demonstrative type and whether the referent was already an attended object on the observed frequency distribution. The resulting model indicated significant main effect of demonstrative type, role and their interaction, as it was discussed in the previous analysis.

The main effect of recent attention was significant, $\chi^2(1)=73.05$, $p<.001$, which is demonstrated in Figure 4.4, where the frequency distribution suggests that participants tended to refer to the objects that they had already referred to in the recent dialogue for all demonstrative types. The interaction between recent attention and demonstrative type was also significant, $\chi^2(2)=197.39$, $p<.001$, which is indicated by the more frequent use of *bu* and *o* for recently attended objects, whereas the reverse pattern was observed in the *şu* case, where participants tended to use *şu* to introduce a new referent into the discourse. The interaction between recent attention and role, and the three-way interaction did not reach significance.

4.7 The Relationship between Role, Demonstrative Use and Gaze Overlap

The annotation scheme also marked whether the speaker and the addressee were dwelling on the same location when one of the demonstratives was uttered. A three-way loglinear analysis was conducted to explore the mutual effects of role, the demonstrative type and whether eye gaze of both partners were overlapping when the utterance were made on the observed frequency distribution. The resulting model indicated significant main effect of demonstrative type, role and their interaction, as it was discussed in the previous analysis. The main effect of gaze overlap was not significant, which reflects the almost identical frequency distributions corresponding to yes and no cases for initial gaze overlap. Moreover, the 2-way interactions between gaze overlap and demonstrative type as well as gaze overlap and role did not reach significance. Therefore, the initial gaze overlap at the time of an utterance including one of the demonstratives *bu*, *şu* and *o* did not have a strong enough influence on the distribution of role and demonstrative cases.

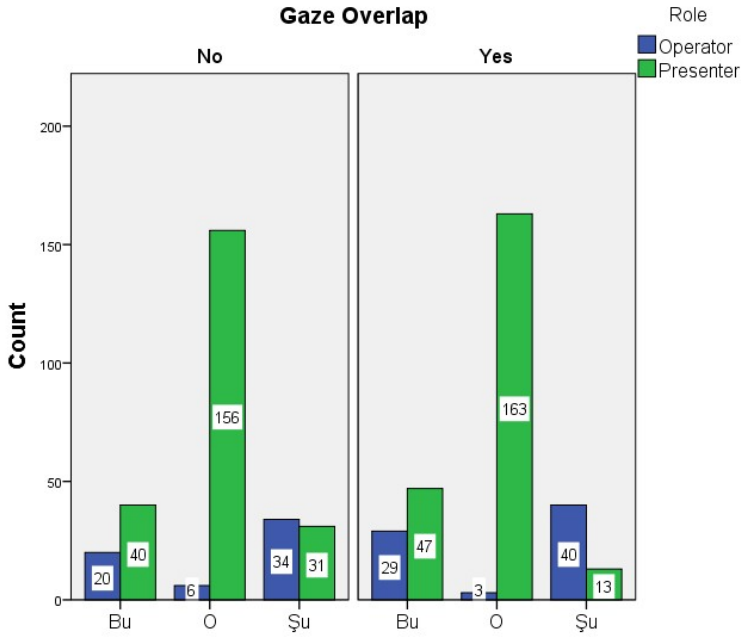


Figure 4.5: Frequency distribution of demonstratives *bu*, *şu*, *o* across role and gaze overlap categories

4.8 The Relationship between Role, Demonstrative Use and Demonstrative Resolution

During the data analysis, we also investigated whether a referring expression use led to an eventual overlap or coordination of eye movements. To this end, we considered a time-window of 2 seconds, which is motivated by related research (Richardson & Dale, 2005). We assumed the reference was resolved if the listener looked at the intended object and/or location within 2 seconds following the speaker's demonstrative utterance.

A three-way loglinear analysis was conducted to explore the mutual effects of role, the demonstrative type and whether the listener looked at the intended object/location within 2 seconds on the observed frequency distribution. The resulting model indicated significant main effect of demonstrative type, role and their interaction, as it was discussed in the previous analysis.

The main effect of gaze coordination was significant, $\chi^2(1)=367.83$, p<.001, which is demonstrated in Figure 4.6 where the frequency distribution suggests that participants tended to exhibit gaze coordination within 2 seconds for all types of demonstratives. The interaction between gaze coordination and demonstrative types did not reach significance. A marginally significant effect was observed for the interaction between gaze coordination and role, $\chi^2(2)$ =342.39, p=.064, which seems to be due to the increase in gaze coordination for the use of *şu* for the operators as compared to presenters. Finally, the three-way interaction did not reach significance.

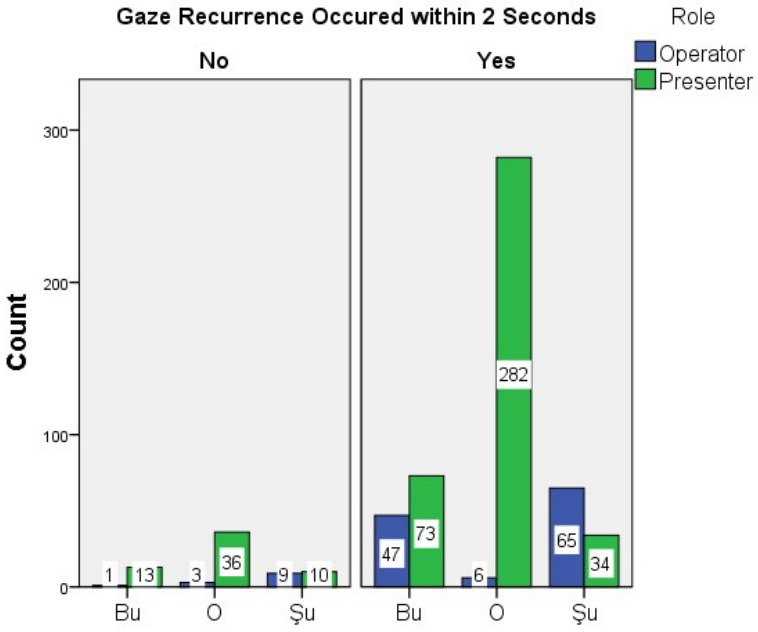


Figure 4.6: Frequency distribution of demonstratives *bu*, *şu*, *o* across role and demonstrative resolution categories

4.9 The Effect of Demonstrative Type, Role and Condition on Gaze Distance

We also measured the distance between the gaze positions of the two partners when the speaker made an utterance including *bu*, *şu*, and *o*. In order to explore whether this distance was influenced by the demonstrative type, visual cue condition and the role of the speaker, we conducted a 3-way ANOVA with initial gaze distance as the dependent variable and demonstrative type, condition and role as the independent variables. Table 4.3 suggests that none of the effects are significant for these independent variables.

