JOURNAL	Computer Assisted Language Learning
CITATION	Veivo, O., & Mutta, M. (2025). Dialogue breakdowns in robot-assisted L2 learning. *Computer Assisted Language Learning*, *38*(1–2), 30–51
DOI	https://doi.org/10.1080/09588221.2022.2158203

Dialogue breakdowns in robot-assisted L2 learning

Outi Veivo
University of Turku, Finland (outi.veivo@utu.fi)

Maarit Mutta
University of Turku, Finland (maarit.mutta@utu.fi)

Abstract

This study focuses on dialogue breakdowns which can occur in robot-assisted language learning (RALL). Our aim is to analyse how children use gaze to resolve these breakdowns, ie. interruptions in the interaction caused by the robot's inability to understand the children and react appropriately. Our corpus consists of 18 video filmed L2 learning situations where 36 primary school children talk pairwise with an educational robot for the first time. Our participants are 10–13 years old mono- and bilingual children from Swedish speaking schools in Finland learning L2 English. After detecting the breakdowns from the data, we use a multimodal analysis to identify most typical gaze patterns during these sequences. Our results show that when breakdowns occur, children's first gaze is directed most frequently towards the robot, but after that, they shift their gaze most often towards the teacher. This result suggests that the children try first to resolve the breakdowns with the robot, but when the problem persists, they also use gaze to seek assistance from other participants present in the learning situation. We interpret this finding to show that children attempt to treat social robots as human-like conversational partners in RALL, but that they turn to other human participants if the robot does not follow the expected interactional behavior.

Keywords: robot-assisted language learning; L2 learning; dialogue breakdown; child-robot interaction; human-robot interaction

1. Introduction

Social robots have recently been adopted in educational contexts (for review, see Belpaeme, Kennedy, Ramachandran, Scassellati, & Tanaka, 2018), including foreign language (L2) learning. What makes a social robot "social" is its capacity to interact autonomously or semi-autonomously with humans respecting norms of human behaviour (Bartneck & Forlizzi, 2004, p. 592). Unlike other types of industrial or service robots, social robots are specifically designed for interaction with humans, and unlike virtual agents or chatbots, they have a physical embodiment (e.g. Randall, 2019). These aspects allow for a multimodal human-robot interaction (HRI) that includes both verbal and nonverbal elements, such as changes of posture, gestures and gaze. This multimodal interaction opens interesting new avenues for L2 learning but it can also present different challenges for human participants.

Robot-assisted language learning (RALL) has been studied from different viewpoints (for reviews, see van den Berghe, Verhagen, Oudgenoeg-Paz, van der Ven, & Leseman, 2019; Randall, 2019). For instance, there is evidence that social robots can reduce language anxiety in speaking (Alemi, Meghdari & Ghazisaedy, 2015) and improve language learners' motivation (Alemi, Meghdari & Haeri, 2017). Learners often enjoy interaction with robots (e.g. Lee, Noh, Lee, Lee, Lee, Sagong & Kim, 2011), because speaking in L2 with the robot can be less stressful than speaking with the teacher or with another learner (Alemi, Meghdari, & Ghazisaedy, 2014). Robots can be useful even with beginning learners, because unlike humans, robots do not get tired of repeating the same sentences or dialogues. This is why they can be used very efficiently also for pronunciation training (Peura, Mutta & Johansson, 2021). Although there is evidence that interest in robots can be due to a novelty effect which can decrease fairly quickly – even in only two weeks (see e.g. Kanda, Hirano, Eaton & Ishiguro, 2004), there is also evidence that the aforementioned affective factors can be important for improving different L2 competences even during longer RALL periods (e.g. Lee et al., 2011).

