Turn-Taking Strategies for Human-Robot Peer-Learning Dialogue

Ranjini Das and **Heather Pon-Barry**
Department of Computer Science
Mount Holyoke College
South Hadley, MA 01075
{das22r,ponbarry}@mtholyoke.edu

Abstract

In this paper, we apply the contribution model of grounding to a corpus of human-human peer-mentoring dialogues. From this analysis, we propose effective turn-taking strategies for human-robot interaction with a teachable robot. Specifically, we focus on (1) how robots can encourage humans to present and (2) how robots can signal that they are going to begin a new presentation. We evaluate the strategies against a corpus of human-robot dialogues and offer three guidelines for teachable robots to follow to achieve more human-like collaborative dialogue.

1 Introduction

Grounding is the process by which two parties coordinate to come to a joint understanding or common ground in a joint project. This involves assuming mutual knowledge, beliefs, and assumptions (Clark, 1996). Since humans use grounding to collaborate in dialogue interactions, robots can look to human grounding patterns to mimic collaboration in a human-like way. In human-robot dialogue with a *teachable* robot, the robot often wants the human to take initiative in presenting material; at the same time, the robot wants to ensure that it can steer the conversation in a natural way. By analyzing a human-human peer-mentoring corpus, we identify turn-level grounding patterns that help achieve these two goals.

First, we observe peer-learning dialogues in a human-human corpus to model how human teachers and learners signal presentation and understanding. In this corpus, both teachers and learners alternately take the floor to offer presentations. While one speaker presents, the other speaker accepts the presentation by displaying evidence of understanding. Our first goal is to understand how a speaker signals to the other speaker to take the floor, such as a teacher encouraging a learner to present an idea, or a learner asking a question that leads the teacher to present an explanation.

Second, speakers may need to shift the floor towards themselves during a conversation. For example, a teacher may have a plan to offer feedback on the learner's work, or a learner may need to explain a problem that confused them. Therefore, our second goal is to understand how a speaker can effectively signal that they are taking the floor.

These two goals are also relevant to human-robot dialogue with a teachable robot: a robot who acts as a peer to a student and prompts the student to teach them the material (Jacq et al., 2016; Lubold et al., 2018b). Because humans engage more deeply with material when they teach it to someone else (Roscoe and Chi, 2007), we want a teachable robot to encourage humans to present material. At the same time, especially when interacting with children, the robot may not always understand or be able to process the human's speech and actions. To handle unexpected, degraded, or out-of-vocabulary input, the robot will sometimes need to take the floor and steer the conversation.

In Section 2 of this paper, we discuss related work. We introduce a human-human peer-mentoring corpus and detail our annotation process in Section 3. In Section 4, we analyze human-human grounding patterns with respect to the two goals: encouraging humans to present, and taking the floor. In Section 5, we introduce and analyze grounding in a corpus of dialogues with a teachable robot. We discuss similarities and differences in the two corpora in Section 6, and offer suggestions for improving human-robot dialogue.

2 Related Work

The **contribution model** of Clark and Schaefer is a widely-used theory of conversational ground-

Proceedings of the SIGDIAL 2018 Conference, pages 119–129,
Melbourne, Australia, 12-14 July 2018.

ing (Clark and Schaefer, 1989; Clark, 1996). The model proposes that collaborative conversations be analyzed in terms of *contribution* units, where each contribution consists of a *presentation phase* followed by an *acceptance phase*. In the presentation phase, Speaker A, the presenter, presents a signal to Speaker B, the acceptor. In the acceptance phase, B, the acceptor, acknowledges that they have understood the signal. This requires positive evidence of understanding from B. The speakers signal back and forth until they have received closure—a sense of mutual understanding.

Traum (1994, 1999) reformulated the contribution model for real-time use by collaborative dialogue agents. In this model, the units of analysis—grounding acts—occur at the utterance level. In human-robot dialogues, Liu et al. (2013) found that incorporating an 'agent-present human-accept' dialogue pattern based on the contribution model into its grounding algorithm led to improved reference resolution. Graesser et al. (2014) used a 'pump-hint-prompt-assertion' dialogue pattern in an intelligent tutoring system, finding learning outcomes comparable to those of human tutors.