Table 4.3: ANOVA results table with role, demonstrative and condition as the independent variables

Source	Type III SS	df	Mean Sq	F	Sig.	Partial η^2
Corrected Model	1234720.264a	17	72630.604	1.002	0.454	0.029
Intercept	9613592.378	1	9613592.378	132.603	0	0.19
Role	11189.243	1	11189.243	0.154	0.695	0
Demonstrative (Dem)	3035.295	2	1517.648	0.021	0.979	0
Condition (Cond)	158863.58	2	79431.79	1.096	0.335	0.004
Role * Dem	168693.883	2	84346.941	1.163	0.313	0.004
Role * Cond	38725.918	2	19362.959	0.267	0.766	0.001
Dem * Cond	427645.033	4	106911.258	1.475	0.208	0.01
Role * Dem * Cond	215628.932	4	53907.233	0.744	0.563	0.005
Error	40889539.31	564	72499.183			
Total	76858687	582				
Corrected Total	42124259.58	581				

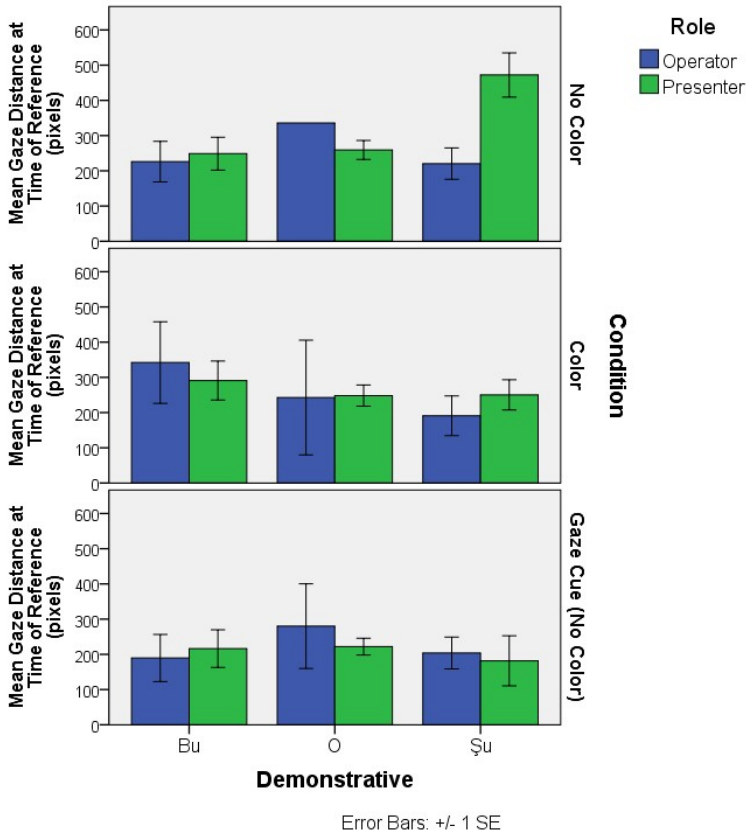


Figure 4.7: Mean gaze distance at the time of reference for each demonstrative type, condition and role

The annotation scheme included whether the speaker attended to the referent (as indicated by the gaze coordinates in the vicinity of the intended object) and also whether the intended referent was to a recently referred object. A 2-way ANOVA including these two binary variables as independent variables and mean gaze distance at time of reference as the dependent

variable found a significant effect of attended, F(1,578)=102.83, p<.001, partial η^2=.51, and significant interaction effect, F(1, 578)=32.12, p<.001, partial η^2=.05. Together with Figure 4.8, this result suggests the the distance between the gaze coordinates are minimal when the speaker attended to the referred object, and when that object was a recently attended object.

Table 4.4: Dependent Variable: Gaze Distance at Time of Reference (pixels)

Source	Type III SS	df	Mean Sq	F	Sig.	Partial η^2
Corrected Model	12976740.940a	3	4325580.312	85.777	0	0.308
Intercept	36496828.4	1	36496828.4	723.738	0	0.556
Attended (Attend)	5185427.169	1	5185427.169	102.828	0	0.151
Recently Attended	151463.788	1	151463.788	3.004	0.084	0.005
Attend * Recently Attend	1619675.5	1	1619675.5	32.118	0	0.053
Error	29147518.64	578	50428.233			
Total	76858687	582				
Corrected Total	42124259.58	581				

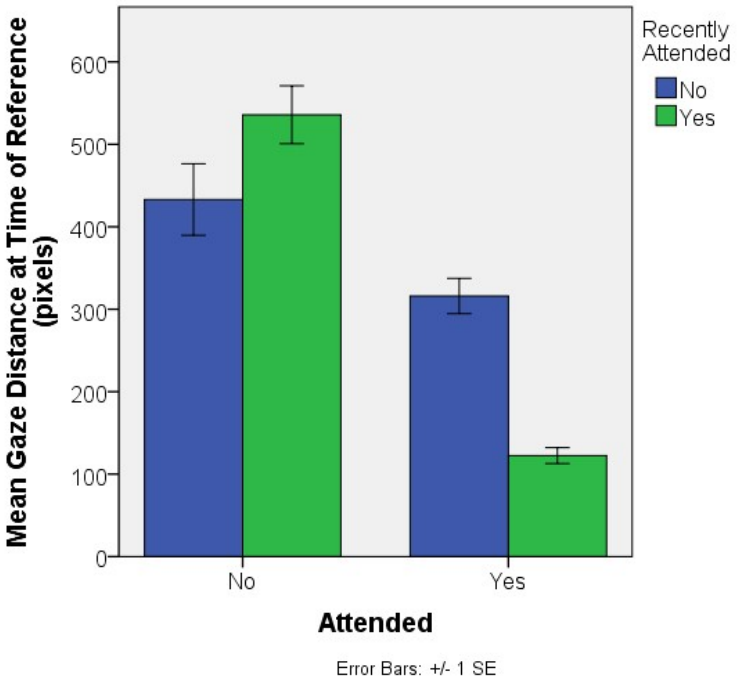


Figure 4.8: Mean gaze distance between eye gaze pozitions

4.10 The Effect of Demonstrative Type, Role and Condition on Demonstrative Resolution Duration

In order to explore whether the time it takes for the listener to gaze over the intended object (within a radius of 100 pixels) after an utterance including *bu*, *şu*, and *o* was influenced by the demonstrative type, visual cue condition and the role of the speaker, we conducted a 3-way ANOVA with the time it took for the listener to gaze over the intended object as the

dependent variable and demonstrative type, condition and role as the independent variables. Table 4.5 suggests that there was a significant effect of role and demonstrative interaction, F(2,491)=6.46, p<.01, partial η^2=.02, as well as role and condition interaction F(2,491)=3.28, p<.05, partial η^2=.01. When the mean time values observed for each demonstrative type, speaker role and condition are observed in Figure 4.9, it is evident that the observed interaction effects are due to the extra effort needed to converge on a same location when the presenter uttered a *şu* demonstrative, but this cost was diminished when the partners were able to see each other's gaze pointers as a cue over the shared space. This feature seemed to help recipients to decode the intended referent more easily. However, these significant effects should be cautiously interpreted as the effect sizes are rather low.