Although many aspects of RALL have been widely studied, relatively little is known about the nature of interaction in RALL situations. The interactional space in RALL includes the participants (i.e. teachers and learners) as well as the material and technological artefacts (e.g. texts, computers, blackboards). These resources are involved in creating a situated organisation of actions (Jakonen & Jauni, 2021) in a multimodal communication frame including language, gestures, gaze and body postures (cf. Mondada, 2019; Jakonen, Veivo, Mutta, Maijala. Honkalammi & Johansson, 2022). Therefore, it is also important to analyse HRI in a multimodal frame. For instance, combining the analysis of verbal interaction to the analysis of non-verbal components such as head movements or gaze can give tools to interpret participants' actions in a specific RALL situation.

There are multiple reasons why the interaction with a robot does not always proceed according to an expected model, for instance due to problems in automatic speech recognition (ASR) systems (Honig & Oron-Gilad, 2018). In the present article, we focus on dialogue breakdowns, moments when the social robot does not react appropriately for the interaction to continue (Uchida, Minato, Koyama & Ishiguro, 2019). More precisely, we are interested in moments when the progressivity of the dialogue in RALL is interrupted because the robot does not react properly to the utterance of the human participant. We focus on the role of gaze in child-robot-interaction (CRI) in RALL with primary school children learning English as an L2. We ask the following research question:

How do children use gaze as a non-verbal action to resolve dialogue breakdowns in RALL?

To answer this question, we analyse the dialogue breakdowns in the multimodal child-robot-interaction qualitatively focusing on the role of gaze during the dialogue breakdowns. We will start by reviewing former studies on the role of gaze in HRI as well as in language learning contexts. We then move on to characterising different types of breakdowns in HRI before describing the method and the results of the present study more in detail.

2. Interaction in RALL

The embodiment of the social robots allows for a multimodal interaction which has been identified as an important factor for motivation in learning contexts (Chang, Lee, Wang & Chen, 2010), especially if these robots are capable of acting supportively and of giving non-verbal feedback (Saerbeck, Schut, Bartneck & Janse, 2010). In this section, we will first review earlier studies on the role of gaze in the multimodal HRI before describing different reasons for breakdowns in HRI. Finally, we reflect on the role of feedback in RALL in parallel with the classic model of classroom interaction (IRE model).

2.1 The role of gaze in HRI

Interaction with social robots is different from interaction with computers, virtual agents or other devices because of the embodiment of the robot. It means that the robot can, for instance, move its head to keep an eye-contact with its interlocutor unlike virtual agents or chatbots. This embodiment allows robots also to use resources such as gestures, body movements and the direction of eye gaze to communicate multimodally (Bonarini, 2020; Bartneck, Belpaeme, Eyssel, Kanda, Keijsers & Šabanović, 2020, pp. 84-85). In the RALL context, a social robot can use these resources for example to indicate acceptance of correct answers by nodding or by holding its thumb up, or alternatively, react to incorrect answers by shaking its head or by holding its thumb down (Saerbeck et al., 2010). In addition to the human-like features of embodiment, social robots can also bring an added dimension to the interaction with robot-specific features, such as lights, sounds, colours or music (Bonarini, 2020).

One of the most interesting features of multimodality in HRI is gaze. Many social robots are programmed to recognize their interlocutor by eye-contact which means that a mutual gaze between the human speaker and the robot is necessary for the interaction to continue. Robot's gaze can also be used to guide attention, to indicate turn-taking or to show different emotions (Saerbeck et al., 2010; Randall, 2019). In human-human interaction (HHI), eye gaze is used to indicate shared attention (Bartneck et al., 2020, p. 83). A speaker usually looks at the listener

(77% of the time) and a listener usually looks at the person who speaks (88% of the time) (Vertegaal, Slagter, van der Veer & Nijholt, 2000). However, mutual eye contact occurs only less than 30% of the time in HHI (Webbink, 1986). When there are multiple participants in a conversation, gaze shifts are often accompanied by head movements, which are more salient than simple eye gazes and therefore important for signalling speech intentions. Those who want to take the turn usually move their head to look at the participant who is speaking and those who prefer to remain silent turn their head to the participant they expect to take the next turn (Jokinen, Nishida & Yamamoto, 2010). Eye contact can also convey information about interlocutors' attitudes and emotions during the conversation (idem). Because of its importance for turn-taking, gaze has also an important function for including other participants in the interactional space (Mondada, 2019). Therefore, absence of eye-contact during a conversation can be interpreted as a sign of disinterest or inattention.