Turn-taking in human-robot interaction involves understanding the cues that signal when it is appropriate for a robot to take a turn (Meena et al., 2014). Integrating factors such as robot gaze, head movement, parts of speech, and semantics into turn-taking models is an active area of research (Chao et al., 2011; Andrist et al., 2014; Johannson and Skantze, 2015), informed by studies of turn-taking in human-human dialogue (Gravano and Hirschberg, 2011). In human-human interaction, turn-taking behaviors vary considerably depending on the task. A better understanding of turn-taking in peer-learning dialogue will help inform the design of effective peer-learning robots.

Robot learning companions have the potential to teach broad populations of learners but an important challenge is maintaining engagement and effectiveness over multiple sessions (Kanda et al., 2004). Social robotic learning companions can motivate students, encourage them to persist with a task, and even promote a growth mindset (Park et al., 2017). Recently, *teachable* robots have flipped the traditional teacher-learner roles, with the goal of improving learning and motivation (Hood et al., 2015). Most of the these robots use spoken utterances as output but do not engage in conversational interaction around the human partner's utterances, if any exist. One exception is a robot that encourages students to think aloud, finding greater long-term learning gains when students articulate their thought process (Ramachandran et al., 2018).

Robots that are physically present have advantages over virtually-present robots and virtual agents. For example, in a game-playing setting with children, a co-present robot companion was found to be more enjoyable and have greater social presence than a virtual version of the same robot (Leite et al., 2008). In a puzzle-solving setting, students learned more with a co-present robot tutor than with a virtual version of the same robot (Leyzberg et al., 2012). A survey by Li (2015) found that in 73% of human-robot interaction studies surveyed, co-present robots were more persuasive, received more attention, and were perceived more positively than virtually-present robots and virtual agents. There may be trade-offs to physical presence; in an interview setting, co-present robots were liked better than virtual agents, but participants disclosed less and remembered less with the co-present robot (Powers et al., 2007). Overall, the literature suggests that physically co-present robots are preferable for relationship-oriented tasks, for interaction with children, and for learning.

3 Peer-Mentoring Dialogue Corpus and Annotation

To develop dialogue strategies for a robot peer-learner to effectively shift the conversational floor, we examine the grounding patterns of human peer-teachers and peer-learners.

Corpus. The human-human peer-mentoring dialogue dataset consists of fifty 10-minute conversations, totaling approximately nine hours. Table 1 summarizes the conversation durations and

Peer-mentoring corpus statistics	Median
Dialogue duration (sec)	596.0
Total turns per dialogue	153.5
Teacher turns per dialogue	76.5
Learner turns per dialogue	76.0
Words per teacher turn	8.0
Words per learner turn	3.0

Table 1: Median duration, number of turns, and turn length data for the corpus of human-human peer-mentoring dialogues (N=50).

Grounding label	Definition	Speaker role
Presentation	A signal or piece of information offered by the presenter	presenter
Probe	Questions such as "When are we meeting?", or a signal made without certainty of positive evidence from the other speaker, such as "You know that assignment..."	either
Backchannel	A short turn to signal understanding, such as "Mm-hmm", "Yeah", and in some cases, laughter	acceptor
Uptake	The acceptor's next relevant turn	acceptor
Answer	A signal to display understanding of the presenter's probe	acceptor
Repetition	A signal to confirm understanding	acceptor
Paraphrase	A signal to confirm understanding	acceptor
Closure	Evidence of the conclusion of a joint project	either

Table 2: Definitions of grounding labels and their associated roles.

turn lengths in this dataset. Audio recordings were collected of conversations between undergraduate computer science students as part of a near-peer mentorship program. The mentees were enrolled in an introductory computer science course. The mentors were mid- and upper-level computer science students. Mentors had multiple mentees and met with each mentee individually each week over the course of a semester to give feedback on completed programming assignments. Because mentors received training on giving effective feedback and encouraging mentees to reflect on their work, we assume that all conversations are examples of effective mentoring. The dataset used in this paper is part of a ongoing data collection project with over 250 dialogues.

The audio recordings of the dialogues were manually transcribed by a commercial transcription service. An excerpt below illustrates an interaction between a mentor and mentee, who we will refer to in this paper as 'teachers' and 'learners' (punctuation is added for clarity).

