



Social Interaction. Video-Based Studies of Human Sociality.
2023 Vol. 6, Issue 1
ISBN: 2446-3620
DOI: 10.7146/si.v6i1.137380

Social Interaction

Video-Based Studies of Human Sociality

Managing Turn-Taking in Human-Robot Interactions: The Case of Projections and Overlaps, and the Anticipation of Turn Design by Human Participants

***Ali Reza Majlesi^{1*}, Ronald Cumbal², Olov Engwall², Sarah Gillet², Silvia
Kunitz³, Gustav Lymer¹, Catrin Norrby¹ & Sylvaine Tuncer⁴***

¹Stockholm University

²Kungliga Tekniska Högskolan

³Linköping University

⁴King's College

Abstract

This study deals with turn-taking in human-robot interactions (HRI). Based on 15 sessions of video-recorded interactions between pairs of human participants and a social robot called Furhat, we explore how human participants orient to violations of the normative order of turn-taking in social interaction and how they handle those violations. As a case in point, we present sequences of HRI to show particular features of turn-taking with the robot and also how the robot may fail to respond to the human participants' bid to take a turn. In these sequences, the participants either complete the turn in progress and ignore the overlap caused by the robot's continuation of its turn, or they cut short their own turn and restart in the next possible turn-transition place. In all cases

* The main author of this paper is the first author; the rest of the co-authors contributed in various ways and are thus listed in alphabetical order.

in our data, the overlaps and failed smooth turn-transitions are oriented to as accountable and in some sense interactionally problematic. The results of the study point not only to improvables in robot engineering, but also to routine practices of projection and the ways in which human subjects orient toward normative expectations of ordinary social interactions, even when conversing with a robot.

Keywords: human-robot interaction, conversation analysis, turn-taking, projection, overlaps

1. Introduction

Within the field of conversation analysis, there has recently been increased research interest in interactional aspects of talk with robots (e.g., Pelikan & Broth, 2016; Tuncer et al., 2022). Although engineers have studied interactional issues including turn-taking in human-robot interactions (henceforth HRI) for a few decades (see Skantze, 2021, for a review), when it comes to naturally occurring social practices in HRI, the details of interaction are yet to be uncovered.

Drawing on ethnomethodological conversation analysis (EMCA) (Sacks, 1995; Schegloff, 2007), in this study we explore how human participants in HRI orient to the systematics of turn-taking organization (Sacks et al., 1974), how they use the projectability of turn constructions to produce their actions (Mondada, 2021), and how such a projectability impacts on interactional practices when engaging with the robot. The study particularly deals with the emergence of troubles in turn-taking in multiparty conversation with a social robot called Furhat (see Gillet et al., 2021). The activity that we examine is a language game, and more specifically the initial part of the interaction (greeting-introduction), involving one robot and two human subjects. In one set of conversations, both participants speak Swedish as their first language (L1). In the other set, one of the participants is an L2 speaker of Swedish. We specifically demonstrate how turn design in the robot's scripted talk (particularly with repeated wordings and designated pauses) causes overlaps, leading to the human participants' engagement in interactional work to manage the trouble. We also show that this interactional work is accompanied by participants adopting a negative stance toward the trouble in question. The overlaps are thus treated as a violation of normative expectations in turn-taking.

2. Turn-Taking and Projectability in Interaction

Interactional partners accountably produce actions (Garfinkel, 1967) and carefully craft their turns at talk as they contribute to the flow of interaction. This means that speakers systematically take turns in particular transition relevance places (TRPs) (Sacks et al., 1974) and they do so with precision timing to fit their turns to the ongoing talk (Jefferson, 1973; Sacks, 1995), which requires them to constantly monitor each other's actions (M. Goodwin, 1980). Although nothing in naturally occurring interactions is pre-scripted (Schegloff, 1986), some rules and particular cues in interactions are employed to allow co-interactants to coordinate turn-taking, that is, to predict the trajectory of the course of action, or how a turn (or a turn-constructional unit [TCU]) within that course of action may be set to begin or end. The precision of timing for coordination in turn-taking is jointly achieved through the use of shared resources to show when to respond to an action, whether in overlap, latching on to an ongoing action, or with some delay, and also how to hold the floor or end a turn (see Sacks et al., 1974: 703).

As shown in EMCA research, in general the projectability of a course of action (i.e., understanding what an action is designed to do) relies on its recognizability, which is often based on the use of four interactional resources: (a) verbal resources (e.g., utterance types, discourse markers); (b) bodily resources (e.g., facial expressions, gestures, gaze direction); (c) vocal resources (e.g., prosody, perturbations in talk); and (d) other contextual resources (e.g., the overall structure of a course of action, material surroundings). Previous research has shown that there is often a combination of different resources (see research on action as a multimodal gestalt, e.g., Mondada, 2016) that provide grounds for co-participants in social interactions to recognize an ongoing course of action, and to contribute with relevant next moves in a timely fashion (e.g., in turn-taking).

Projectability is based on the normative procedure of unfolding activities and how human actions are organized to form a coherent gestalt. Actions are made and treated as relevant to one another and the emergence of one action affords and constrains the relevant next. This relational bundle of one action (any gestalt of, e.g., linguistic terms with gestures, vocalizations, etc.) with another action is instantiated in EMCA in the concept of paired actions (Sacks, 1995), or adjacency pairs (Schegloff, 2007). Projectability is also reflected in the rule of relevance in the turn-taking system, that is, how the next turn is made conditionally relevant by the prior turn (Sacks et al., 1974; Schegloff, 2007). Certain practices are used as normative conduct to transition from one action to another or from one turn to another, for instance, when the format of a turn (e.g., a question) makes the format of another turn (e.g., answer) relevant; when an address term, or even just gaze, is used for nominating the next speaker (Sacks et al., 1974: 717); when hand raising is used as a request for talk or bidding for the floor (Mehan, 1979: 91); and when pointing is used to project the transition to the next activity (Mondada, 2006).

The co-occurrence of social actions (e.g., adjacency pairs) forms both a routine and also a normative order. Therefore, the routine organization of an activity can also become a resource for understanding and predicting a routine at its onset (the routine order of action of everyday life is the basis for the recognition of a conduct as “reasonable” conduct, see Garfinkel, 1967: 279; see also Schegloff, 1986, on the routine as achievement). Psathas (1999) points to *sequence types* as routine accomplishments in everyday social activities, and how the typicality of sequences may be used as a resource for the projectability of conduct within those activities (see also early discussions of omnirelevance of some conduct in particular contexts in Sacks’ lecture, 1995: 515). Routine conduct (e.g., a particular way of talking or the production of a turn) may therefore lead or allude to understanding the type of sequence or action it may generate and therefore the prediction of what may come next in the sequence. The examples may be “summon-answer sequences,” “identification-recognition sequences,” or “greetings exchange sequences” (Psathas, 1999: 142), or “introduction rounds” (Sacks, 1995: 72). Further, studies have also shown that such routine conduct may fail to be followed in particular circumstances, such as technology-mediated

interactions where participants have limited visual access to each others' bodies (e.g., Schönfeldt & Golato, 2003) and conversations involving people with cognitive and communicative difficulties (see e.g., Walton et al., 2020).

In line with the studies above, in this paper we explore how normative orders as achieved routines in social situations are used as resources for recognizing and projecting social actions and thus for producing particular conduct in HRI. We specifically show how troubles in talk can emerge and how human participants seek to remedy these troubles, sometimes also displaying an emotional stance in response to breaches of the normative expectation of ordered ordinary conversation.