In educational context, teachers are known to rely on gaze in allocating turns in the classroom (e.g. Kääntä, 2012). Similarly, pupils are known to use gaze in the classroom to show interest or emotions, to ask questions and to obtain approval (Okita, Ng-Thow-Hing & Sarvadevabhatla, 2011). In HRI, eye contact has an important role, because it is often used as a means for the robot to identify its interlocutors. Interestingly, eye-contact with the robot can also shape human interlocutor's experience, e.g. help to remember what the robot is saying (Mutlu, Forlizzi & Hodgins, 2006). In addition to being a sign of involvement in the interaction, eye-contact can also indicate a need for verbal or non-verbal feedback.

To sum up, multimodality and eye-contact are crucial elements of interaction with social robots. Although there is evidence that these features can be useful in RALL contexts, relatively little is known about how children resolve problematic situations in RALL. To be able to analyse these situations more in detail, we now move on to describe different types of breakdowns that can occur in HRI.

2.2 Breakdowns in HRI

Like in any interaction, there can be different breakdowns in the interaction between humans and social robots. In human computer interaction (HCI), breakdowns have been defined as interruptions in using computer applications caused by some unexpected dysfunction within the application (Bødker, 1995). In HRI, breakdowns can be viewed in a broader context and divided for example into technical failures and interaction failures (Honig & Oron-Gilad, 2018).

Technical failures have gained more attention than interaction failures in the HRI literature. They can be further divided into problems in the design or implementation of the software and to problems related to the hardware, such as dysfunctions in the sensors or problems with its internet connection (for a review see Honig & Oron-Gilad, 2018). In the RALL context, especially with children, technical failures in the robots' automatic speech recognition (ASR) can be of special interest. ASR systems work ideally recognizing the speech of fluent native speakers in noiseless environments. Also, although ASR systems are constantly developing, they are usually trained with adult speaker corpora and therefore less accurate with children (Potamianos & Narayanan, 2003). Compared to adults, children are more heterogeneous as a group, because they can be at different developmental stages in their language learning and because they can have a lot of variation in their speech patterns (Gurunath Shivakumar & Georgiou, 2018). Children learning an L2 may have many different shortcomings in their language competences which can be a supplementary challenge to ASR (see also Randall, 2019). For instance, even proficient L2 speakers can have a foreign accent which can lead to failures in HRI. Accentedness can influence both intelligibility - how much of the utterance is actually understood - and comprehensibility - how easy or difficult it is for the listener to understand the speaker (Derwing & Munro, 2009, pp. 478-480). In HHI, the degree of intelligibility may vary: some parts of the utterance can be understood whereas other parts are not understood. In HRI, intelligibility is a binary concept: either the robot understands the speaker or it does not. This means that even a minor problem in L2 speech production - which might be easily overcome in HHI - may result in total unintelligibility in HRI, especially in

CRI. Indeed, previous studies have confirmed that non-native accented speech is a challenge for ASR systems (e.g. Derwing, Munro & Carbonaro, 2000; Moussali & Cardoso, 2016).

Second, breakdowns in RALL can be caused by failures in the robot's interaction with its environment. These failures include for instance problems in the robot's sensors to detect obstacles in its environment (Honig & Oron-Gilad, 2018). In the present study, however, we are more interested in *dialogue breakdowns,* interruptions in the flow of conversation caused by the robot's failure to react appropriately to human participants. We rely on Uchida et al. (2019) and define dialogue breakdowns as moments where the robot gives an inappropriate response which interrupts the expected dialogue. Dialogue breakdowns are typical in conversations with dialogue systems such as the ones required by social robots, because their capacities for automatic speech recognition (ASR) and natural language processing are limited (Uchida et al., 2019, p. 1). Recently, there has been a growing interest also in how dialogue breakdowns could be avoided not only by improving robots' dialogue capacities but also by programming the robots to encourage users to avoid and recover from dialogue breakdowns (*idem.*).