> **TEACHER:** So then you might have like a Point2D trunk start which would then update within that method down below
> **LEARNER:** What do you mean by ...
> **TEACHER:** So like up here instead of putting say like public int tx1 you might write something like—
> **LEARNER:** Oh you mean in uh as a parameter—
> **TEACHER:** Yeah like just put 'public Point2D trunk start' and then you just end it
> **LEARNER:** Yeah yeah I got that

Annotation. Our approach to annotation is motivated by the grounding actions proposed in Clark's model of collaborative dialogue (Clark, 1996), and also by the turn-level unit of analysis in Traum's model (see Section 2). The set of grounding labels, shown in Table 2, is designed to be applicable to both human-human and human-robot corpora. The annotation guidelines and the annotated data are publicly available [1].

In our annotation model, at any time, one speaker has the presenter role, and the other is the acceptor. The roles are associated with a set of grounding actions, which characterize individual dialogue turns. Only the presenter's turns can be labeled as *presentation*[2]. Labels such as *uptake*, *answer*, and *backchannel*[3] typically indicate shorter signals to confirm understanding, and occur in turns by the acceptor. Two labels can occur with both presenters and acceptors: *probe* and *closure*. Each turn is labeled with one or sometimes two grounding labels.

We manually annotated each dialogue turn in the peer-mentoring corpus with one or two grounding labels as well as the identity of the current presenter. This annotation was performed by a single annotator. The counts of each grounding label for teachers and for learners are shown in Table 3. We note that presentation is the most frequent label for teachers, while backchannel is the most frequent label for learners.

[1] http://www.ponbarry.com/PeerLearningDialogueGrounding/

[2] This differs from Clark's model, where contributions in the acceptance phase can also be presentations.

[3] We consider spoken backchannels to be dialogue turns to minimize complexity in the human-robot setting, where we consider all robot utterances to be dialogue turns.

Grounding label	Teacher	Learner
presentation	2475	999
probe	517	507
backchannel	957	1793
uptake	356	701
answer	125	357
repetition	12	26
paraphrase	7	16
closure	205	214
TOTAL	4654	4613

Table 3: Grounding label counts for teacher turns and learner turns in the human-human peer-mentoring corpus.

4 Peer-Mentoring Dialogue Analysis

To support our goal of designing effective turn-taking strategies for a teachable robot, we use the corpus of human-human peer-mentoring dialogues to answer two questions: (1) how do humans encourage their partners to present? and (2) how do humans signal that they are going to shift the floor towards themselves? To frame the decision of whether to focus on teacher strategies, learner strategies, or both, we begin by examining initiative patterns in the corpus.

4.1 Initiative and presentation

Expecting that perceived initiative is closely related to the number of presentation turns, we label each dialogue in the peer-mentoring corpus with a perceived initiative score from 1 to 5 (1=high learner-initiative; 5=high teacher-initiative). We compare the initiative ratings with the count of each speaker's presentation turns as a proportion of their total turns in the dialogue. This is shown in Figure 1. For learners, the proportion of presentation turns is highest when they are perceived to have high initiative. However, teachers present for roughly the same proportion of turns regardless of initiative label. This analysis suggests that learners might assume greater initiative if they are encouraged to present.

4.2 Encouraging partner to present

To analyze how one speaker encourages their partner to present, we consider two cases: (a) when the partner does *not* currently have the floor, and (b) when the partner does currently have the floor.

To understand how human mentors and mentees encourage their partners to present when that partner **does not hold the floor** (i.e., to take the floor), we identify all turns with a presentation label that are at the start of a floor shift. A floor shift occurs when a presentation turn shifts the presenter role from one speaker to the other. We examine what the partner's grounding label was in the preceding turn. In other words, if Speaker B has taken the floor by beginning a presentation, what was Speaker A's last grounding action? An annotated example exchange is shown below.

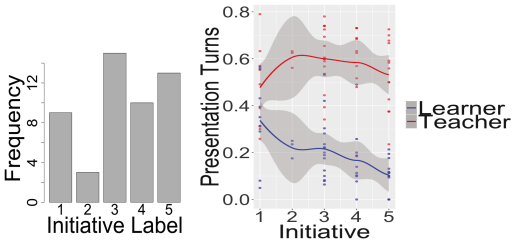

Figure 1: (left) Distribution of initiative labels. (right) Proportion of presentation turns in the conversation compared with conversation initiative.