3. Turn-Taking in HRI: State of the Art

Studies of dialogue systems show how robotic engineers have been concerned with improving turn-taking design for robots in the past few decades (see Skantze, 2021, for a review). However, detailed studies of human conduct for modeling speech systems in HRI have only recently been significantly developed (Skantze, 2021). The problems that engineers have been tackling include finding methods through which the coordination of turn-taking between humans and virtual or physical agents could run smoothly. In his review of studies of turn-taking, Skantze (2021) provides an account of the central issues in conversational coordination, highlighting in particular projectability in naturally occurring interactions (e.g., predicting the trajectory of a course of action; see section 2 above) and the use of multimodal resources as coordination cues (e.g., as signals of turn initiation and turn completion). What seems to be part of the difficulty, that we too intend to raise in this study, is the sensitivity of human participants to the normative order of conversation.

Previous studies (e.g., Fischer, 2011a, b) have shown that human participants in HRI adapt to the interactional abilities of the robots. Studies of HRI involving humanoid robots have shown how human participants may “adjust their turn designs with respect to what they progressively discover” in interaction with the robot (Pelikan & Broth, 2016: 4927). According to Pelikan and Broth (2016: 4929), this adaptation includes the shortening of turns, using simpler words, and employing clearer prosodic marking. Other studies (e.g., Yamazaki et al., 2008) show how human participants in interaction with humanoid robots display sensitivity not only to verbal actions but also to the gaze and head movements of the robot. Such sensitivity may cause confusion for human participants if robotic head movements are not timed in relation to the production of TCUs and TRPs. This supports the results of many other studies pointing to the significance of attending to the multimodal nature of interaction and timing of the turn-taking organization in the design of robotic behavior (e.g., Thomaz & Chao, 2011; Lala et al., 2019; see also Skantze, 2021 for a review of studies on turn-taking in HRI).

In line with humans' ability to adapt to the specifics of the social situation, studies have shown how humans' expectations about HRI change as they engage in interactions with robots (Kwon et al., 2019; Tuncer et al., 2023) and how they adjust their own behavior during the interaction (see Pelikan & Broth, 2016; Tuncer et al., 2023). Human participants often rely on experience of their own social interactions when making sense of the robot's behavior. This includes an orientation to routine normative orders of coordination in talk (Fischer, 2011a; 2011b) as well as to the display of emotions (Fischer et al., 2019). Human participants seem to take into account emotional aspects of interaction in HRI, and they often make sense of emotional displays by the robots to make proper responses (Pelikan et al., 2020). As Pelikan et al. (2020: 468) conclude, "people make sense of the emotion displays in relation to preceding actions and treat them as projecting specific ways to continue the interaction".

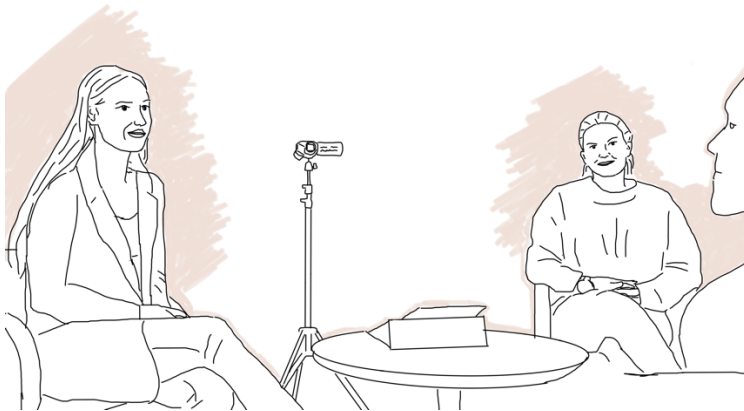
In this study, in line with the studies mentioned above, we will examine how human subjects make sense of the robot's behavior. In addition, we will highlight how human subjects, regardless of their linguistic backgrounds, can immediately react to breaches of normative expectations in talk, and display emotions and stance regarding those breaches in HRI. Such stance taking, we argue, also highlights the contribution of the robot, and thus its agency, in the emergence of trouble in HRI.

4. Data

In this study, we analyze video-recorded interactions with Furhat, a humanoid robot in the shape of a bust, that is, head, chest, and shoulders. Furhat's face is a projection which allows facial movements and variation in the appearance of the face. Through the projection system, the movement of the eyes and lips are visible. Moreover, the robot is able to move its head and thus direct its face and gaze toward a particular co-participant.

The set-up of the interactions consists of a language game designed for Furhat to interact with two human conversational partners. In one set of conversations, both human participants speak Swedish as L1, while in the other set one participant speaks Swedish as L1 and the other speaks Swedish as L2. Even if all participants were, to some extent, familiar with and may have used agentive technology in their daily life, this particular situation, that is, playing a game with Furhat, was new to all participants (although a fraction of the participants reported that they interacted with robots regularly). For this study, we analyzed 15 sessions, 13 with an L1-L2 pairing and 2 sessions with an L1-L1 pairing. The configuration of participants in the set-up is shown in Figure 1.

Figure 1. *Furhat interacting with two human participants (Furhat to the right)*



The whole conversation can be divided into three phases: (a) greetings and introductions; (b) game instructions; and (c) game play. Our study focuses on the first phase, where the robot initiates the interaction by taking the floor and inviting the participants to introduce themselves; all examples come from this phase. The reason to focus on the first phase is that it is characterized by the scripted behavior of the robot, which unfolds and pans out similarly in all sequences and which, from an analytic perspective, allows for comparing the sequences to each other. In this phase, Furhat's actions follow a script, with the researchers allowing the script to proceed through a Wizard of Oz system¹ in only two places in conversation. Excerpt 1 below is a transcribed rendition of the script on which Furhat relies during conversations. The two places where the researcher intervenes can be found after lines 09 and 13, that is, after the first and the second set of two questions ("what's your name?" and "and where do you come from?"; see lines 07–09 and 11–13) that are addressed to each human participant in turn. These are the designated places (lines 10 and 14) in the design of the script where human participants are expected to take the floor and introduce themselves. Specific annotations are used to show gaze (gz) and other embodied conduct, for example, hand movements (h), facial expressions (fe), and closed eyes (ce). Furhat's gaze symbol is a delta sign (Δ) and the remaining symbols are used for human participants (*, ∞ for showing gaze, and, ‡, \diamond for other bodily movements. Particularly, for head movement, we use the symbol \cap). Lines with analytic interests are highlighted by an arrow sign (\rightarrow).

¹The robot works as an automated system and in only two places in conversation does the researcher/experimenter intervene in the talk and take control of the robot's conduct. The experimenter simulating the robot's or the agent's behavior in experimental settings is known as Wizard of Oz experiment/system.