In RALL, there may be several reasons for dialogue breakdowns. Most importantly, the language system of the young L2 learners is still developing, and therefore there may be different shortcomings in their L2 competences. These include shortcomings at all levels of speech production, such as pronunciation problems, gaps in vocabulary knowledge or insufficient knowledge about syntactic rules. There is ample evidence that L2 speakers rely on a variety of different verbal strategies (from repeating their utterances to using their L1) to resolve situations caused by these shortcomings (e.g. Poulisse, 2019). However, in CRI, these shortcomings may lead to dialogue breakdowns if the conversation is based on pre-programmed patterns which give little or no space for using communication strategies or for adjusting the utterances. There is evidence that when children communicate with the robot in pairs, they are skillful in handling problematic RALL situations by seeking and giving verbal advice to their

peers (Honkalammi, Veivo & Johansson, 2022). They may help their peers spontaneously or when asked to do so, for instance by translating the robot's utterances, by hinting the correct answers or by telling how to continue the interaction with the robot.

2.3. The role of feedback in RALL

When interacting in L2, learners need feedback for their utterances to be able to judge whether their hypotheses about the use of L2 are correct, and to evaluate what they already know and what they still need to acquire. In classroom discourse, it is usually the teacher who gives this feedback to the learners. Question-and-answer sequences in classroom discourse are known to follow the so called IRE model (for initiation-response-evaluation; or IRF: initiation-response-feedback), where the teacher initiates the activity, the learners respond to the teacher and finally the teacher evaluates the answer by accepting or rejecting it (e.g. Sinclair & Coulthard, 1975; Lyle, 2008; Walsh, 2015, pp.17-19). This discourse structure is quite common as it concerns 60% of the teaching related talk in classrooms (Lyle, 2008; p. 225).

In the RALL context, feedback is equally important for proceeding in different pedagogical activities (cf. Maijala & Mutta, 2021), but it can be the robot who is in charge of evaluating the answers and giving the appropriate feedback to the learners. As stated above, robots have resources which allow for many forms of feedback in the interaction. For instance, showing multicoloured blinking eyes in response to a correct answer to reward children's correct answers has been shown to increase motivation in RALL (Ahtinen & Kaipainen, 2020). If no feedback is available or if the feedback is insufficient, dialogue breakdown may occur.

Although dialogue breakdowns have been studied in other robot interaction contexts, such as in museums with adults (Arend, Sunnen & Claire, 2017) or in map reading or game scenarios with children (Serholt, 2018), relatively little is known about breakdowns in the RALL context. Therefore, the main objective of the present study is to analyse dialogue breakdowns in RALL from the language learner's perspective. Our rationale for examining dialogue breakdowns is

therefore not to find ways to remedy these issues by programming the robot differently (cf. Uchida et al., 2019). Instead, we are interested in the human participant's actions during dialogue breakdowns caused by the robot's incapacity to understand the human participant's utterance. By "understand" we mean the robot's ability to recognize spoken utterances and to react appropriately for the interaction to progress. With this delimitation, we exclude from our analysis dialogue breakdowns resulting from the lack of human's motivation to continue to communicate with the robot (cf. Matsui, Tani, Sasai & Gunji, 2021). Also, we are not interested in dialogue breakdowns caused by human participants' comprehension problems, but focus on the breakdowns which occur when the robot does not give the expected feedback by accepting the response given by the children communicating with the robot in their L2. Therefore, our focus is on children's behaviour in RALL situations to find out how the children use gaze to resolve these problematic moments which interrupt the progressivity of the interaction. We hypothesise that if the interaction does not follow the typical IRE model, with the robot showing acceptance of the learner's answers, the learners will try to invite other participants in the interactional space to find ways to progress in the activity.