Table 4.5: Dependent Variable: Demonstrative Resolution Time

Source	Type III SS	df	Mean Sq	F	Sig.	Partial η^2
Corrected Model	7.342a	16	0.459	2.468	0.001	0.074
Intercept	11.486	1	11.486	61.762	0	0.112
Role	0.001	1	0.001	0.007	0.934	0
Demonstrative (Dem)	0.942	2	0.471	2.532	0.08	0.01
Condition (Cond)	0.471	2	0.235	1.266	0.283	0.005
Role * Dem	2.404	2	1.202	6.464	0.002	0.026
Role * Cond	1.221	2	0.611	3.284	0.038	0.013
Dem * Cond	1.432	4	0.358	1.925	0.105	0.015
Role * Dem * Cond	1.32	3	0.44	2.366	0.07	0.014
Error	91.311	491	0.186			
Total	127	508				
Corrected Total	98.654	507				

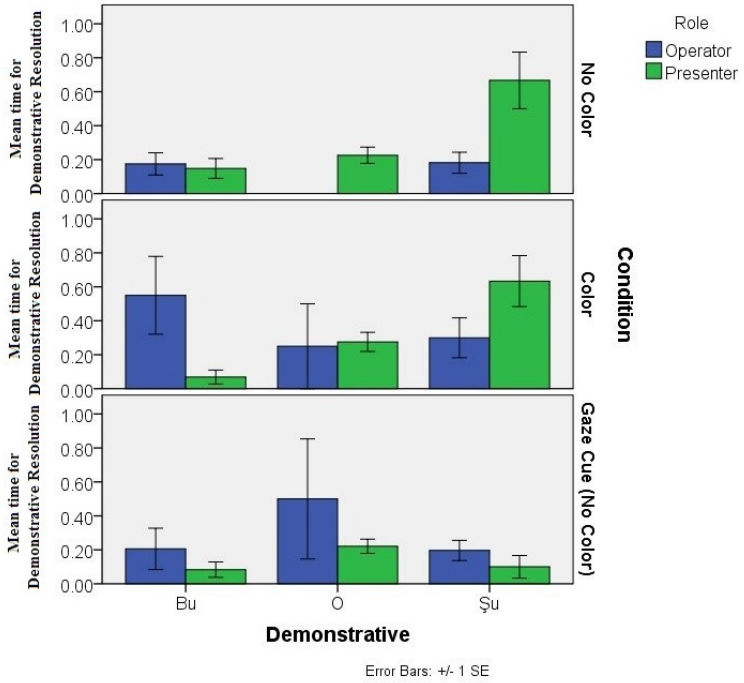


Figure 4.9: Mean demonstrative resolution time according to conditions

We also investigated whether the speaker attended to the referent (as indicated by the gaze coordinates in the vicinity of the intended object) and also whether the intended referent was to a recently referred object. A 2-way ANOVA including these two binary variables as independent variables and mean time as the dependent variable found a significant effect of attended, F(1, 79.36)=26.17, p<.001, partial η^2=.05, significant effect of recently attended, F(1, 79.36)=83.27, p<.001, partial η^2=.14, and significant interaction effect, F(1, 79.36)=32.12, p<.01, partial η^2=.02. Together with Figure 4.10, this result suggests the distance between the gaze coordinates are minimal when the speaker had attended to the referred object, and when that object was a recently attended object. Together with the bar chart these results suggest that if a referent is towards an already referred object, participants tend to converge on the same object more quickly, when this object is initially attended by the speaker it takes shorter to converge and finally there is a significant decrease in time for recently attended cases depending on whether the referent is attended by the speaker.

Table 4.6: Dependent Variable: Demonstrative Resolution Time

Source	Type III SS	df	Mean Sq	F	Sig.	Partial η^2
Corrected Model	19.299a	3	6.433	40.857	0	0.196
Intercept	38.844	1	38.844	246.705	0	0.329
Attended	4.12	1	4.12	26.167	0	0.049
RecentlyAttended	13.111	1	13.111	83.274	0	0.142
Attended * RecentlyAttended	1.274	1	1.274	8.094	0.005	0.016
Error	79.355	504	0.157			
Total	127	508				
Corrected Total	98.654	507				

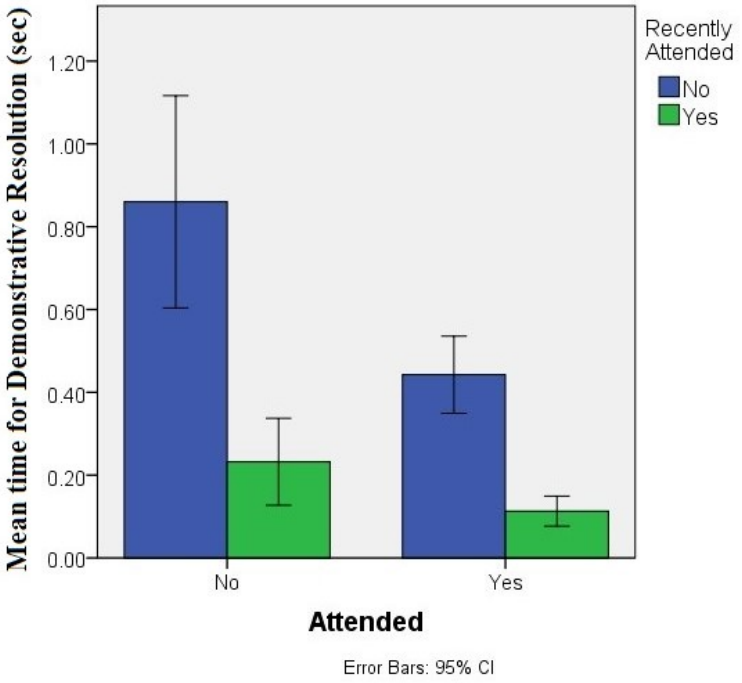


Figure 4.10: Mean time for demonstrative resolution

CHAPTER 5

DISCUSSION

This chapter elaborates on the results presented in the previous chapter and discusses them in detail in relation with the research questions below:

- Do demonstrative expressions differ in terms of their frequency distribution and functional use in a situated dialogue?
- Do demonstrative expressions differ in terms of the role of the participants?
- Do demonstrative expressions differ in terms of cue conditions?
- Can eye gaze be used as a visual cue to resolve demonstrative reference?
- Is there a temporal relation between eye gaze and demonstrative use?
- Do demonstrative expressions differ in terms of the presence of joint attention?