Excerpt 1. *Furhat's scripted contribution to the conversation*

Ex.#01: Furhat's scripted contribution to the conversation

```
01. FUR .hh uhum uhum ((clears throat))
02. (0.6)
03. FUR Δ∩hej∩
    hi
FUR ∩head-turn∩
FUR >>gz-in-btwn-->Δgz-tw-LEFT-PARTY-->
04. (0.6)
05. FUR ja heter furhat
    my name is furhat
06. (1.7)
07. FUR Δ=[va] heter du?
    what's your name?
    FUR ->Δgz--tw-RIGHT-PARTY-->
08. (0.3)
09. FUR och (.) vart kommer du ifrån?
    and where do you come from?
10. (long pause - controlled by the researcher)
11. FUR å Δva heter du
    'n what's your name?
    FUR ->Δgz-tw-LEFT-PARTY-->
12. (0.3)
13. FUR o[ch (.) vart kommer du] ifrån?=
    and where do you come from?=
14. (long pause - controlled by the researcher)
15. FUR trevligt Δå träffas
    nice to meet you
    FUR ->Δgz-tw-in-btwn-->>
16. (1.8)
17. FUR vi ska spela spelet me andra ord (.)
    we're going to play "in other words" ((game's name))
```

As we can see in Excerpt 1, Furhat begins with a designed inbreath before making a sound resembling coughing or clearing the throat. The inbreath and coughing sounds function as a way of drawing the participants' attention; that is, they work as a summons (see Schegloff, 1968) to begin the conversation.

Furhat then begins the conversation proper after a pause of 0.6 seconds (line 02), by saying "hi" and introducing itself ("my name is Furhat", line 05). Furhat is programmed to begin the conversation by moving its head from facing the space between the participants (in lines 01–02) toward the participant on its left while saying "hi" (line 03). Then, after another 0.6 seconds (line 04), it continues to introduce itself (line 05). The script continues with a longer pause of 1.7 seconds (line 06) before Furhat moves its head toward the participant on its right and asks for the participant's name (line 07) and where they come from (line 09). Between these two questions, there is also a pause of 0.3 seconds. The first set of questions ends with the researcher gaining control to prevent Furhat from continuing the conversation until the first participant has answered. This Wizard of Oz method is applied because of the unpredictability of the length of the participant's answer in the response slot in line 10. Once the participant's answer is complete, the researcher (who is sitting in another room watching the conversation) resumes the conversation through a command. After asking two questions to the person on the right, the same set of questions is posed to the person on the left, after which the researcher regains control of the robot (line 14); when the replies are produced, Furhat resumes the conversation by

command. This phase of the conversation comes to an end with the robot saying “nice to meet you” (line 15). The robot then, after a pause of 1.8 seconds (line 16), moves on to instruct the participants about the game (line 17).

5. Data Analysis

In what follows, we will analyze how the design of Furhat’s introductory script leads to interactional troubles in (a) the coordination for turn transition (Sacks et al., 1974); and (b) the projectability of the trajectory of ongoing actions (Mondada, 2021). The transcription follows Jefferson’s (2004) conventions for verbal exchanges and Mondada’s (e.g., 2016) for multimodal annotations. In what follows, we will begin chronologically with the analysis of the greeting sequence in our data (section 5.1). We then analyze the first set of introduction questions that the robot poses to the person on the right (section 5.2) before turning to the person on the left to ask the same set of questions (section 5.3).

5.1 Normative order of ordinary conversation and the emergence of overlap

In section 5.1, we aim to uncover how the robot’s behavior may elicit overlapping responses from human participants and how these overlaps are oriented to by participants as breaching the normative order of ordinary conversation. In the first section, we focus on the reciprocation of greetings. In section 5.2, we examine troubles that emerge in the coordination of turn transitions.

5.1.1 Delayed reciprocation in the exchange of greetings

The design of Furhat’s scripted turns in the first phase of the conversation (greeting-introduction) seems to follow a one-TCU-one-turn principle. After each TCU, a pause is inserted. The pauses are often longer than inter-turn pauses in ordinary human-human interactions (the median of short gaps in and between turns in everyday conversation is 100–300 ms, see Levinson & Torreira, 2015) and thus may potentially be treated as a TRP by the conversational partners. For instance, after the first greeting “hi” (line 03, Excerpt 1), there is always a long pause of 600 ms (line 04, Excerpt 1). Likewise, after Furhat has introduced itself (line 05, Excerpt 1), there is also a long pause (line 06, Excerpt 1), which can be interpreted as inviting the initiation of a reciprocal action (see e.g., Excerpt 2, line 07). However, in none of our 15 cases do human participants use the pause in the robot’s dialogue system in line 04 to reciprocate Furhat’s greeting. In 11 cases, human participants do reciprocate, but this is not done immediately after its greeting; instead, a greeting is produced either (nine cases) after Furhat introduces itself (line 05) or (two cases) after the first set of questions posed by Furhat. In four cases, a reciprocation of Furhat’s greeting is never produced (even

if some signs of nonverbal attempts to take the floor may be observed; see below).

Below, we present two of those nine cases in which the human participants' bid for taking the floor comes after Furhat's introduction (line 05) when there is a long pause of 1.7 seconds in the robot's dialogue system.

Excerpt 2.

Participants: Livia (LIV), Stefan (STE), Furhat (FUR)

EX.#02

Participants: Livia (LIV), Stefan (STE), Furhat (FUR)

```
01. FUR .hh uhum uhum ((clears its throat))
02.     (0.6)
03. FUR →                Δhhejh
                        hi
                        FUR                hhead-turnh
                        FUR                >>gz-in-btwn-->Δgz-tw-STE-->
04.     (0.6)
05. FUR ja heter furhat
      my name is furhat
06.     (1.6)
07. STE → †tsk he[j ]
          tsk hi
          †smiles-->
08. FUR Δ=[va] heter du?
          what's your name?
      FUR ->Δgz--tw--LIV-->
09. LIV → he[j]∞
      STE ∞gz-tw-LIV-->
10. FUR [o]ch (.) ∞vart kommer du ifrån?†
          and where do you come from?
      STE ->†
      STE ->∞gz-back-to-FUR-->
```

In the example above (Excerpt 2), Stefan produces a response to the greeting but only after Furhat has introduced itself. In terms of adjacency pairs (Sacks et al., 1974), it seems Stefan's "hi" in line 07 is produced late, after 1.6 seconds (line 06), compared to a regular response to greetings in everyday conversations, which would normally be produced with minimal delay (see Sacks et al., 1974). The second greeting, from Livia, is produced when Furhat turns to her and asks her name (line 08). A greeting in response to a question does not answer the question and could only mean that Liv begins her turn with the greeting as a bid to take the turn. In an overlap with Liv's turn, however, Furhat continues holding the floor, expands its own turn, and poses a second question (line 10).

The next extract is also one of the nine cases where human participants greet Furhat after the self-introduction in line 05.

Excerpt 3.

Participants: Olva (OLV), Lilian (LIL), Furhat (FUR)

Ex.03

Participants: Olva (OLV), Lilian (LIL), Furhat (FUR)

```
01. FUR .hh uhum uhum ((clears its throat))
02.     (0.6)
03. FUR → Δhhejh
           hi
           FUR     hhead-turnh
           FUR >>gz-in-btwn-->Δgz-tw-LIL-->
04.     (0.6)
05. FUR ja heter furhat
           my name is furhat
06.     (1.4)
07. OLV → [hej]=
           hi
08. LIL → [hhejh]
           hi
09. FUR Δ=va heter du?
           what's your name?
           FUR ->Δgz--tw--OLV-->
```

Like in the previous example, the human participants in Excerpt 3 greet Furhat with a delayed response, as it was not produced in the expected sequential position as a second pair part to the first greeting in line 03, but only after Furhat's self-introduction (line 05). Here, both participants reciprocate the greeting in overlap (lines 07–08).