> **A:** But don't put it off because it's a big project *(presentation)*
> **B:** I can tell cause it's broken down into two parts *(uptake/presentation)*
> **A:** Mh-mmm *(backchannel)*

We find that when a speaker takes the floor, their partner is most frequently presenting in the preceding turn: 0.554 and 0.618 for teachers and learners, respectively. The next most frequent grounding label is probe (see all values in Table 4, section (a)).

To understand how human mentors and mentees encourage their partners to present when that partner **already has the floor** (i.e., to continue presenting), we identify all turns with a presentation label that are *not* at the at start of a floor shift. We examine what the partner's grounding label was in the preceding turn. In other words, if Speaker B already has the floor and then has a presentation turn, what is Speaker A doing before B's presentation that encourages B to continue to present? An annotated example exchange is shown below.

> **B:** It'll be the same problems *(presentation)*
> **A:** Mh-mmm *(backchannel)*
> **B:** So you should prepare in the same way you did last semester *(presentation)*

	N	present.	probe	backch.	uptake	ans.	clos.
(a) Encourage presentation - at shift in floor							
Grounding by T before partner presentation	139	**0.554**	0.266	0.122	0.000	0.035	0.024
Grounding by L before partner presentation	136	**0.618**	0.162	0.140	0.015	0.050	0.015
(b) Encourage presentation - no shift in floor							
Grounding by T before partner presentation	995	0.089	0.125	**0.542**	0.181	0.035	0.024
Grounding by L before partner presentation	2453	0.046	0.104	**0.604**	0.166	0.056	0.015
(c) Signal a shift in floor							
Grounding by T at floor shift	136	1.00	0.132	0.007	**0.596**	0.257	0.007
Grounding by L at floor shift	139	1.00	0.115	0.007	**0.547**	0.317	0.014

Table 4: Normalized frequencies of grounding turn labels for teachers (T) and learners (L); for (a) grounding preceding a presentation by partner at a shift in floor, (b) grounding preceding a presentation by partner, with no shift in floor, and (c) grounding accompanying a presentation at a shift in floor. Presentations are most frequent for (a), backchannels are most frequent for (b), and uptakes are most frequent for (c), as indicated by bolded values. Paraphrases and repetitions have values < 0.01 and are omitted from the table.

When there is no floor shift, we find, unsurprisingly, that the most frequent grounding label preceding presentation turns is a backchannel: 0.542 of the turns for teachers, 0.604 of the turns for learners. The next most frequent labels are uptakes and probes (see all values in Table 4, section (b)).

This data suggests that a robot should consider presenting or probing to encourage a partner who does not have the floor to present, and should consider backchannels to encourage a partner who already has the floor to continue presenting. We note, however, that the overall label frequencies are a factor. After considering next-turn probabilities conditioned on the preceding labels, we expect that probes might be more effective than presentations at encouraging a partner to take the floor.

4.3 Signaling taking the floor

To understand how human mentors and mentees naturally take the floor and become the presenter, we look at the grounding labels of dialogue turns at shifts in the conversational floor. All floor shifts begin with a *presentation* turn; most also have a second grounding label. If there is no accompanying grounding label, we report the grounding label of the speaker's previous turn.

We find that when a speaker takes the floor, the grounding label most frequently accompanying the presentation label is uptake: 0.596 and 0.547 for teachers and learners, respectively. The next most frequent grounding labels are answer and probe (see all values in Table 4, section (c)). This suggests that a robot that wants to take the floor might consider an uptake, answer, or probe in conjunction with their presentation.

5 Comparison with Human-Robot Dialogue Interaction

To understand if the grounding strategies we observed in the human-human corpus are effective in human-robot interaction, we perform a preliminary empirical analysis using dialogue data from a teachable robot interaction experiment conducted in a Wizard-of-Oz (WOZ) style. Section 5.1 describes the dialogue data; Section 5.2 presents our empirical analysis.

5.1 Human-robot dialogue data

The human-robot dialogue data consists of transcripts from a teachable robot interaction experiment where the robot was operated by a human Wizard. In this WOZ experiment, human students interacted in a learning-by-teaching context (Ploetzner et al., 1999) with Nico, a social, teachable, NAO robot. The human participants were peer teachers while Nico behaved as a peer learner, working to solve mathematics word problems.

The human-robot corpus includes dialogue transcripts from twenty college-age participants who each engaged in four problem-solving dialogues with Nico in the WOZ experiment (Chaffey et al., 2018). Table 5 summarizes the dialogue durations and turn lengths in this human-robot dialogue corpus.