We now move on to describe how the data was collected and analysed, before presenting the results of our study.

3. Methodology

3.1 Corpus and participants

The data analysed in this study is a sub-corpus of a larger RALL-corpus with 111 participants collected in spring 2019. This larger corpus (13 hrs 42 min) includes recordings from 42 RALL-situations with two to four child participants. The sub-corpus analysed in the present study includes only situations with two child participants (n = 18, duration = 5 hrs 33 min). The participants of the present study (n = 36) were 10 to 13 years old children (average: 11.5 years) - 19 boys, 17 girls - from two Swedish speaking primary schools in southern Finland. At the time of the study, they had been learning English as a foreign language at school for 1,5 to 3,5

years. All participants were Swedish - Finnish bilinguals, but they used languages differently at home: 19 participants used only Swedish, 2 participants used only Finnish, 8 used dominantly Swedish but also Finnish, 5 used dominantly Finnish but also Swedish, 2 used dominantly Swedish but also another language. All participants were granted informed consent from their parents and from local school authorities to take part in the study before data collection.

3.2. Procedure and data collection

The robot used in this study was NAO, a small humanoid robot (by Softbank Robotics). This robot is not designed only for educational purposes, but it is widely used in different RALL contexts (see e.g. Randall, 2019). NAO is programmable with the Choregraphe software, but the RALL situations of the present study were based on the Elias robot application (by Curious technologies) which contains a set of pre-programmed lessons on different themes at different proficiency levels. The subparts of these lessons follow the common action chain of classroom discourse based on the repetitive actions of initiation, response and evaluation (IRE model) (e.g. Lyle, 2008). Whereas in the classroom it is the teacher who initiates a communicative sequence and waits for the learners to respond before evaluating the response by accepting or refusing it, in this application, it is most often the robot who initiates the sequence. The pre-programmed lessons follow a fairly rigid structure and give little space for variation, although sometimes alternatives for correct answers are programmed within the activities (e.g. *dad* or *father*). In this application, the robot is therefore not an independent actor who would be able to participate autonomously to verbal interaction nor to adapt to non pre-programmed interaction situations.

Elias application lessons focus on vocabulary learning. At first, the words are presented with visual support on the support device. Then the learners have to repeat the words, reply to the robot's questions relying on the visual cues, use the words in sentences and to discuss freely with the robot. The exercises are self-paced. The course of these RALL-lessons is depicted in Figure 1.

<Please add Figure 1 here>

As Figure 1 shows, the subparts of the lessons are mostly initiated by the robot. At the final stage, it is the learner who initiates the interaction, but the robot is only capable of reacting to the utterances within the limitations of rehearsed vocabulary and sentence structures.

The learner has to have eye-contact with the robot in order for the robot to listen to the speaker. When the learner gives a correct answer, the robot approves it by nodding and by repeating the correct answer. If the learner is especially good at pronouncing the correct answer, the robot cheers this by showing the so-called candy eyes – by blinking its eyes with lights of different colours. If the answer is incorrect, the robot reacts with a simple nod.

In the RALL situations of the present study, the participants met the robot in a quiet room pairwise. The children had seen the robot in the presence of the whole class, but this was their first encounter with the robot in a learning situation where they had to speak English with it. The learning situation consisted of one thematic Elias application lesson on family words (e.g. *grandmother, sister, brother*). The length of the learning situations varied between 14 and 29 minutes because they were self-spaced and because the length depended on the amount of repetitions, dialogue breakdowns and false answers. During the lesson, the participants alternated talking with the robot. A visiting teacher (unknown to the participants before this lesson) assisted in the RALL-lesson as a tutor. The teacher helped the learners to interact with the robot by giving instructions or advice when need be. At the end of the RALL situation, the teacher asked the children how they felt about talking with the robot. The responses to this question were also video filmed. The organisation of the interactional space in the analysed RALL situations is depicted in Figure 2.

