# MREDI: Multimodal Reference in Dialogue: Annotator's manual

## Introduction

Please make sure you have the following available:

1. A DVD disc with video files;
2. 4 files containing maps;
3. A set of Excel files with dialogue transcripts.

## Maps

Each dialogue in the corpus involves a director explaining to their partner, the follower, how to get from one landmark to the other on a shared map, in the exact same order that is shown on the director's private map, through the exact same landmarks. The partner's task, on the other hand, is to draw the route that the director explains on their own copy of the map. There are exactly 18 landmarks on a map.

There are **4 different maps**, as follows:

1. A map consisting of circles, where the landmarks are *individual circles*.
2. A map consisting of circles where the landmarks are *groups of circles*.
3. A map consisting of squares, where the landmarks are *individual squares*.
4. A map consisting of squares, where the landmarks are *groups of squares*.

You should have each one of these maps in a separate file. The **itinerary on each map is marked by numbering the landmarks** in the order in which they should be visited. Some landmarks have **two numbers**. That means that they are visited twice. For example, if a place on the map has number 2 and number 5, that means it is both the second and the fifth stopping point on the itinerary.

## Participants

Participants worked in pairs and are given a numeric ID. This is reflected in the name of the video file for a participant. For example S1+S2 means "participant 1 and participant 2".

Every pair of participants did all four maps, and switched roles from one map to the other. Therefore, for every pair you annotate, you will have 4 videos, labelled map1 through map4. Since they switched roles, for every pair of participants S1 and S2, there are 2 dialogues where S1 is director, and a further two where S2 is director.

Different pairs of participants did the maps in different orders. These are shown below:

| order no. | map1 | map2 | map3 | map4 |
|---|---|---|---|---|
| 1 | IC | GC | IS | GS |
| 2 | IS | GS | IC | GC |
| 3 | IC | GS | IS | GC |
| 4 | IS | GC | IC | GS |
| 5 | GC | IC | GS | IS |
| 6 | GS | IS | GC | IC |
| 7 | GS | IC | GC | IS |
| 8 | GC | IS | GS | IC |
| 9 | IC | IS | GC | GS |
| 10 | IS | IC | GS | GC |
| 11 | IC | IS | GS | GC |
| 12 | IS | IC | GC | GS |
| 13 | GC | GS | IC | IS |
| 14 | GS | GC | IS | IC |
| 15 | GS | GC | IC | IS |
| 16 | GC | GS | IS | IC |

You will know which order corresponds to which participant pair, because it is indicated in the file name of the video files. For example, O2_S3S4 means "participants 3 and 4 did the maps in order 2".

## Annotation

During the dialogue, the two interlocutors may talk about a given landmark over several turns. We divide the dialogue up into **stations** where every station is a segment of the dialogue in which the interlocutors are talking about one landmark. For example, *station 1* is the part of the dialogue where they talk about the first landmark. Here is an example:

| **Utterance** | | | **Station** |
|---|---|---|---|
| 1 | D | Right you start off the **smallest** one **beside the start** | 1 |
| 2 | M | Yeah | 1 |
| 3 | D | And then you take the, you take a diagonal right through the sort of pathway | 2 |
| 4 | M | Yeah straight [XXX space] | 2 |
| 5 | D | [Until] Until you hit the next biggest one or the biggest one (2) that you can see straight ahead of you | 2 |
| 6 | M | Well so I go from the little one straight in front the start | 2 |
| 7 | D | And then right down that passage way | 2 |
| 8 | M | Pass the 3 | 2 |
| 9 | D | Yes straight down | 2 |

| | | | |
|---|---|---|---|
| 10 | M | Until that open space and keep going [right in diagonal until] | 2 |
| 11 | D | [Yeah until you see] that big one the biggest one right in the middle way | 2 |
| 12 | M | I'm doing okay so  the big one di- directly diagonal from another big one at the start? | 2 |
| 13 | D | Yeah, yeah that one | 2 |

In this example, the first two turns are about the first landmark (station 1); the next 11 turns are about the second landmark (station 2). Note that:

- Utterances are numbered consecutively.
- Every utterance is marked D ("director") or M ("follower"). **You should only annotate the director's utterances**.
- You will need to identify the stations, that is, you will need to indicate, for each dialogue utterance, which of the 18 stations it belongs to. Stations have to be in order and go from 1 to 18. In other words, you need to segment the dialogue into sequences of consecutive utterances that are about the same station. Typically, there will be one unique segment for each station and the segments will have the same order as the stations of the itenary  (though there are exceptions to this rule, see below under Linguistic Features - Identity).