Contrary to the traditional egocentric account, we aim to look at demonstratives within a social and multimodal perspective. We analyzed a corpus of Turkish situated dialogues where pairs of participants were required to collaborate while solving a tangram puzzle through networked computers. Considering reference as a multimodal enterprise, the corpus included the recordings of eye gaze along with speech. Both modalities were aligned in a video format. The task design employed to build the corpus we analyzed was a computer-mediated environment. Located back to back, the partners collaborated over a shared screen through a software. The environment the puzzles were presented was a two dimensional and well defined problem space. Naturally, it did not reflect the complexity of a visual setting of a real life setting. However, our findings are still significant since they have the potential to offer a base for multimodal reference resolution studies focusing on eye gaze as a visual cue in Turkish. Also, similar tasks have been used to produce aforementioned corpora of naturalistically uttered referring expressions, which reveal essential findings for multimodal interfaces.

After careful transcription and annotation of approximately 30% of the corpus, we identified 587 instances of demonstratives *bu*, *şu*, and *o* in 1256 utterances. Overall, there are 140 *bu* (23.9%), 119 *şu* (20.3%),and 328 *o* (55.9%) cases. Therefore, the distribution of demonstratives in our study differs from the studies investigating demonstratives in text (Çallı, 2012) and in conversational data (Küntay & Özyürek, 2006). While *bu* is the most frequent demonstrative in Çallı (2012)'s analysis, Küntay & Özyürek (2006) find out adults use *şu* more

frequently in coversational data. However, Çallı also explores the exophoric uses in the corpus and finds out *o* is the most frequent demonstrative with a reference outside the text, which is in line with our finding.

We investigated whether the role has an influence on the type of demonstrative speakers choose to use. The results show that the role they assume has a strong effect on the type they prefer. While presenters tend to use *o* more than *bu* and *şu*, operators use *şu* more than other demonstratives. The reason behind this distribution could be the nature of the task. The roles the participants assume is a determinant of who is leading the task since the target shape is only visible to the presenter. The operator applies the instructions coming from the presenter and sometimes makes suggestions about possible moves.

The pairs were exposed to three different conditions during the experiment: no-cue, color-cue, and gaze-cue. The results reveal that the condition has a slight effect on the demonstrative use. The number of demonstratives used by the speakers in color-cue condition is somewhat less than the no-cue and gaze-cue conditions. The reason behind this could be color as a feature of an object is salient. Referring expressions are generally ambiguous. However, if possible referents in the surrounding have distinctive colors, this decreases the complexity of the visual scene, therefore the need to use extra reference.

Moreover, the condition also influences the role. For instance, the operator's use of *şu* in the gaze-cue condition is more frequent than other conditions. The gaze cue, here, is effective because the operator can follow where the presenter's gaze is located on the shared screen. Therefore, she can act on her own initiative and suggest a possible move when the presenter is visually attending close to the relevant piece. Similarly, there is an increase in the presenter's use of *bu* in the color cue condition compared to others. The role the speaker assumes is essential in the task. The presenter is normally leading the moves. Again, since color is salient, the presenter can show more initiative, by referring easier and faster and suggesting more moves.

The eye gaze can be a cue to resolve temporal ambiguities. To this end, we examined whether the speaker visually attends the intended piece when she uses a demonstrative reference. We found out that there is a significant relationship between visual attention and demonstrative use. The eye gaze of the speaker is generally over or around the object when they refer to it with a demonstrative regardless of the type. This finding is in line with Meyer et al. (1998)'s study, which suggests that people glance at the objects before naming them. It seems the direction of the eye gaze is informative in the sense that it reveals the focus of attention of the speaker and hence foreshadows the object of reference. Hanna & Brennan (2007) claim such eye movements are natural accompaniment to reference production and can be used during reference resolution by the addressee. Our findings support their claim, revealing the tendency to visually attend the referent at the time of utterance.

However, there also seems a significant interaction between visual attention and the role of the speaker. The presenters did not attend the referent 70 times when they used the demonstrative *o*, and 21 times when they used *bu*. This might occur because the presenter needs to compute a step ahead after referring to an object in the previous move. Griffin & Bock (2000) show when multiple objects are being named, speech is produced about one object while the next object is fixated. This might be causing the effect of the speaker role. Trying to achieve the goal, the presenter is always keeping an eye on the target shape, revisiting it all the time with saccades.

Recent attendance to a referent by the interlocutors can determine the next demonstrative choice of the speaker. To explore this, we examined the recent presence of joint attention between interlocutors on an intended object. We observed that the participants were very likely to refer to objects they had already referred to regardless of the demonstrative type. In other words, they establish a joint attention on an object after the first reference and continue referring to it. Our results indicate that the presence of recent joint attention has a significant effect on the demonstrative type the speaker use for the referent the next time. In general, the addresser prefers the demonstratives *bu* and *o* when the interlocutors recently attended the object. On the other hand, the addresser uses the demonstrative *şu* to introduce a new entity into the discourse. These findings support Özyürek & Kita (2000)'s analysis of the Turkish demonstrative system.

However, our observations contrast with their distinction in the use of *bu* and *o* based on proximity. They state *bu* is used for proximal objects whereas *o* for distal referents. We argue it is not primarily distance which differentiates their use although demonstratives encode distance. We detected that the speaker used firstly *bu* and then *o* to refer to the same entity several times in the same discourse segment. We claim what governs their use before distance is the cognitive status of the referent in the addressee's mind, as Gundel et al. (1993) suggested. However, we should note that unlike Özyürek and Kita's analysis which is based on typical examples of *bu*, *şu*, *o* use in hypothetical face to face dialogues, whereas our examples are sampled from a computer-mediated collaborative puzzle solving task, which may have an effect on the way we distinguish the use of *bu* and *o* on recently attended objects. A follow up study including a similar task conducted in a face to face setting seems to be necessary to explore the generality of this distinction.

We measured the distance between the eye gaze of the addresser and the addressee to find out if there is a gaze overlap at the time demonstrative utterance. In other words, we aim to look at whether the demonstrative choice is driven by the eye gaze distance. The results show that there is not a significant relationship between them. However, early studies on Turkish demonstratives state physical distance is central. It appears distance is; however, not physical but psychological, as suggested by Enfield (2003). The choice of proximal or distal demonstrative by the speaker seems to depend on the perception and interpretation of the interlocutors.