In two out of our 15 cases, the addressed human participant greets Furhat only after the first set of questions is posed and not earlier (compared to the nine abovementioned cases). In the four remaining cases where there is no greeting, there are obvious cues that indicate that human participants are about to initiate a turn when Furhat greets and introduces itself. These cues include nonverbal and embodied actions such as initial inbreath, mouth opening and even hearable chuckles or laughter. The next example is one of these cases.

Excerpt 4.

Participants: Emil (EMI), Albina (ALB), Furhat (FUR)

Ex.04.

Participants: Emil (EMI), Albina (ALB), Furhat (FUR)

```
05. FUR .hh uhuh uhuh ((clears its throat))
06. (0.6)
07. FUR Δhej
      hi
      FUR Δhead-turn
      FUR >>gz-in-btwn->Δgz-tw-ALB-->
08. (0.6)
09. FUR ja heter furhat
      my name is furhat
10. (0.4)
11. EMI hh. ((giggles))
12. ALB Δhihihi ((chuckles))#
      #IMG02
      ALB Δtilts her head; ce; opens mouth with a broad smile-->
13. FUR Δva heter du?
      what's your name?
      Fur ->Δgz--tw-EMI-->
      ALB ->Δ
```

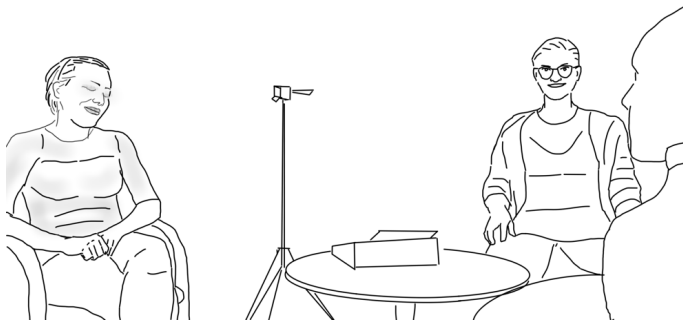
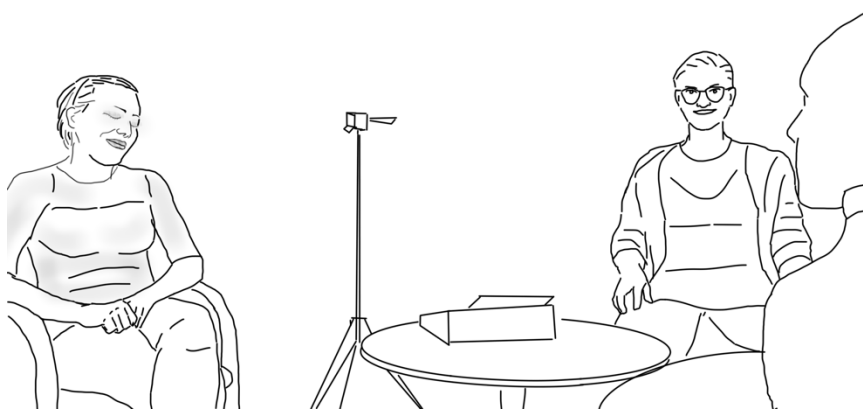


Image2: After Furhat's greeting and self-introduction, Albina chuckles and closes her eyes

Figure 2. After Furhat's greeting and self-introduction, Albina chuckles and closes her eyes



In the example above, although no response is produced, neither after Furhat's greeting (line 07), nor after its self-introduction (line 09), Albina shows some

readiness to talk by tilting her head, chuckling, closing her eyes (Figure 2), and also opening her mouth (line 12). At this point, however, Furhat turns toward Emil and poses the first question to him (line 13).

As shown in the excerpts above (Excerpts 02–04), as in all 15 cases, after the robot produces its greeting (line 03), and after a pause of 0.6 seconds (600 ms), Furhat continues by introducing itself. As none of the 15 participants produces the second part of the greeting after line 03, it is safe to say that this interactional space is generally not treated as appropriate to return the greeting. Nonverbal aspects of Furhat’s behavior may contribute to this treatment. Before Furhat begins to talk, its gaze is directed toward the space between the participants (line 07, Excerpt 4). Exactly at the same time as it says “hi”, the robot turns toward the participant on its left (e.g., Albina in Excerpt 4, line 07). Its head movement is only completed after the greeting term “hi” is produced and finished. Producing the greeting while the head is turned toward a participant does not seem to be enough for the addressee to reciprocate the greeting.

On the whole, in all cases where a response to Furhat’s greeting was either produced with delay, after Furhat’s self-introduction (in 11 cases), or not produced at all (in 4 cases), there are accounts internal to the context of interaction: all 15 cases point to the fact that in the greeting sequence, the human participants’ expectations are not met. The robot’s head turn and greeting do not tally with the expected order in everyday conversation, where the face-to-face configuration of the interactants is expected to be established before the greeting is produced (see e.g., Nilsson et al., 2018). Here, the robot’s greeting is made *while* it turns its head toward the addressed participant. This causes confusion in terms of how the addressed participant should proceed, which leads to delayed greetings (with long pauses after Furhat introduces itself, see e.g., Excerpt 2 and 3), or no greeting at all (Excerpt 4).

5.1.2 The problem of coordination of turn transitions

As evident in the above excerpts, Furhat launches its greeting and the question about the participant’s name, regardless of whether its greeting has been reciprocated (line 07 in the scripted talk, see Excerpt 1). To recap, Furhat initiates a greeting with the person on its left (line 03, see Excerpts 1–3), introduces itself (line 05, see Excerpts 1–3) and subsequently turns toward the human participant on the right, simultaneously beginning to ask: “what’s your name?”. This action evokes responses from human participants in at least six cases out of 15. Both the robot’s verbal and nonverbal behavior provides for the relevance of a response after line 07; the change of gaze direction along with the direct question function as a turn-allocation practice (Sacks et al., 1974). The point just after the question “what’s your name?”, where there is also a scripted pause (line 08, Excerpt 1), is treated in six cases as a TRP. However, after 0.3 seconds of silence, Furhat is programmed to continue with a second question formulated

with an initial “and”: “and where do you come from?” (line 09, see Excerpts 1–4), even though its immediate prior actions—turning toward the person on the right and asking their name followed by a short pause – mark a relevant place for the coordination of a turn transition. Following the scripted nature of Furhat’s talk in this phase of the activity, the delivery of the second question is continued regardless of whether the participant produces a response directly after the first question, and regardless of any other displays of response initiation (e.g., the participant opening their mouth or moving their torso forward, demonstrating readiness to talk, etc.).

In all six cases in which the participants produced a response to the first question, an inevitable overlap was caused by Furhat’s continuation with the second question.

Excerpt 5

Participants: Livia (LIV), Stefan (STE), Furhat (FUR)

Ex.#05.

Participants: Livia (LIV), Stefan (STE), Furhat (FUR)

```

08. FUR →      Δ=[va ] heter du?
                what's your name?
    FUR      ->Δgz--tw--LIV-->
09. LIV → he[j]∞
    STE      ∞gz-tw-LIV-->
10. FUR →      [o]ch (.) ∞vart kommer du ifrån?‡
                and           where do you come from?
    STE
    STE      ->‡
                ->∞gz-back-to-FUR-->
11.          (0.4)
12. LIV      ja∞ heter livia och ja kommer från ∞ungern,
                my name is livia and I come from Hungary,
    STE      ∞gz-tw-LIV-----∞,,, ((gz-tw-Fur))
13.          (1.0)

```

As shown in this example, after the question “what’s your name?” (line 08), Livia starts to respond with a greeting (line 09), which is produced in overlap with the robot’s delivery of the second question (“and where do you come from?”, line 10). Livia however responds (line 12) to both questions (posed in lines 08 and 10) after Furhat has finished its turn. A similar pattern can be observed in the next example, where Sara provides the answer (line 10) to Furhat’s first question (line 08) in overlap with Furhat’s second question (line 11).