<Please add Figure 2 here>

As Figure 2 shows, the interactional space in the RALL situations was organised so that human participants were sitting on the floor and the robot was standing between them. The two learners were facing the robot whereas the teacher sat behind the robot. The laptop used as a support device was on the floor next to the robot. The participant next to the laptop operated it. This setup meant that the learners had to lean towards the robot when they were talking with it to keep eye contact. In addition, one or two research assistants were present for managing the data collection with two video cameras and an audio recorder.

3.3 Data analysis

The video filmed RALL situations were first transcribed using a simplified version of the multimodal transcription conventions based on Jefferson (2004) and on Mondada (2019) focusing on the content of the interaction. Second, we moved on to detect all dialogue breakdowns moments where the robot did not respond to the learner's utterance as expected. They were operationalised as moments in the interaction where the learner had to repeat the response to the robot's utterance more than once. For this purpose, the two video recordings of the learning situations with their transcriptions were manually analysed by two research assistants and verified by the two authors. When the dialogue breakdowns had been detected, we examined the breakdown moments more in detail to calculate the frequencies of their possible causes. Third, we moved on to analyse children's gaze behaviour after the dialogue breakdowns to shed light on the role of gaze in resolving them. We analysed the video recordings from the camera headed towards the learners in detail to reveal where the learners directed their gaze immediately after the breakdown (first gaze) and where they looked at after that (second gaze). After identifying five main gaze directions, the frequencies of the first and second gaze towards them were calculated. Fourth, on the basis of these frequencies, we established a categorization of three most typical gaze patterns during dialogue breakdowns, selected representative examples of them from the data, and examined them qualitatively in more detail. Finally, we compared children's comments on RALL at the end of the lesson to the number of dialogue breakdowns.

4. Results

In the analysed 18 RALL situations with 36 participants, we identified altogether 132 dialogue breakdowns. The number of dialogue breakdowns in one RALL situation varied between 2 and 22 and the number of breakdowns per speaker between 1 and 16. This means that the nature of the RALL situations varied greatly between participants: some had a fairly smooth conversation experience with the robot whereas for others the interaction was constantly interrupted.

We then moved on to analyse the dialogue breakdowns more in detail to identify why the robot did not react to the speaker as expected. In the great majority of cases (63%), the breakdown occurred because the robot faced a technical failure. The most important reason for a breakdown to occur was a technical problem in the robot's automatic speech recognition system (37%). In other words, even if the L2 learner produced an answer which the robot was supposed to accept as a correct answer, the robot did not recognize it. There were also other types of technical breakdowns: breakdowns due to dysfunctions of the robot's software, such as asking questions from a wrong thematic lesson or being unresponsive during the free conversation or breakdowns due to problems in the robot's internet connection (26%). A smaller proportion of the breakdowns (37%) was related to the participants. Most often, the L2 learners did not pronounce the words correctly (27%), and this is why the robot was unable to understand them. Problems in forming correct phrases (6%) or problems in using the correct words (2%) were much less frequent. In only 2% of the cases there was a dialogue breakdown because the child did not speak loudly enough. Taken together, a big majority (66%) of dialogue breakdowns were related to sound and to speech recognition. This shows that the robot's ASR system is not always functioning as it should, even if the L2 learners would be pronouncing the words correctly and using a sufficient speech volume. This is in line with the results of Honig and Oron-Gilad (2018) and Potamianos and Narayanan (2003), who showed that problems in ASR constituted a major challenge in HRI and especially in CRI.

The analysis of children's gaze behaviour during dialogue breakdown moments revealed five main gaze directions for the first and second gazes: robot, teacher, peer, computer and other artefact in the interactional space. We started by examining the proportions of first and second gazes to different participants and artefacts in the interactional space. These proportions are depicted in Figure 3.