On the other hand, when we looked at the relationship between the mean gaze distance and the speaker attendance to a referent which was recently in the joint attention of the pair, we found out that the speaker's visual attendance had a significant effect on gaze distance. This result suggests that the eye gaze distance at the time of demonstrative utterance is minimal if the interlocutors have already established a joint attention on the intended object.

We also investigated whether a demonstrative use by the addresser ended up with demonstrative resolution by the addressee (i.e. gaze coordination within 2 seconds). We observed a significant effect of demonstrative use on gaze coordination. That is to say, the listener tends to look at the object of reference within two seconds after demonstrative utterance. This finding is compatible with the findings of Richardson & Dale (2005), which stated that the eye movements of speakers and listeners are linked in the context of a joint picture viewing task. In particular, they claim that a listener is likely to attend a picture referred by the speaker about 2 seconds after the speaker uttered the reference. In our design, the listener looked at the referred object generally in less than 2 seconds.

In addition, when we explored which factors influenced the reference resolution time, we found out that the interaction between the role of the speaker, the type of the demonstrative and the condition had a marginally significant effect. This occurs because when the presenter uses the demonstrative *şu*, it requires some effort for the listener to disambiguate the intended object. However, the gaze-cue condition seems to somewhat eliminate this extra cost thanks to the extra information it provides to the listener about where the speaker is currently looking at. Moreover, we looked at the relationship between the mean resolution time and the speaker attendance to a referent which was recently co-attended by the pair. The results suggest that the listener was able to disambiguate the demonstrative reference more quickly when the presenter visually attended the object of demonstrative reference which was already in joint attention since the gaze distance in such cases was minimal.

CHAPTER 6

CONCLUSION

This chapter summarizes the findings of the study and suggests directions for future research.

6.1 Summary of the Findings

This study, to the best our knowledge, is among the few studies which investigated the Turkish demonstratives within a situated and distributed perspective. By incorporating dual eye trackers into the Tangram problem solving task originally developed by Tokunaga et al. (2010), Acartürk & Çakır (2012) conducted an experiment to build a corpus of Turkish referring expressions. We selected 12 of the videos including speech and eye gaze of two pairs. After transcribing and annotating the data, we conducted several statistical tests to analyze the relationship between demonstrative use and eye gaze. The results indicate the following:

- In our data, *o* is the most frequently used demonstrative, followed by *bu* and then *şu*.
- The speaker role plays a significant role on demonstrative type. While presenters tend to use *o* more than *bu* and *şu*, operators prefer to use *şu* more than others.
- The different cue conditions have a slight effect on the number of and type of demonstratives used by the participants. While the color cue causes a decrease in the number of demonstratives, the gaze cue causes an increase in the number of *şu* used by the operator.
- There is a temporal relationship between visual attention and demonstrative use. The eye gaze of the speaker is generally over or around the object when they refer to it with a demonstrative regardless of the type.
- The recent joint attention on the intended object influences the speaker's choice of next demonstrative use.
- The eye gaze distance between the participants seem to have no effect on demonstrative use.
- Demonstrative use has a significant effect on gaze coordination. The listeners generally look at the intended referent in less than 2 seconds.

Our findings suggest demonstrative reference does not occur randomly, rather follows certain cognitive principles, which questions the prevailing speaker-centric view. They clearly show it is a joint activity in which both the speaker and the listener play active roles. They also indicate eye gaze proves to be a significant visual cue, which is temporally linked with demonstrative use.

6.2 Limitations and Directions for Future Research

The design of the experiment placed certain limitations on our study. For instance, we could not investigate whether the listener made use of the speaker's eye gaze as a cue during demonstrative disambiguation since they were situated in different locations. Another limitation was the task environment. It was a computer mediated problem-solving environment which did not reflect the distance perception of a natural setting properly. However, it might yield important findings for multimodal interfaces. Eye trackers employed in the experiment also have certain limitations. As well known in the eye tracking literature, eye trackers suffer from accuracy and precision problems. It is difficult to identify the exact location of the eye gaze, but we employed 100-pixel criterion to overcome this issue.

To overcome those limitations, a follow up study which includes a similar problem-solving task in a face to face setting should be carried out using glass eye trackers. Such equipment will allow us to examine the interaction between the speaker and the listener in a more naturalistic situated dialogue.

Bibliography

Acartürk, C., & Çakır, M. P. (2012). Towards building a corpus of turkish referring expressions. In *Proc. 1st workshop on language resources and technologies for turkic languages* (pp. 1–5).

Bach, K. (2008). On referring and not referring. *Reference: interdisciplinary perspectives*, 13–58.

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

Brennan, S. E., & Clark, H. H. (1996). Conceptual pacts and lexical choice in conversation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22*(6), 1482.

Byron, D. K., & Fosler-Lussier, E. (2006). The osu quake 2004 corpus of two-party situated problem-solving dialogs. In *Proceedings of the 15th language resources and evaluation conference.*

Çallı, A. B. S. (2012). Demonstrative anaphora in turkish: A corpus based analysis. In *First workshop on language resources and technologies for turkic languages* (p. 33).

Clark, H. H. (2003). Pointing and placing. In *Pointing* (pp. 251–276). Psychology Press.

Clark, H. H., & Bangerter, A. (2004). Changing ideas about reference. In *Experimental pragmatics* (pp. 25–49). Springer.

Clark, H. H., & Wilkes-Gibbs, D. (1990). Referring as a collaborative process. *Intentions in communication*, 463–493.

Coventry, K. R., Valdés, B., Castillo, A., & Guijarro-Fuentes, P. (2008). Language within your reach: Near–far perceptual space and spatial demonstratives. *Cognition*, *108*(3), 889–895.

Diessel, H. (1999). *Demonstratives: Form, function and grammaticalization* (Vol. 42). John Benjamins Publishing.

Diessel, H. (2006). Demonstratives, joint attention, and the emergence of grammar. *Cognitive linguistics*, *17*(4), 463–489.

Diessel, H. (2014). Demonstratives, frames of reference, and semantic universals of space. *Language and Linguistics Compass*, *8*(3), 116–132.

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

Dixon, R. M. (2003). Demonstratives: A cross-linguistic typology. *Studies in Language. International Journal sponsored by the Foundation "Foundations of Language"*, *27*(1), 61–112.

Eilan, N., Hoerl, C., McCormack, T., Roessler, J., et al. (2005). *Joint attention: Communication and other minds: Issues in philosophy and psychology*. Oxford University Press on Demand.

Enfield, N. J. (2003). Demonstratives in space and interaction: Data from lao speakers and implications for semantic analysis. *Language*, *79*(1), 82–117.