Excerpt 6

Participants: Sara (SAR), Tom (TOM), Furhat (FUR)

Ex.#06.

Participants: Sara (SAR), Tom (TOM), Furhat (FUR)

```
08. FUR → Δ[va ] heter du?  
           what's your name?  
           Fur ->Δgz--tw--SAR-->  
09.       (0.3)  
10. SAR → [sara ]  
11. FUR → [och (.)] vart kommer du ifrån?  
           and     where do you come from?  
12.       (0.9)  
13. SAR   från stockholm  
           from stockholm  
14.       (0.7)
```

As observed in both examples above (Excerpts 5 and 6), human participants produce turns at talk and allocate the turns in HRI (at least in this initial phase of the conversation) as they normatively would do in everyday interactions (Sacks et al., 1974). In other words, the pause after Furhat's first question (line 08, Excerpt 6) is treated by human participants as a TRP, as the question is a syntactically and pragmatically complete TCU. However, the scripted continuation of the talk by Furhat, after a short pause of 0.3 seconds, leads to overlap between the response to the first question (line 10, Excerpt 6) and the posing of the second question (line 11, Excerpt 6).

In sum, we have shown so far how human participants' delayed responses (in 11 cases), or the lack of response altogether (in four cases), as well as their attempts to answer Furhat's first question (in six cases) can be explained by considering the normative order of conversation, such as the conditional relevance of providing a response to immediately preceding question (see Sacks et al., 1974).

5.2 Turns and sequence type as resources of projection

In this section, we aim not only to highlight the emergence of overlaps in the studied interactions, but also more concretely to point to the resources used for the projectability of the next relevant action (which have led to the emergence of overlaps).

In what follows, we present the continuation of the conversation after Furhat has asked for the name and the origin of the participant on the right (and note that a response is provided in all 15 cases), and when it turns to the participant on the left to pose the same set of questions. When addressing the person on the left, Furhat initiates with "and" ("and what's your name?", see Excerpt 7 below, line 12). The use of "and" as a connector in turn-initial position signals continuity, projecting that the new action it prefaces ties to the previous action (see also Mazeland, 2013, on *en*, 'and', in Dutch). Indeed, after the *and* preface, the first question about the participant's name is formulated with exactly the same

linguistic resources as the first question in the first set (“and what’s your name?”). With that connector and the verbal repetition, a general logic of an interactional pattern is presented; the pattern is based on the logic that the first set of questions as a request for self-introduction constitutes a precedent for the second set. In other words, after Furhat asks for the name and the origin of the first participant, turning to the next participant and producing the first question already projects what Furhat’s next move may be. This projection is observably recognized by the participant in all 15 cases; as they promptly begin to reply immediately after the first question (“and what’s your name?”). They do not wait for Furhat to produce its second question. Without any exception, in all 15 cases, the human participants begin to produce an answer to the first question. And in all 15 cases, Furhat’s transition from one question (“and what’s your name?”) to the next (“and where do you come from?”) overlaps with the responses provided by human participants to the first question. For example, observe the overlap in lines 14 and 15 in Excerpt 7.

Excerpt 7

Participants: David (DAV), Erik (ERI), Furhat (FUR)

Ex.#07.

Participants: David (DAV), Erik (ERI), Furhat (FUR)

```

12. FUR   å va ~heter du ((asking the participant on the left))
          'n what's your name?
          ERI   ->°gz-tw-FUR-->>
13.      (0.3)
14. ERI → ◊[ja hete-]
          my name i-
          ERI   ◊smile-->
15. FUR → [och (.) ]*vart kommer du i[frå]n?=
          and      where do you come from?=
          DAV     ->*gz-tw-ERI-->

```

When Furhat turns its head, directs its gaze toward Erik, and issues the question “and what’s your name?” (line 12), after a short pause of 0.3 seconds (line 13), Erik begins to respond (line 14). However, as dictated by the script, Furhat continues to ask the second question (line 15). As a result, an overlap between lines 14 and 15. This pattern is consistently observed in all 15 cases, including the following example in which the recipient (Osborn, line 15) not only answers the question about his name, but keeps the floor, likely anticipating the second question and beginning to provide information about his country of origin: “n I co-” (line 15, a response to a second question which is not yet been asked but appears in line 16).

Excerpt 8

Participants: RITA (SAR), OSBORN (OSB), Furhat (FUR)

Ex.#08.

Participants: RITA (SAR), OSBORN (OSB), Furhat (FUR)

```
14. FUR   å va heter du
          'n what's your name?
15. OSB → *ja h[eter osborn å ja ko-]
          my name is osborn 'n I co-
RIT ->*gz-tw-OSB-->
16. FUR   [och      (.)      va]rt #kommer du∅ ifrån?=  
          and                where do you come from?=  
OSB                                     #smiles-->  
RIT                                     ∅chuckles-->
17. OSB   ja heter osborn å ja kommer från sverige
          my name is osborn 'n I come from sweden
18.       (*0.8 + 0.3)
RIT ->*gz-tw-FUR-->>
19. FUR   trevligt ∆å träffas
          nice to meet you
Fur      ->∆gz-tw-in-btw-->>
20.       (1.6)
21. OSB   de[samma]
          you too
```

As is shown in Excerpt 8, the human participants' response to the question about their name (line 14) not only addresses that question but also projects the second question (about their origin) even before it is produced (the one that Furhat produces in overlap, "and where do you come from?", line 16). The example above (Excerpt 8), along with all the examples in our dataset, point to the fact that the first question-answer sequence provides a resource for the anticipation of what is expected in the second sequence, and the participants act accordingly.

5.3 Two practices to manage overlapping talk

In our dataset, we observe that human participants use two distinctive practices to manage the overlap occurring between the answer to the question "and what's your name?" and Furhat's second question "and where do you come from?". The participant may (a) abort their turn and restart or (b) ignore the overlap and produce a complete reply. The former can be observed in the clear majority of cases (13 out of 15) and the latter in only two cases.

5.3.1 Restarts

In response to the overlap, human participants may cut off their own answer and resume their turn after the next TRP, which occurs after the completion of the second question by Furhat. This shows how Furhat's conduct is consequential and that it is recognized and addressed by human participants.

In the following extract, which is the continuation of the previous example (Excerpt 9), Eric cuts off his utterance ("my name i-", line 14) when his response

co-occurs with Furhat's delivery of the second question (line 15). However, he restarts his turn after Furhat's second question and repeats the same wording: "my name is erik 'n I come from sweden too" (line 16).

Excerpt 9

Participants: David (DAV), Erik (ERI), Furhat (FUR)

Ex.#09.

Participants: David (DAV), Erik (ERI), Furhat (FUR)

```

12. FUR  å va  heter du ((asking the participant on the left))
        'n what's your name?
    ERI  -> gz-tw-FUR-->>
13.      (0.3)
14. ERI  →  [ja hete-]
        my name i-
    ERI   smile-->
15. FUR  →  [och (.) ]* vart kommer du i+[fr n]?=
        and      where do you come from?=
    DAV  ->*gz-tw-ERI-->
16. ERI  →  ja heter erik   ja kommer fr n  sverige* ocks 
        my name is erik 'n I come from sweden too
    ERI  -> 
    DAV  ->*gz-tw-FUR-->>
17.      (1.4)

```

The same happens in the following example, in which Furhat's second question "and where do you come from?" co-occurs with Leonie's response to the first question. Leonie cuts off her response (line 16) and restarts (line 18), repeating the answer to the first question before answering the second question.