The WOZ experiment aided in the development of an autonomous version of the teachable robot

Human-robot corpus statistics	Median
Total turns per dialogue	202.5
Human teacher turns per dialogue	101.5
Robot learner turns per dialogue	100.0
Words per human turn	10.0
Words per robot turn	5.0

Table 5: Median number of turns and turn lengths for the corpus of teachable robot (WOZ experiment) dialogues (N=20).

aimed at middle-school students (Lubold et al., 2018a,b).

WOZ experiment overview. Participants were told that their goal was to help Nico solve a set of mathematics problems. Prior to the interaction, they received worked-out problem solutions. During the interaction, a tablet user interface displayed the problem, highlighting one step at time. Nico, controlled by the Wizard, took initiative in leading the dialogue, asking for help about how to approach the problem sub-parts (e.g., *"How do I figure out how much paint to mix?"*). Participants responded by explaining their reasoning (e.g., *"We want to have six cans of green paint so we mix three cans of yellow paint and three cans of blue paint because..."*). Nico's actions included text-to-speech output, gestures such as scratching its head, and updates to values in the tablet interface. Figure 2 shows a student teaching Nico.

Figure 2: Nico, a teachable robot, being taught by a student.

Wizard behavior. A human Wizard operated Nico behind the scenes, selecting dialogue responses and corresponding gesture movements from a pre-defined set. If necessary, they had the ability to input additional phrases. If the participant did not explain their reasoning, the Wizard prompted them to try again (e.g., *"Could you explain that better?"*). The Wizard was not instructed to model specific grounding behaviors.

5.2 Empirical analysis

We analyze the human-robot dialogue transcripts asking the same questions as in Section 4, but from the robot perspective: (1) how does the robot encourage the human to present, and (2) how does the robot signal that it is taking the floor?

5.2.1 Encouraging partner to present

Based on our analysis of the human-human dialogues, we hypothesize that effective strategies for a robot to use when encouraging their partner to present, e.g., to elaborate or to explain, are: *presentation* and *probe* if their partner does not have floor, and *backchannel* if their partner already has the floor.

To evaluate the extent to which the human-robot dialogues reflect these strategies, we identify the following robot dialogue phrases (fixed phrases or templates, available to the Wizard):

- presentation: "Okay, we [perform math operation][4]", "So now we [perform math operation]?"
- probe: "How did we get that number?", "What do we do next?", "Could you give me a hint?"
- backchannel: "Okay"

For each grounding category (presentation, probe, and backchannel) we manually annotate 50 dialogue exchanges surrounding the queried phrases. Each exchange is five turns in length. We label each turn in the exchange with one or more grounding labels, as we did for the human-human corpus. For presentations and probes, the dialogue exchanges are in contexts where the human partner does not have the floor in the preceding turn. Two examples are shown in Appendix A. We test if presentations and probes result in the human partner taking the floor. For backchannels, the dialogue exchanges are in contexts where the human partner has the floor in the preceding turn. We test if backchannels result in the human partner keeping the floor.

Following presentations, 36% of the exchanges had a presentation in the human's first turn after the robot presentation. Following probes, 74% of the exchanges had a presentation in the human's first turn after the robot probe. Following backchannels, 68% of the exchanges had a presentation in the human's first turn after the robot

[4] {add/subtract/multiply/divide} x {and/from/with/by} y.

	Partner turn is presentation	Median turn length (num words)
Following robot presentation		
on the 1st turn	36.0%	13.0
on the 2nd turn	20.0%	19.5
Following robot probe		
on the 1st turn	74.0%	25.0
on the 2nd turn	18.0%	30.0
Following robot backchannel		
on the 1st turn	68.0%	20.0
on the 2nd turn	20.0%	30.0

Table 6: Success in encouraging human to present in the first turn, or second turn following robot presentations, probes, and backchannels; median human turn lengths for presentations.

backchannel. Table 6 summarizes this data, reports turn lengths, and reports on occurence of presentations in the subsequent turn (if the first turn was not a presentation). Not only are probes more effective than presentations at getting the human to present, the subsequent human presentation turns are also longer.

5.2.2 Taking the floor

Based on our analysis of the human-human dialogues, we hypothesize that effective strategies for a robot to use when taking the floor from their partner are: *uptake*, *answer*, and *probe*.