Friedrich, A., & Palmer, A. (2014). Centering theory in natural text: a large-scale corpus study. In *Konvens* (pp. 137–144).

Gordon, P. C., Grosz, B. J., & Gilliom, L. A. (1993). Pronouns, names, and the centering of attention in discourse. *Cognitive science*, *17*(3), 311–347.

Grice, H. P. (1975). Logic and conversation. *1975*, 41–58.

Griffin, Z. M., & Bock, K. (2000). What the eyes say about speaking. *Psychological science*, *11*(4), 274–279.

Grosz, B. J., Joshi, A. K., & Weinstein, S. (1983). Providing a unified account of definite noun phrases in discourse. In *Proceedings of the 21st annual meeting on association for computational linguistics* (pp. 44–50).

Gundel, J. K., & Hedberg, N. (2008). *Reference: interdisciplinary perspectives*. Oxford University Press.

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

Halliday, M. A., & Hasan, R. (1976). Cohesion in. *English, Longman, London*.

Hanna, J. E., & Brennan, S. E. (2007). Speakers' eye gaze disambiguates referring expressions early during face-to-face conversation. *Journal of Memory and Language*, *57*(4), 596–615.

Hayes, A. F., & Krippendorff, K. (2007). Answering the call for a standard reliability measure for coding data. *Communication methods and measures*, *1*(1), 77–89.

Henderson, J. M. (2003). Human gaze control during real-world scene perception. *Trends in cognitive sciences*, *7*(11), 498–504.

Hollan, J., Hutchins, E., & Kirsh, D. (2000). Distributed cognition: toward a new foundation for human-computer interaction research. *ACM Transactions on Computer-Human Interaction (TOCHI)*, *7*(2), 174–196.

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

Hutchins, E. (1995a). *Cognition in the wild*. MIT press.

Hutchins, E. (1995b). How a cockpit remembers its speeds. *Cognitive science*, *19*(3), 265–288.

Just, M. A., & Carpenter, P. A. (1976). Eye fixations and cognitive processes. *Cognitive psychology*, *8*(4), 441–480.

Kılıçaslan, Y., Güner, E. S., & Yıldırım, S. (2009). Learning-based pronoun resolution for turkish with a comparative evaluation. *Computer Speech & Language*, *23*(3), 311–331.

Kornfilt, J. (1997). Turkish. descriptive grammars. *London and New York*.

Krippendorff, K. (2004). Reliability in content analysis: Some common misconceptions and recommendations. *Human communication research*, *30*(3), 411–433.

Kuntay, A., & Ozyurek, A. (2002). Joint attention and the development of the use of demonstrative pronouns in turkish. In *26th annual boston university conference on language development* (pp. 336–347).

Küntay, A. C., & Özyürek, A. (2006). Learning to use demonstratives in conversation: what do language specific strategies in turkish reveal? *Journal of Child Language*, *33*(2), 303–320.

Kuriyama, N., Terai, A., Yasuhara, M., Tokunaga, T., Yamagishi, K., & Kusumi, T. (2011). Gaze matching of referring expressions in collaborative problem solving. In *Proceedings of international workshop on dual eye tracking in cscw (duet 2011).*

Levinson, S. C. (1983). Pragmatics. cambridge textbooks in linguistics. *Cambridge/New York*.

Lyons, J. (1977). Semantics (vols i & ii). *Cambridge CUP*.

Meyer, A. S., Sleiderink, A. M., & Levelt, W. J. (1998). Viewing and naming objects: Eye movements during noun phrase production. *Cognition*, *66*(2), B25–B33.

Nüssli, M.-A. (2011). Dual eye-tracking methods for the study of remote collaborative problem solving.

Olson, D. R. (1970). Language and thought: Aspects of a cognitive theory of semantics. *Psychological review*, *77*(4), 257.

Özyürek, A., & Kita, S. (2000). Attention manipulation in the situational use of turkish and japanese demonstratives. In *Linguistic society of america conference, chicago, in: Kuntay, a (2007) disassembling the puzzle using the crosslinguistic methodology: Koç üniversitesi dergisi (2).*

Peeters, D., & Özyürek, A. (2016). This and that revisited: A social and multimodal approach to spatial demonstratives. *Frontiers in psychology*, *7*, 222.

Piwek, P., Beun, R.-J., & Cremers, A. (2008). 'proximal'and'distal'in language and cognition: Evidence from deictic demonstratives in dutch. *Journal of Pragmatics*, *40*(4), 694–718.

Pusch, M., & Loomis, J. (2001). Judging another person's facing direction using peripheral vision. *Journal of Vision*, *1*(3), 288–288.

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.

Richardson, D. C., Dale, R., & Kirkham, N. Z. (2007). The art of conversation is coordination. *Psychological science*, *18*(5), 407–413.

Say, B., Zeyrek, D., Oflazer, K., & Özge, U. (2002). Development of a corpus and a treebank for present-day written turkish. In *Proceedings of the eleventh international conference of turkish linguistics* (pp. 183–192).

Spanger, P., Masaaki, Y., Ryu, I., & Takenobu, T. (2009). A japanese corpus of referring expressions used in a situated collaboration task. In *Proceedings of the 12th european workshop on natural language generation* (pp. 110–113).

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.

Stoia, L., Shockley, D. M., Byron, D. K., & Fosler-Lussier, E. (2008). Scare: a situated corpus with annotated referring expressions. In *Lrec*.

Takenobu, T., Ryu, I., Asuka, T., & Naoko, K. (2012). The rex corpora: A collection of multimodal corpora of referring expressions in collaborative problem solving dialogues. In *Proceedings of the eighth international conference on language resources and evaluation (lrec 2012). european language resources association (elra).*

Tokunaga, T., Iida, R., Yasuhara, M., Terai, A., 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 resouces* (pp. 38–46).

Turan, Ü. D. (1996). Null vs. overt subjects in turkish discourse: A centering analysis. *IRCS Technical Reports Series*, 95.

Yüksel, Ö., & Bozsahin, C. (2002). Contextually appropriate reference generation. *Natural Language Engineering*, *8*(1), 69–89.

Yule, G. (1996). *Pragmatics/george yule.* Oxford: Oxford University Press.

Zeyrek, D., Demirşahin, I., Sevdik-Çalli, A., Balaban, H. Ö., Yalçinkaya, İ., & Turan, Ü. D. (2010). The annotation scheme of the turkish discourse bank and an evaluation of inconsistent annotations. In *Proceedings of the fourth linguistic annotation workshop* (pp. 282–289).

Zupnik, Y.-J. (1994). A pragmatic analysis of the use of person deixis in political discourse. *Journal of pragmatics*, *21*(4), 339–383.