Excerpt 10

Participants: Elias (ELI), Leonie (LEO), Furhat (FUR)

Ex.10

Participants: Elias (ELI), Leonie (LEO), Furhat (FUR)

```

15. FUR     va heter du?
        'n what's your name?
    FUR  -> gz-tw-LEO-->
16. LEO  →  a: [ja heter leon-]
17. FUR  →  [och (.) *vart] kommer du  ifr n?=  

        and      where do you come from?
    ELI  ->*gz-tw-LEO-->
    ELI  ->#broad smile-->
18. LEO  →  =e ja heter leonie   ja kommer fr n tyskland
        e my name is leonie 'n I come from germany
19.      (1.3)

```

5.3.2 Ignoring the overlap

Another strategy that human participants use is to fully provide the answer to Furhat's first question and ignore the overlap caused by Furhat's production of the second question. Even if such an overlap could jeopardize mutual understanding, in neither of the cases when such a practice is used do human participants fail to hear and respond to Furhat's second question, despite its production in overlap with their own answers. The following example shows how one of those cases unfolds.

Excerpt 11

Participants: Tove (TOV), Sophie (SOP), Furhat (FUR)

Ex.#11.

Participants: Tove (TOV), Sophie (SOP), Furhat (FUR)

```
13. FUR  å va heter *du? ((asking the participant on the left))
        'n what's your name?
        TOV                                ->*gz-tw-SOP-->
14. SOP → ja he[te:r e: sophie]
        my name i:s e: sophie
15. FUR → [och (.) #*vart ] kommer du ifrån?*
        and where do you come from?
        TOV                                ->*gz-tw-FUR-----*gz-tw-SOP-->
        TOV                                †smiles and nods-->
16. SOP  och ja kommer från tyskland
        and I come from germany
```

In the example, when Sophie is addressed by Furhat (line 13), she completes her turn despite the overlap (line 14) and even ties her response to Furhat's second question to the answer she provided in response to the first question, with the help of the connector "and": "and I come from germany" (line 16).

5.4 Overlap as a breach of interactional norms

As mentioned above, in all 15 cases, the second set of questions is characterized by an overlap between the participant's replies to the first question ("and what's your name?") and the second question ("and where do you come from?"). In all cases, human participants treat the overlap as a violation of coordination for transition in turn-taking and also as a violation of normative expectations concerning the projectability of an ongoing action. By violation of expectations, we mean that the human participants treat the overlap as an extraordinary event which was not expected to happen. The violation is understood by virtue of the ways in which the human participants manage the overlaps. Treating the overlap as a violation of normative expectations is also evidence of treating Furhat as an agentive entity in interaction whose conduct is consequential in terms of contributing to the organization of order in HRI. In the previous section, we

showed that, in 13 cases out of 15, the human participants cut off their unfolding turns and restarted in the next possible TRP (which is after Furhat's second question "and where do you come from?"). In all 15 cases, including those two cases where the human participants ignore the overlap, there are embodied actions indicating that either one or both human participants take a stance on the overlap and the failure of the smooth transition of speakership. The human participants seem to react to the robot's interruption of the delivery of the turn in progress, something they evidently display in all 15 cases through embodied conduct, showing a stance of embarrassment, amusement, or frustration over the violation of normative expectations of ordinary conversation. Changes of facial expressions, for instance, occur in all 15 cases in connection with the overlap and its management. Such changes include shifting from a serious facial expression to smiling (often broadly) in eight cases, hearable laughter in five cases, doing a "surprise face" through protruding lips and raised eyebrows in one case, and in two other cases drawing down the corner of the lips. In some cases, frustration is displayed not only through a broad smile, chuckles, or other types of lip movement (e.g., lip parting. See various studies on laughter, e.g., Glenn & Holt, 2013; Petitjean & González-Martínez, 2015), but also through closing eyes or through a gesture very similar to eye-rolling.

In the following example, when receiving the first question ("and what's your name?", line 13), Regina immediately begins to respond (line 14). Her response, however, overlaps with Furhat's second question ("and where do you come from?", line 15). Regina immediately reacts to the overlap with a facial expression and what resembles an eye roll (Figure 3).

Excerpt 12

Participants: Regina (REG), Marlon (MAR), Furhat (FUR)

Ex.#12

Participants: Regina (REG), Marlon (MAR), Furhat (FUR)

```
13. FUR  å Δva heter du
        'n what's your name?
        Fur  ->Δgz-tw-REG-->
14. REG  *j[a heter ]
        my name is
        MAR->*gz-tw-REG-->
15. FUR  [och (.) ] vart∞ kom*mer du ∞‡◊ifrå◊*n?=#
        and      where do you come from?=
        REG                      ∞gz-away/fe--∞ce-->
        REG                      #smile-->
                                   #IMG03
        MAR                      ->*gz-tw-FUR----*gz-tw-REG-->
        MAR                      ◊raises eyebrows◊
16. REG  ◊ja heter ∞regina  å ja kommer från sverige,‡
        my name is regina 'n I come from sweden,‡
        REG          ->∞gz-tw-FUR-->>
        MAR  ◊chuckles/smiles-->
        REG
        MAR
17.      (0.6)
18. FUR  trevligt Δå träffas◊
        nice to meet you
        Fur  ->Δgz-tw-in-btw-->>
        MAR          ->◊
```



Image3: Regina does an 'eye roll' and a facial expression, indicating 'trouble' in interaction

Figure 3. Regina does an “eye roll” and a facial expression, indicating trouble in interaction



Immediately after the overlap (line 15), Regina shows a facial expression which could be interpreted as a reaction to an embarrassing event in interaction, exactly at the moment when she was interrupted (line 14). She makes a face, closes her eyes, and smiles (line 15; see eye roll in the display of exasperation or the embodied display of dissent in Clift, 2021), and resumes her turn after Furhat finishes the second question (line 15). She restarts by introducing herself once more and answering the second question (line 16). Her conversational partner, Marlene, chuckles at the event (line 16). Similar ways of indicating embarrassment or frustration are observed in all other cases. Here is another example (Excerpt 13). This time, when a human participant, Tom, is interrupted by Furhat (line 18), he chuckles (line 19), and the second human participant joins in his laughter (line 20).

Excerpt 13

Participants: Sara (SAR), Tom (TOM), Furhat (FUR)

Ex.#13

Participants: Sara (SAR), Tom (TOM), Furhat (FUR)

15. FUR å va heter du
'n what's your name?
16. (0.2)
17. TOM *[ja-]
I
18. FUR [och] (.) vart kommer du i#[från?]=
and where do you come from?=
19. TOM → ‡[hhh.]hihi#
#IMG04
20. SAR → hihi[hi]
21. TOM [ja] heter tom å ja kommer från belgien
my name is tom 'n I come from belgium
22. (1.1)
23. FUR trevligt Δå träffas
nice to meet you
Fur ->Δgz-tw-in-btw-->>
24. (1.7)



Image 04: Tom chuckles toward the end of Furhat's turn

Figure 4. *Tom chuckles toward the end of Furhat's turn*



The reaction toward the overlap which observably hinders the progression of the turn is exhibited first through a cut-off in Tom's attempt to respond to Furhat's first question (line 17). Then, Furhat's second question (line 18) is followed by Tom's chuckles (line 19, Figure 4). Although Tom eventually responds to both questions at once (line 21) after Furhat poses the second question, there are nonverbal reactions as evidence of stance-taking by Tom and Sara, pointing to their display of embarrassment and/or amusement, and perhaps frustration with the

emergence of a trouble in the transition between turns at talk. A similar event occurs in the next example (Excerpt 14).