## How to annotate

Use the excel workbook corresponding to each dialogue pair you're annotating. Each workbook will contain 4 sheets, numbered map1 through map4, corresponding to the 4 maps for the dialogue pair in the order in which they're done.

When you're annotating, you should keep the relevant map in front of you, as it is crucial for you to know (a) what target is being talked about; (b) what station the dialogue has arrived at.

Among the features to code are a number of linguistic features, that is, features related to the content of the utterance. To code these, you will generally not need to look at the video, but only at the transcript. There are also gestural features, those related to how participants point to objects on the shared map. For these, you will need to use the video.

Features are of two kinds:

1. **Frequency***:* for these, you need to mark up the **number of times a particular feature occurs**. Most features are of this type.
2. **Boolean**: for these, you need to mark up **whether or not the feature occurs at least once**.

## Features to code 1: Linguistic Features

Here are the feature definitions:

| **Variable** | **Type** | **Description** | **Comment** | **Examples/Counterexamples** |
|---|---|---|---|---|
| Relative position (RP) | Frequency | The number of times the position of the target landmark is described relative to another landmark. | Defined as a reference to the location of a target in relation to another object, when the other object is explicitly mentioned. Note that this criterion is **syntactic**, i.e. there has to be an explicit relation expressed syntactically between the target and another landmark. Typically, speakers will use prepositional phrases like *below X, near X* and so on. | • The blue square just below the red square<br>• The 5 little ones at the start<br>**But not:**<br>• south-west<br>• Some squares just on their own |
| Absolute position | Frequency | Mention of target position using an | Defined as a reference to location of a target using absolute coordinates. | • The blue circle down at the bottom |

| | | | | |
|---|---|---|---|---|
| (AP) | | absolute locative frame of reference, i.e., without using another landmark to locate the target. | | **But not:**<br>• The blue circle near the red one |
| Frequency of references on path (FP) | Frequency | Number of references to **non-target** landmarks which are mentioned as being on the path to the target. | Any reference to an object which is not the target. This means that AP and RP are also classified as FP, but not every FP is classified as an AP or RP. | • And you're going to go east to the first tiny square, past the blue one.<br>This includes references to a configuration of objects and descriptions of space delimited by objects, e.g.:<br>• the sort of pathway, on its own<br>• the red one with no others around it |
| Size (S) | Frequency | Mention of size of target | Note that this only applies to descriptions of the target, not of other landmarks which may be mentioned. | • The large one<br>• the group of small circles<br>The one below has only **one** size mention; the second one ("big") is for the non-target, so it isn't counted:<br>• The large square near the big red square |
| Shape (Sh) | Frequency | Mention of target shape (which is usually circle or square) | Note that this only applies to descriptions of the target, not of other landmarks which may be mentioned. | • The circles at the bottom<br>The one below has only **one** size mention; the second one ("square" in "red square") is for the non-target, so it |

| | | | | |
|---|---|---|---|---|
| | | | | isn't counted:<br>• The blue square near the red square |
| Color (C) | Frequency | Mention of color of target | Note that this only applies to descriptions of the target, not of other landmarks which may be mentioned. | This example has only one mention of target colour:<br>• The red ones next to the blue ones |
| Deixis (D) | Frequency | Use of a deictic reference. | This feature applies to any deictic expression, whether it's demonstrative or something else. This **excludes** existential constructions (e.g. ***there's*** *a big red square at the top*). | • Over here<br>• down there<br>• this one<br>• those squares |
| Identity (I) | Frequency | Statement of identity between the current target and a previous or later target, whether implicit or explicit. | This includes cases where somebody refers to an object and anticipates a later reference to it. This may be implicit. | • Number 4 is the blue square, which is also number 17.<br>• The red square, the same one we saw at number 5.<br>The example below has two identity references:<br>• Go back to the red one again. |
| Directions (DIR) | Frequency | Direction-giving. | Any direction which is within the universe of the map (i.e. does not make reference to the subjects' own 3D universe). The variable is restricted to expressions giving directions to be followed by the addressee. These will | The example below has 3 occurrences of Direction:<br>• Take a right, go across and straight down |

| | | | | |
|---|---|---|---|---|
| | | | typically be in the *imperative*. | |

The example snippet above is shown below with the linguistic features annotated. Blank cells are equal to zero.