Appendix A

A SAMPLE ANNOTATION ACCORDING TO THE GUIDELINE

A.1 AN EXCERPT FOR THE SAMPLE ANNOTATION:

The excerpt below was taken from one of the twelve video transcripts in which the participants were trying to achieve the target shape vase in the color-cue condition. The excerpt was composed of two utterances. The speaker of the first utterance was the operator (with the yellow eye gaze in the video) and the second one was the presenter (with the pink eye gaze in the video).

1. **Operator:** *Şu* olabilir.

   "[That] might fit."

2. **Presenter:** *O* daha mantıklı.

   "[It] is more logical."

With the first utterance, the operator made a suggestion about a possible move, referring to the square in the working space with the demonstrative *şu*. In the second utterance, the presenter supported this suggestion referring to the square with the demonstrative *o*. Both utterances were annotated below in order to demonstrate how the annotation procedure was carried out.

A.2 SAMPLE ANNOTATION FOR THE FIRST UTTERANCE

1. **The role of the speaker:**

   The speaker of the demonstrative *şu* was the operator. Therefore, the role is coded as "O".

2. **The type of the demonstrative:**

   The demonstrative was *şu*. Therefore, "1" is entered into the related column.

3. **The Referent:**

   The demonstrative *şu* refers to the square. Therefore, it is coded with "s".

4. **Attended by the speaker:**

The eye gaze of the speaker (yellow) is over the square at the time of demonstrative utterance. In other words, the referent is attended by the speaker (See Figure A.1). Therefore, "1" is entered into the related column.

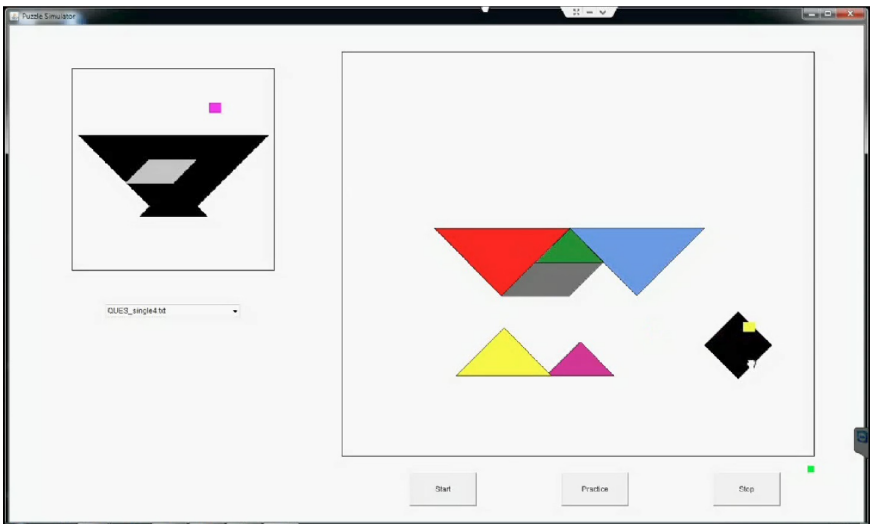


Figure A.1: The speaker (yellow eye gaze) visually attends the square.

5. **Recently attended by the pair:**

   The pair did not visually attend the square before, so there is no joint attention on the referent. Therefore, "0" is entered into the related column.

6. **Gaze Overlap at the time of utterance:**

   The distance between eye gaze locations is more than 100 pixels, so there is no gaze overlap. Therefore "0" is entered into the related column (See Figure A.2).

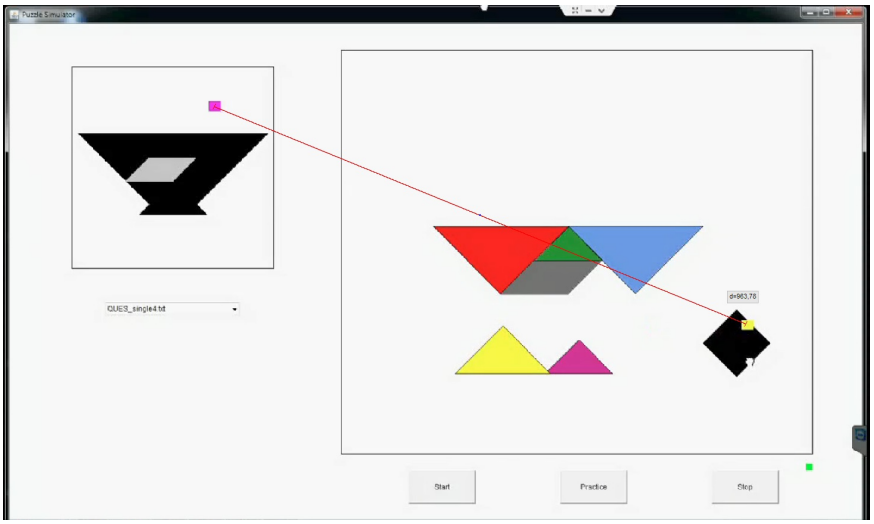


Figure A.2: The distance between eye gaze locations at the time of utterance

7. **Gaze distance at the time of utterance:**

   The distance between eye gaze locations is also recorded in pixels. Therefore, "963" is entered into the related column (See Figure A.2).

8. **Demonstrative resolution within 2 seconds after utterance:**

   The eye gaze of the listener (pink) falls over the square within 2 seconds. Therefore, we assume the listener resolves the demonstrative and enter "1" into the related column (See Figure A.3).

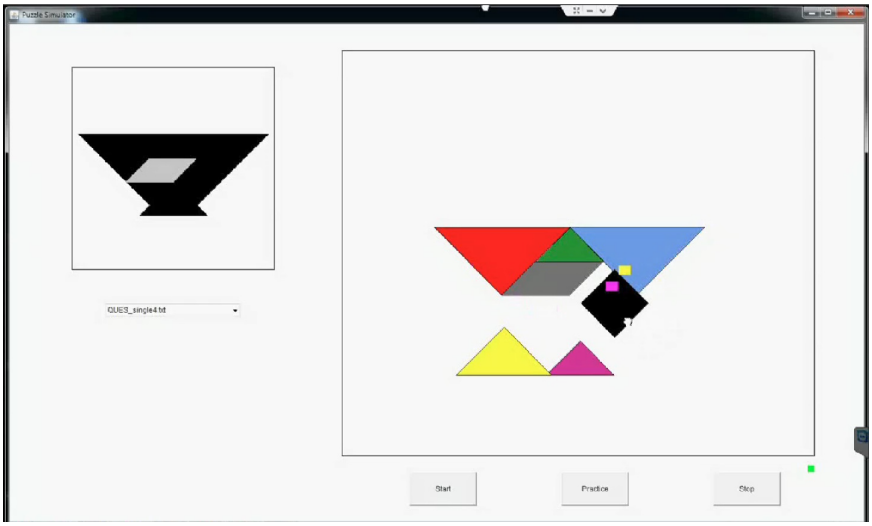


Figure A.3: The listener (pink eye gaze) visually attends the square.