Excerpt 14

Participants: Linnea (LIN), Klara (KLA), Furhat (FUR)

Ex.#14

Participants: Linnea (LIN), Klara (KLA), Furhat (FUR)

18. FUR Δå va heter du?*
- 'n what's your name?
- FUR ->Δgz-tw-LEO-->
- LIN ->*gz-tw-KLA-->
19. (0.4)
20. KLA [ja heter klara]
21. FUR [och (.) vart k*om]mer du ifrån?
- and where do you come from?
- LIN ->*gz-tw-FUR-->
22. LIN → #he.he. †
- LIN ‡chuckles‡
23. KLA → e: *som sagt ja heter klara å
- e: as said my name is klara 'n
- LIN ->*gz-tw-KLA-->
24. KLA ja kommer ifrån (XX)∅ ut*anför stockholm
- I come from XX outside stockholm
- LIN ->*gz-tw-FUR-->>
- KLA ->∅stops smiling
25. (1.1)
26. FUR trevlit å träffas
- nice to meet you
27. (1.8)

When Furhat asks Klara to introduce herself, upon her introduction Furhat enters into Klara's turn by producing the second question ("and where do you come from?", line 21). This is followed by Linnea's chuckle (line 22) as a reaction to the overlap. As Klara produces the response to both questions (lines 23–24), restarting her turn, she clarifies that she had already introduced herself (which was not attended to by Furhat in the first place). She says, "as said my name is klara 'n I come from (xx) outside stockholm" (lines 23–24). By explicitly marking the provided information as a repetition of something that has already been said, Klara highlights the non-normative character of the produced sequence.

6. Conclusion

There are certain resources that are used for recognizing and anticipating the trajectory of actions: the design of a turn, such as its format (e.g., linguistic forms or embodied movements), the action that it contributes to accomplishing (e.g., sequence type or the type of activity), and the ways in which the turn (and the whole action) is delivered, for instance, the prosody or other embodied features in action production (see e.g., Mondada, 2006 on end of the turn or closing of the sequence). In this study, we have shown how those resources are used by human participants anticipating the trajectory of upcoming action and responding

accordingly. In the analyzed interactions, the human participants' action projections are not consistently met and responded to by the social robot, and this leads to breaches of the normative expectations about the order of ordinary conversation, and thus to troubles in talk. The troubles that we have analyzed in this study are concerned with the coordination and the transition of turns, resulting in the production of overlaps. We have shown how human participants react toward interactional troubles and how these troubles are managed.

In the initial phase of conversation, during the greeting sequence, we have shown how human participants either do not produce any responses to Furhat's greeting or do so with a significant delay. We have shown how the embodied behavior of the robot, namely gazing in between participants, may elicit uncertainty about addressivity. In the first set of questions and answers, we have also shown how the design of questions and the lexico-syntactical elements in the turn (e.g., "what is your name?") are understood as indicating the transition place at least in nine cases out of 15. Furthermore, the analysis highlights the fact that sequence types, such as rounds of introductions, are resources for participants to project the expected next action. In our data, once the participants are exposed to the first set of questions, the beginning of the second set of questions ("and what's your name?") arguably makes the whole action recognizable and answerable for the second participant. This is based on the fact that in all 15 cases, the human participants respond to the first question without pause or any hesitation. However, they get interrupted by the robot producing a second question according to its scripted talk. This causes overlaps in interaction in all 15 cases. Although the participants have been exposed to the first set of questions, they, nonetheless, do not wait for Furhat to complete the second set of questions. A possible explanation is that the normative expectations of ordinary conversation, which here is responding according to the action projection, override any idiosyncratic characteristics of the robot's behavior. Although Furhat's scripted talk could give hints to the participants that they have to wait for Furhat to complete its turn before they are given the opportunity to talk, the human participants nonetheless start their turns in the midst of Furhat's turn. In other words, as the human participants recognize the introduction-round in the first set of questions, they respond to it without the second set of questions being complete. We could also conclude that by orienting to the sequence type of introduction-rounds as a resource for the projection of the next relevant action, the human participants treat the exact repetition of the second set of questions as unnecessary and therefore unexpected.

On the whole, the study demonstrates how the projectability of turn-completions, which usually is naturally used for the minimization of gaps and overlaps in human-human talk (Sacks et al., 1974), may sometimes get lost in talk with robots. In the studied sequences, the robot's action-designs do not allow early turn-transitions when an action is recognized in the midst of its production and/or when the completion of a turn is projected by human participants. In our analyzed data, this leads to overlaps to which the human participants in all our 15 cases

react by displaying some form of marked stance, for instance, through laughter and/or various facial expressions. These displays are done together with the use of one of two practices to manage the overlap: cutting short the turn in response to the question followed by a restart at the next TRP (13/15 cases) or continuing with the delivery of their turn, thereby ignoring the overlap (2/15 cases).

One of the aims of the design of the word game with Furhat was also to study the possible differences between L2 and L1 speakers in interaction with the social robot. However, from an interactional perspective, in our analysis we have found that both Swedish and non-Swedish participants interacted with the robot according to the normative expectations of ordinary conversation. Independent of the language being a first or a second, it seems that the underlying pattern of order in conversation leads to a specific set of expectations that individuals display in interaction regardless of the partner's human or non-human nature.

The results of the study point to routine conversational practices of projection and participants' orientations toward normative expectations of ordinary social interactions in human-robot interaction. This also has implications for robot design. First, the study re-asserts what previous research has already demonstrated (e.g., Fischer, 2011b; Gillet et al., 2021; Tuncer et al., 2022; Pelikan & Broth, 2016): that is, the significance of attending to nonverbal behavior, both in terms of for instance, the robot's head movement and gaze directions, and also the significance of the robot's ability to interpret, and respond to, human subjects' embodied actions. Second, scripted talk in HRI, if intended to follow the human turn-taking system, should also attend to the details of naturally occurring conversations. For instance, the sequential placement of pauses is crucial in interaction, because the participants may or may not treat them as TRPs. Third, with regard to human participants' ability to project the upcoming actions, as we have shown in our study, routine activities such as introduction rounds are managed with reference to normative expectations regarding how those routines are normally carried out. Therefore, the repetitions (e.g., repeating the same set of questions verbatim in introduction rounds) could easily be predicted, and responded to, by human participants. Therefore, the scripted behavior of the robot could lead to unnecessary overlaps and interruptions, to which the human participants react in various marked ways (e.g. displaying embarrassment, frustration and/or amusement). Fourth, and more importantly, human participants' reaction to the emergence of overlaps shows that they take a stance on the trouble in talk in HRI and thus recognize the contribution of the robot to the emergence of the interactional troubles. Although we cannot be certain that the stance human participants take is only toward the accountability of actions made by the robot itself (and thus its agency), the agency of the robot in the construction of the action and its contribution to the emergence of the trouble is recognized by the participants. Future research is encouraged for further investigating practical implications of the recognition of the robot's agency regarding the details of practices used in HRI from a socio-interactional perspective.