| Turn | Role | **Map 1 (single squares)** | Station | VE | DIR | RP | AP | FP | S | Sh | C | D | I | ELO |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | D | Right you start off the **smallest** one **beside the start** | 1 | | | 1 | | 1 | 1 | | | | | |
| 2 | M | Yeah | 1 | | | | | | | | | | | |
| 3 | D | And then you take the, you take a diagonal right through the sort of pathway | 2 | | 1 | | | 1 | | | | | | |
| 4 | M | Yeah straight [XXX space] | 2 | | | | | | | | | | | |
| 5 | D | [Until] Until you hit the next biggest one or the biggest one (2) that you can see straight ahead of you | 2 | | | | | | 2 | | | | | |
| 6 | M | Well so I go from the little one straight in front the start | 2 | | | | | | | | | | | |
| 7 | D | And then right down that passage way | 2 | | 1 | | | 1 | | | | | | 2 |
| 8 | M | Pass the 3 | 2 | | | | | | | | | | | |
| 9 | D | Yes straight down | 2 | | 1 | | | | | | | | | |
| 10 | M | Until that open space and keep going [right in diagonal until] | 2 | | | | | | | | | | | |
| 11 | D | [Yeah until you see] that big one the biggest one right in the middle way | 2 | | | | 1 | | 2 | | | | | |
| 12 | M | I'm doing okay so  the big one di- directly diagonal from another big one at the start? | 2 | | | | | | | | | | | |
| 13 | D | Yeah, yeah that one | 2 | | | | | | | | | 1 | | |

## Gestural features

Gestural features cover all the **pointing gestures** made by the director. A pointing gesture is one where somebody is explicitly indicating an object on the map using a movement of the arm and/or hand. This excludes so-called iconic gestures, where a person uses the hands to indicate a shape, and also other movements of the arms/hands which are not pointing.

| Variable | Type | Description |
|---|---|---|
| Elbow on table (ELO) | Frequency | Number of times a speaker points with her elbow on the table. |
| Elbow off table (ELF) | Frequency | Number of times a speaker points with her elbow off the table. |
| Full extension (EXF) | Frequency | Number of times a speaker points with her arm fully extended. |
| Partial Extension (EXP) | Frequency | Number of times a speaker points with her arm partially extended. |
| Gaze at shared map (GZ) | Boolean | Whether both participants are looking at the shared map at least once during an utterance. |

**Some important things to note about pointing and gestures:**

1. **Frequency:** Within a particular station, you should always observe that ELF+ELO = EXF+EXP. This is because any ELF or ELO gesture will also be an EXF or an EXP gesture, since an act of pointing is classified in a dual way, in terms of whther the elbow is on or off the table, and whether it is full or partial.

2. **Coding:** For each station, we only annotate pointing gestures to the target. When people are talking about a different object, and point to it, do not count these gestures. Sometimes, you'll need to make an educated guess about whether a gesture is to the target landmark or not. Use your intuition, based on what you see in the video and what you read in the transcript.

**Criteria for recognising gestures:**

1. The classification of a pointing gesture occurs at its *peak*, i.e. the point of greatest extension of the arm during the actual gestural motion.
2. If a gesture is directed at the target by the speaker, we count it as a pointing gesture, even if the gesture also has other functions (e.g., it might additionally be iconic or indicate the direction of a path).
3. Criterion for recognising pointing gestures: the gesture is only meaningful in relation to the map. This discounts the sorts of gestures obtained when two participants face each other and use their hands to accentuate speech or make iconic gestures.
4. In case a gesture is ambiguous, i.e it is not clear whether it is used to point to the target or not, it is useful to cross-reference the transcript: if the dialogue at that point is talking about a non-target object, and a gesture is not clearly a target gesture, it is quite likely that it isn't.

**Single vs. multiple gestures:**

1. Some gestures can have multiple beats (i.e. the arm or hand is partially extended, partially retracted and then extended again, multiple times). A gesture with multiple beats but pointing towards the same object **without retraction** is counted as a single gesture. Note that, in many cases, retraction will not involve a complete move of the arm back to the table or the person's side. Therefore, retraction here is counted as **significant movement towards the rest position**. For example, once the arm is (partially) extended, a flick of the finger or a movement of the palm of the hand does **not** count as a new gesture, because the flick or movement of the hand is not a retraction.