To evaluate the extent to which the human-robot dialogues utilize these grounding acts, we identify four dialogue phrases that the robot uses to take the floor and steer the conversation. The first two selected phrases are navigation instructions, labelled as uptakes. In these, the robot takes the floor to explicitly steer the conversation towards the next problem step. We did not find any suitable robot phrases at floor shifts that we considered to be answers. The second two phrases are questions about the partner's attitudes towards the material. These are labelled as probes, and serve to indirectly steer the conversation away from the previous topic. The dialogue phrases are as follows:

- uptake: "Please tap the 'next' button for me so we can move on to the next step", "Please press the 'back' button"
- probe: "Do you like math?", "Have you done problems like this before?"

We manually annotate 45 dialogue exchanges surrounding each of the queried categories. As above, we label each turn in the exchange with one or more grounding labels. Two examples are shown in Appendix A.

We find that navigation instruction uptakes succeed in taking the floor immediately in 97.8% of the exchanges. For the probes about attitudes towards math, we evaluate their success in shifting the floor by reporting how long the partner continues answering the question that the robot posed, and how verbose those answers are (see Table 7). We find that in 35% of the exchanges, partners continue to answer the question for only one turn; in 60% of the exchanges they stay on-topic for two turns. The average length of these turns is 5.5 and 8.0, respectively.

6 Discussion

In the human-human peer-mentoring dialogue corpus, we find that human speakers encourage partners to take the floor most frequently via presentations or probes. In the human-robot dialogue corpus, we find that probes are more successful than presentations in getting partners to take the floor and also result in longer turn lengths. We note that our analysis is limited by the set of robot phrases queried. To more accurately assess the success of probes versus presentations in human-robot dialogue, we would need to annotate all instances of these two grounding actions in the corpus.

Speakers in the peer-mentoring dialogue corpus encourage partners to keep the floor most frequently by backchanneling. Therefore, it seems that providing a simple acknowledgement of the partner's signal is an effective way to ensure that they continue to present. In the human-robot

	Partner accepts floor shift	Median turn length (num words)
Following robot instruction about UI navigation		
on the 1st turn	97.8%	19.5
on the 2nd turn	0.02%	20.0
Following robot probe about math attitudes		
on the 1st turn	35%	5.5
on the 2nd turn	60%	8.0

Table 7: Success in getting human to accept floor shift following robot instructions about user interface (UI) navigation and probes about math attitudes; median human turn lengths if floor shift is accepted.

dialogue corpus, we find that backchannels are successful in encouraging a partner to hold the floor. Partners present within the next two turns 88% of the time. However, we find that the robot backchannels occur on average in 8.9% of its total turns in a conversation, whereas learners in human-human conversations backchannel for 40.8% of their turns. By incorporating more backchannels in the robot's dialogues (see Kawahara et al. (2016)), we could encourage presentations more often, and also make the robot's dialogue more similar to that of human learners. Backchannels could also take non-verbal form, such as nodding. However, we should be cautious of using backchannels too liberally if they are not a result of true understanding, since they could break down trust between robot and human.

In the human-human corpus, we find that speakers use uptakes, answers, and probes as signals that they are taking the floor. Uptakes are the most frequently used grounding label in this regard. This reinforces the idea that speakers take more initiative when taking the floor because they must produce a relevant turn without being explicitly prompted for it.

In the human-robot dialogue corpus, we find that uptakes in the form of instructions to the human partner are successful in shifting the floor. Due to the nature of the human-robot dialogue, we could not find instances of the robot using answers at floor shifts. Instead, the robot used probes to take the conversation floor. These are less successful than instructions in immediately shifting the floor, but this may be due to the unexpectedness of these questions; participants may have been caught off guard.

To achieve more human-like collaborative dialogue, we suggest that teachable robots consider using the following turn-taking strategies:

- When human partners are not taking initiative, probe partners to encourage them to talk more and take the floor.
- Backchannel more frequently while human partners are presenting to encourage partners to talk more and to better articulate their thoughts and explanations.
- Use uptakes, answers, and probes to take the floor. These can be useful when the conversation has gotten off-course and the robot wants to steer it to a different topic.