<Please add Figure 3 here>

Our results show that at the time of the dialogue breakdowns (1st gaze), children most often keep their eyes directed to the robot (86%), but if they deviate their gaze from the robot, they most often look at the teacher present in the situation (7%). They may also look at the computer screen (4%) to check the instructions or elsewhere in the interactional space (e.g. wall, up in the air) (3%) and very rarely to their pair (1%). Some children shifted their gaze very rapidly to several participants or objects in the interactional space during the breakdowns, but only the target of their first gaze is included in these figures.

When we examined the direction of the participants' gaze immediately after the dialogue breakdown (2nd gaze), we could notice that most often the children turned their head to look at the teacher (41%) or kept their gaze headed towards the robot (33%). If they looked at their teacher, they were looking for help or for support from the teacher. If they continued to look at the robot, they were repeating their utterance to the robot. Some of the children turned their head towards their pair (12%). This was because they were seeking support from the other participant, or the other participant had given them some advice on how to overcome the communication problem. If they were trying to figure out the right answer by themselves, they deviated their gaze from the robot towards the computer screen (8%) or towards other artefacts in the interactional space (6%).

To illustrate the role of gaze during the dialogue breakdowns, we present three extracts corresponding to the most typical gaze behaviour patterns in the data, namely gaze to robot (Extract 1), gaze to robot and teacher (Extract 2), and multiple gaze directions (Extract 3). All the extracts concern the first part of the Elias lesson on family words, repeating individual words after the robot. The following abbreviations are used in the transcriptions: T = teacher, R = robot, C1 = child 1, C2 = child 2, gz = gaze.

In Extract 1, the learner C1 says the word *brother* after seeing the corresponding picture on the computer screen and engages in a firm eye-contact with the robot.

Extract 1. Gaze towards the robot.

```
O1 R brother
02 C1 brother
                + gz to robot
03 R            *nods*
04 C1 *coughs in her arm*
05 C1 brother
                + gz to robot
06 R            *nods*
07 C1 brother
                + gz down towards robot
08 R            *nods*
09 C1 brother
                + gz to robot
10  R           brother
+ sound effect and blinking eyes as a reward for a correct answer
```

C1 repeats four times the word *brother* (lines 2, 5, 7 and 9) before the robot recognizes and approves it with a sound effect. The robot also rewards the learner non-verbally by blinking its multicoloured eyes at the same time. This rewarding feedback closes the IRE sequence initiated by the robot. In this extract, the learner interacts solely with the robot, as he keeps his

gaze headed towards the robot the whole time. This shows that despite the fact that the robot does not understand what the child is saying, the child does not try to resolve the problem by looking at other participants. Instead, by continuing to look at the robot, the child shows that he includes only the robot in the interactional space.

Extract 2 is an example of a dialogue breakdown situation where the learner (C1) looks both at the robot and at the teacher.

Extract 2. Gaze towards the robot and the teacher.

01 R **cousin**
02 C1 cousin
+ gz to robot
03 R ***nods***
04 C1 cousin
+ gz to robot
05 R ***nods***
C1 gz to teacher
06 T **[prova bara]**
'just try'
07 C1 [cousin]
+ gz to robot
08 R ***nods***
09 T **ännu**
'again'
C1 gz to teacher
10 C1 cousin
+ gz to robot
11 R ***nods***
C1 gz to teacher
12 C1 cousin
+ gz to robot
13 R ***nods***
+ putting hands on hips

C1 gz to teacher
14 T han tittar på dig då
'he looks at your then'
+ moving the robot slightly towards the learner
15 C1 mm-hm
+ gz to robot
16 C1 cousin
+ gz to robot
17 R cousin
+ sound effect as reward of correct answer
C1 gz to teacher
18 T okej
'okay'