Acknowledgements

We greatly appreciate the comments we received on the earlier version of this paper from the two anonymous reviewers. This study was planned as a pilot project funded by Stockholm University (grant nr. SU FV-2.1.1-0086-19) and is now part of the ongoing project 'Interaction with Social Robots for Education: Robot-Assisted Learning for Students with Diverse Language Backgrounds,' funded by The Swedish Research Council (grant nr. 2022–03265).

References

- Clift, R. (2021). Embodiment in dissent: the eye roll as an interactional practice. *Research on Language and Social Interaction*, 54(3), 261–276. DOI: 10.1080/08351813.2021.1936858.
- Fischer, K., Jung, M., Jensen, L. C. & aus der Wieschen, M. V. (2019). Emotion Expression in HRI – When and Why. Conference Proceedings: *14th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, 29–38.
- Fischer, K. (2011b). How People Talk with Robots: Designing Dialogue to Reduce User Uncertainty. *AI Magazine*, 32(4), 31–38. DOI:10.1609/aimag.v32i4.2377.
- Garfinkel, H. (1967). *Studies in Ethnomethodology*. New Jersey: Prentice-Hall.
- Gillet, S., Cumbal, R., Pereira, A., Lopes, J., Engwall, O. & Leite, I. (2021). Robot gaze can mediate participation imbalance in groups with different skill levels. In *Proceedings of the 2021 ACM/IEEE International Conference on Human-Robot Interaction (HRI '21)*. Association for Computing Machinery, New York, NY, USA, 303–311. <https://doi.org/10.1145/3434073.3444670>.
- Goodwin, M. H. (1980) Processes of mutual monitoring implicated in the production of description sequences. *Sociological Inquiry*, 303–317.
- Jefferson, G. (1973) A Case of Precision Timing in Ordinary Conversation: Overlapped Tag-positioned Address Terms in Closing Sequences, *Semiotica* 9: 47–96.
- Jefferson, G. (2004). Glossary of transcript symbols with an introduction. In G. H. Lerner (Ed.), *Conversation analysis: Studies from the first generation* (pp. 13–31). John Benjamins.
- Kwon, M., Jung, M. F. & Knepper, R. A. (2019). Human Expectations of Social Robots. Conference Proceedings, *The Eleventh ACM/IEEE International Conference on Human Robot Interaction*, 463-464.
- Lala, D., Inoue, K. & Kawahara, T. (2019). Smooth Turn-taking by a Robot Using an Online Continuous Model to Generate Turn-taking Cues. Conference

Proceedings, *ICMI '19*, October 14–18, ACM ISBN 978-1-4503-6860-5/19/10.

- Levinson, S.C., & Torreira, F. (2015). Timing in turn-taking and its implications for processing models of language. *Frontiers in Psychology*, 6:731, doi: 10.3389/fpsyg.2015.00731.
- Mazeland, H. (2013). Grammar in conversation. In J. Sidnell and T. Stivers (Eds), *The Handbook of conversation analysis*, (pp.475–491). Willey-Blackwell.
- Mehan, H. (1979). *Learning lessons: Social organization in the classroom*. Cambridge, MA: Harvard University Press.
- Mondada, L. (2006). Participants' online analysis and multimodal practices: Projecting the end of the turn and closing of the sequence.
- Mondada, L. (2016). Challenges of multimodality: Language and the body in social interaction. *Journal of Sociolinguistics*, 20(3): 336–366.
- Mondada, L. (2021). How Early can Embodied Responses be? Issues in Time and Sequentiality. *Discourse Processes*, 58(4), 397–418.
- Nilsson, J, Norrthon, S, Lindström, J. & Wide, C. (2018). Greetings as social action in Finland Swedish and Sweden Swedish service encounters - a pluricentric perspective. *Intercultura Pragmatics*, 15, 57–88.
- Pelikan, H. & Broth, M. (2016). Why that Nao? How humans adapt to a conventional humanoid robot in taking turns-at-talk. Conference Proceedings, *CHI'16*, May 07-12, San Jose, CA, USA. DOI: <http://dx.doi.org/10.1145/2858036.2858478>.
- Pelikan, H. R. M., Broth, M. & Keevallik, L. (2020). "Are You Sad, Cozmo?" How Humans Make Sense of a Home Robot's Emotion Displays. Conference Proceedings: *HRI '20*, March 23–26, 2020, Cambridge, United Kingdom, 461–470.
- Petitjean, C. & González-Martínez, E. (2015). Laughing and smiling to manage trouble in French-language classroom interaction. *Classroom Discourse*, 6(2), 89–106.
- Glenn, P. & Holt, E. (2013). *Studies of laughter in interaction*. London: Bloomsbury.
- Psathas, G. (1999). Studying the organization in action: Membership categorization and interaction analysis. *Human Studies*, 22: 139–162.
- Sacks, H. (1995). *Lectures on conversation. Volumes I & II*. Blackwell publishing.
- Sacks, H., Schegloff, E. A. & Jefferson, G. (1974) A simplest systematics for the organization of turn-taking for conversation. *Language*, 50(4): 696–735.

- Schegloff, E.A. (1968). Sequencing in conversational openings. *American Anthropologist*, 70(6), 1075–1095.
- Schegloff, E.A. (1986). The routine as achievement. *Human Studies*, 9, 111–151.
- Schegloff, E.A. (2007). *Sequence organization in interaction. A primer in conversation analysis*. Cambridge University Press.
- Schönfeldt, J. & Golato, A. (2003). Repair in chats: A conversation analytic approach. *Research on Language and Social Interaction*, 36(3), 241–284. DOI: 10.1207/ S15327973RLSI3603_02.
- Skantze, G. (2021). Turn-taking in Conversational Systems and Human-Robot Interaction: A Review. *Computer Science & Language*, 67. 101178. DOI: <https://doi.org/10.1016/j.csl.2020.101178>.
- Thomaz, A. L. & Chao, C. (2011). Turn-taking Based on Information Flow for Fluent Human-Robot Interaction. *AI Magazine*, 32(4), 53–63.
- Tuncer, S., Gillet, S. & Leite, I. (2022). Robot-Mediated inclusive processes in groups of children: From gaze aversion to mutual smiling. *Frontiers in Robotics and AI*, 9:729146. DOI: 10.3389/frobt.2022.729146.
- Tuncer, S. Licoppe, C., Luff, P. & Heath, C. (2023). Recipient-design in human-robot interaction: the emergent assessment of a robot's competence. *AI & Society*. DOI: <https://doi.org/10.1007/s00146-022-01608-7>
- Walton, C., Antaki, C. & Finlay, W.M.L. (2020). Difficulties facing people with intellectual disability in conversation: Initiation, co-ordination, and the problem of asymmetric competence. In. R. Wilkinson, J. P. Rae & G. Rasmussen (Eds). *Atypical Interaction: The impact of communicative impairments within everyday talk*. Springer International Publishing AG (pp. 93–127).
- Yamazaki, A., Yamazaki, K., Kuno, K., Burdelski, M., Kawashima, M., & Kuzuoka, H. (2008). Precision Timing in Human-Robot Interaction: Coordination of Head Movement and Utterance. Conference Proceedings: *CHI 2008*, April 5–10, 2008, Florence, Italy. ACM 978-1-60558-011-1/08/04.