7 Conclusion

To inform turn-taking strategies for teachable robots, we annotate and analyze grounding patterns in a corpus of human-human peer-mentoring dialogues and a corpus of human-robot dialogues (Wizard-controlled). In the human-human dialogues, we identify grounding actions that may encourage dialogue partners to take initiative in teaching, while steering the conversation naturally. We find that some of these grounding actions are present in the corpus of human-robot dialogues, but that others are absent, or present to a lesser degree. This suggests future research to investigate whether student outcomes might improve if robot interactions could be designed to encourage more human-like collaborative dialogue.

Acknowledgments

The authors wish to thank Tricia Chaffey, Hyeji Kim, and Emilia Nobrega for their contributions as well as Nichola Lubold and the anonymous SIGDIAL reviewers for their thoughtful feedback. This material is based upon work supported by the National Science Foundation under Grant No. IIS-1637947.

References

Sean Andrist, Xiang Zhi Tan, Michael Gleicher, and Bilge Mutlu. 2014. Conversational gaze aversion for humanlike robots. In *HRI '14 Proceedings of the 2014 ACM/IEEE International Conference on Human-Robot Interaction*, pages 25–32. ACM.

Tricia Chaffey, Hyeji Kim, Emilia Nobrega, Nichola Lubold, and Heather Pon-Barry. 2018. Dyadic stance in natural language communication with a teachable robot. In *HRI '18 Companion: 2018 ACM/IEEE International Conference on Human-Robot Interaction Companion*, pages 85–86. ACM.

Crystal Chao, Jinhan Lee, Momotaz Begum, and Andrea L Thomaz. 2011. Simon plays simon says: The timing of turn-taking in an imitation game. In *Proceedings of RO-MAN*, pages 235–240. IEEE.

Herbert H. Clark. 1996. *Using Language*. Cambridge University Press.

Herbert H Clark and Edward F Schaefer. 1989. Contributing to discourse. *Cognitive Science*, 13(2):259–294.

Arthur C. Graesser, Haiying Li, and Carol Forsyth. 2014. Learning by communicating in natural language with conversational agents. *Current Directions in Psychological Science*, 23(5):374–380.

Agustín Gravano and Julia Hirschberg. 2011. Turn-taking cues in task-oriented dialogue. *Computer Speech & Language*, 25(3):601–634.

Deanna Hood, Séverin Lemaignan, and Pierre Dillenbourg. 2015. When children teach a robot to write: An autonomous teachable humanoid which uses simulated handwriting. In *HRI '15: Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction*, pages 83–90. ACM.

Alexis Jacq, Severin Lemaignan, Fernando Garcia, Pierre Dillenbourg, and Ana Paiva. 2016. Building successful long child-robot interactions in a learning context. In *HRI '16: Proceedings of the Eleventh Annual ACM/IEEE International Conference on Human-Robot Interaction*, pages 239–246. ACM.

Martin Johannson and Gabriel Skantze. 2015. Opportunities and obligations to take turns in collaborative multi-party human-robot interaction. In *Proceedings of the 16th Annual Meeting of the Special Interest Group on Discourse and Dialogue*, pages 305–314. Association for Computational Linguistics.

Takayuki Kanda, Children A. Field Trial, Takayuki K, Hiroshi Ishiguro, Takayuki Hirano, and Daniel Eaton. 2004. Interactive robots as social partners and peer tutors for children: A field trial. *Human-Computer Interaction*, 19(1):61–84.

Tatsuya Kawahara, Takashi Yamaguchi, Koji Inoue, Katsuya Takanashi, and Nigel G. Ward. 2016. Prediction and generation of backchannel forms for attentive listening systems. In *Proceedings of Interspeech*.

Iolanda Leite, Andre Pereira, Carlos Martinho, and Ana Paiva. 2008. Are emotional robots more fun to play with? In *Proceedings of the 17th IEEE International Symposium on Robot and Human Interactive Communication*, pages 77–82.

Daniel Leyzberg, Samuel Spaulding, Mariya Toneva, and Brian Scassellati. 2012. The physical presence of a robot tutor increases cognitive learning gains. *Proceedings of the Annual Meeting of the Cognitive Science Society*, 34(34).

Jamy Li. 2015. The benefit of being physically present: A survey of experimental works comparing copresent robots, telepresent robots and virtual agents. *International Journal of Human-Computer Studies*, 77:23–37.