9. **Demonstrative resolution duration:**

   It takes almost half a second for the listener to look at the square. Therefore, we enter 0.5 into the related column.

A.3 SAMPLE ANNOTATION FOR THE SECOND UTTERANCE

1. **The role of the speaker:**

   The speaker of the demonstrative *o* is the presenter. Therefore, the role is coded as "P".

2. **The type of the demonstrative:**

   The demonstrative is *o*. Therefore, "1" is entered into the related column.

3. **The Referent:**

   The demonstrative *o* refers to the square. Therefore, it is coded with "s".

4. **Attended by the speaker:**

   The eye gaze of the speaker (pink) is over the square at the time of demonstrative utterance. In other words, the referent is attended by the speaker and the square (See Figure A.4). Therefore, "1" is entered into the related column.

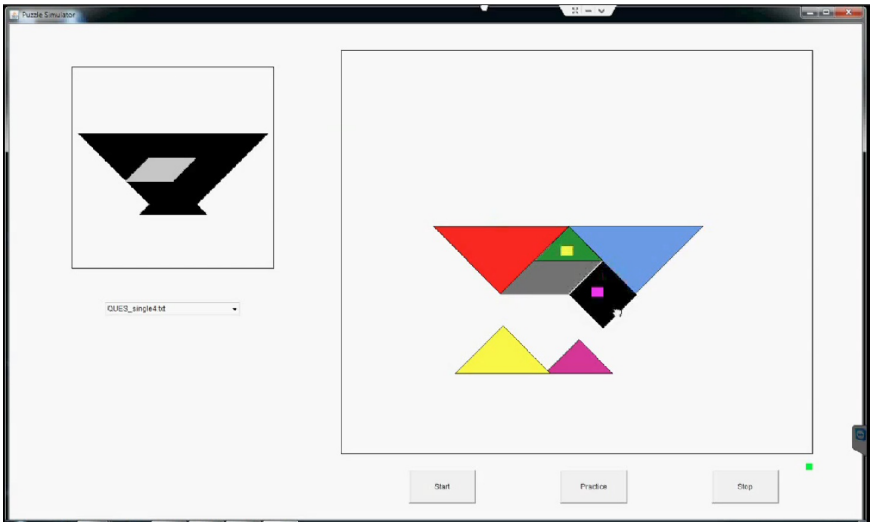


Figure A.4: The speaker (pink eye gaze) visually attends the square.

5. **Recently attended by the pair:**

   The pair visually attended the square in the first utterance, so they established joint attention on the referent. Therefore, "1" is entered into the related column.

6. **Gaze Overlap at the time of utterance:**

   The distance between eye gaze locations is more than 100 pixels, so there is no gaze overlap. Therefore "0" is entered into the related column (See Figure A.5).

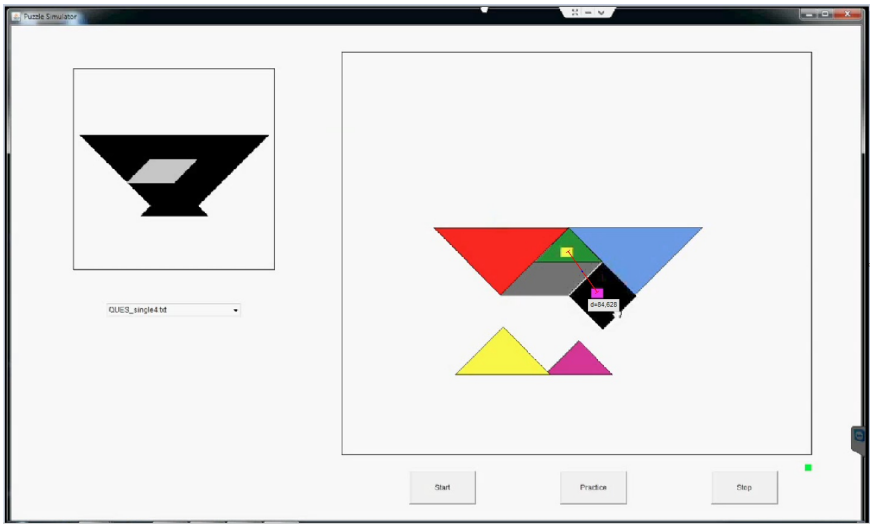


Figure A.5: The distance between eye gaze locations at the time of utterance

7. **Gaze distance at the time of utterance:**

   The distance between eye gaze locations is also recorded in pixels. Therefore, "84" is entered into the relevant column (See Figure A.5).

8. **Demonstrative resolution within 2 seconds after utterance:**

   The eye gaze of the listener (yellow) falls over the square within 2 seconds. Therefore, we assume the listener resolves the demonstrative reference and enter "1" into the related column (See Figure A.6).

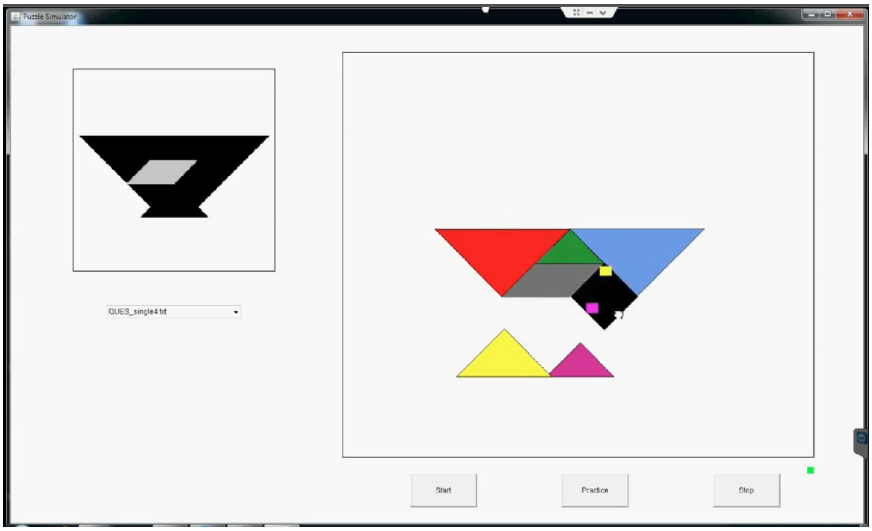


Figure A.6: The listener (yellow eye gaze) visually attends the square.

9. **Demonstrative resolution duration:**

   It takes almost half a second for the listener to look at the square. Therefore, we enter 0.5 into the related column.