Changsong Liu, Rui Fang, Lanbo She, and Joyce Chai. 2013. Modeling collaborative referring for situated referential grounding. In *Proceedings of the SIGDIAL 2013 Conference*, pages 78–86. Association for Computational Linguistics.

Nichola Lubold, Erin Walker, Heather Pon-Barry, Yuliana Flores, and Amy Ogan. 2018a. Using iterative design to create efficacy-building social experiences with a teachable robot. In *Proceedings of the International Conference for the Learning Sciences (ICLS 2018)*.

Nichola Lubold, Erin Walker, Heather Pon-Barry, and Amy Ogan. 2018b. Automated pitch convergence improves learning in a social, teachable robot for middle school mathematics. In *Proceedings of the 19th International Conference on Artificial Intelligence in Education (AIED 2018)*.

Raveesh Meena, Gabriel Skantze, and Joakim Gustafson. 2014. Data-driven models for timing feedback responses in a map task dialogue system. *Computer Speech & Language*, 28(4):903–922.

Hae Won Park, Rinat Rosenberg-Kima, Maor Rosenberg, Goren Gordon, and Cynthia Breazeal. 2017. Growing growth mindset with a social robot peer. In *HRI '17: Proceedings of the 2017 ACM/IEEE International Conference on Human-Robot Interaction*, pages 137–145. ACM.

Rolf Ploetzner, Pierre Dillenbourg, Michael Praier, and David Traum. 1999. Learning by explaining to oneself and to others. In Pierre Dillenbourg, editor, *Collaborative-learning: Cognitive and computational approaches*, pages 103–121. Elsevier.

Aaron Powers, Sara Kiesler, Susan Fussell, and Cristen Torrey. 2007. Comparing a computer agent with a humanoid robot. In *HRI '07: Proceedings of the ACM/IEEE International Conference on Human-Robot Interaction*, pages 145–152. ACM.

Aditi Ramachandran, Chien-Ming Huang, Edward Gartland, and Brian Scassellati. 2018. Thinking aloud with a tutoring robot to enhance learning. In *HRI '18: Proceedings of the 2018 ACM/IEEE International Conference on Human-Robot Interaction*, pages 59–68. ACM.

Rod D. Roscoe and Michelene T.H. Chi. 2007. Understanding tutor learning: Knowledge-building and knowledge-telling in peer tutors explanations and questions. *Review of Educational Research*, 77(4):534–574.

David Traum. 1994. *A Computational Theory of Grounding in Natural Language Conversation*. Ph.D. thesis, Computer Science Dept., University of Rochester.

David R Traum. 1999. Computational models of grounding in collaborative systems. In *AAAI Fall Symposium on Psychological Models of Communication in Collaborative Systems*, pages 124–131.

A Supplemental human-robot dialogue examples (annotated)

Example 1: Robot presentation (see 2nd turn)		
HUMAN:	Correct so what would you be multiplying by?	*probe*
ROBOT:	We multiply two-point-five and fifty	*presentation*
HUMAN:	Very good so what does that produce?	*probe*
ROBOT:	The answer is one-fifty	*presentation*
HUMAN:	Very good good job	*uptake*
Example 2: Robot probe (see 2nd turn)		
HUMAN:	Yes	*answer*
ROBOT:	How did we get that number?	*probe*
HUMAN:	Sixty times two-point-five is one-fifty	*answer/presentation*
ROBOT:	Hmm why did we do that?	*probe*
HUMAN:	Because we knew how many miles he goes in one one hour and we wanted to know how many miles he goes in two hours	*presentation*
Example 3: Robot uptake - navigation instruction (see 2nd turn)		
HUMAN:	Mh-mmm	*backchannel*
ROBOT:	Please tap the next button for me so we can move on to the next step	*uptake/presentation*
HUMAN:	Okay now we're gonna work on your torso	*uptake/presentation*
ROBOT:	Hmmm that part seems complicated to me can you explain that again?	*probe*
HUMAN:	So we know that with a surface area of six the volume is two	*presentation*
Example 4: Robot probe - attitude question (see 2nd turn)		
HUMAN:	That is correct	*presentation*
ROBOT:	Do you like math?	*probe*
HUMAN:	Eh sometimes	*answer*
ROBOT:	Was it hard when you first started?	*probe*
HUMAN:	Oh yeah it took me a long time to get multiplications right I was a late bloomer	*answer*