In this Extract, the learner C1 repeats the word *cousin* after the robot and looks at the robot in the eyes. As the robot does approve the answer, C1 shifts her gaze towards the teacher (line 5), and the teacher encourages her verbally by saying *just try* (line 6). C1 repeats the word again, gazing at the robot (line 7) and the teacher continues to encourage her verbally (line 9). This pattern repeats itself several times until the teacher asks whether the robot is in the right posture and turns it slightly towards C1 (line 14). This seems to help and finally the robot approves the word with a sound effect (line 17). After this, C1 shifts her gaze towards the teacher who also approves the right answer by saying *okey* and the turn ends. This example shows that when the dialogue breakdown occurs, the child first tries to resolve the situation by herself, but when she is not successful, she shifts her gaze towards the teacher. This gaze shift can be interpreted as a means to seek help from the teacher. By doing this, she also includes the teacher in the interactional space. The teacher accepts this invitation and starts to help the child in this problematic situation both verbally and with her actions by moving the robot towards the child. This example shows that in a more problematic dialogue breakdown requiring multiple repetitions, the child seeks for feedback not only from the robot but also from the teacher, even at the end of the IRE sequence.

In Extracts 1 and 2, the child is finally understood by the robot, and the turns end either in the robot's or the teacher's approval. In Extract 3, however, the child's turn does not end as expected, and the other learner (C2) offers his help.

Extract 3. Gaze towards multiple directions.

01 R **aunt**
02 C1 aunt *sniffing and laughing a little bit*
+ gz to wall
+ gz to robot
+ gz to floor
03 R ***nods***
04 T **en gång till**
'once more'
05 C1 aunt
+ gz to robot
06 R ***nods***
C1 gz to teacher
07 C1 jag vet inte hur (uttala)
'i don't know how (to say)'
+ gz to computer
08 T no han säger aunt [[us pronunciation]]
'no he says aunt'
09 C2 *presses the button 'repeat'*
+ gz to *computer*
10 R **aunt**
11 C1 *shows C2 with the hand*
+ gz to C2
12 C1 säger du(.)det går inte bra
'you say it(.)it does not go well'
+ gz to C2
13 C2 ska jag säga
'should i say'
+ gz to teacher

14 T	**joo**	
	'yes'	
	+ turn the robot towards C2	

This sequence starts with the robot saying the word *aunt* with the US pronunciation. The child (C1) repeats this word with the UK pronunciation (line 2), and after this she looks in multiple directions, first at the wall, then at the robot and finally shifts her gaze towards the floor. At the same time, she shows her uneasiness with gestures and laughter. She may feel uneasy because already before this turn, she has had problems in getting the robot to understand the words that she repeats (e.g. *sister*), because she talks with a silent and shy voice. She can also be uncomfortable if she realises that her own pronunciation is not similar to the model given by the robot.

When the robot continues to nod as a sign of disapproval (line 6), C1 looks at the teacher and affirms in her first language that she does not know how to pronounce the word (line 7). The teacher helps her and pronounces the word with the US pronunciation (line 8) and after that C1 uses the computer to listen to the robot saying the word again (line 10). After this, she abandons and gives the turn to C2 first non-verbally and then verbally (lines 11-12). C2 asks if he should repeat the word instead of C1 and looks at the teacher. She confirms verbally and turns the robot towards C2 (line 14) and C2 takes the turn. As in Extract 2, the child who faces the dialogue breakdown uses gaze to seek help from other participants, now both from the teacher and from the other child. This example shows that the dialogue breakdowns are not always resolved by closing the IRE sequence by acceptive feedback from the robot or from the teacher. Instead, the breakdown leads to handing the turn over to the other learner.

At the end of the RALL situation, the teacher asked the participants how they felt about learning English with the robot. Their comments on this first experience of RALL were mainly positive (32 out of 36 learners). They described their experience as *fun*, *good*, *nice*, *cool*, *interesting* or *easy*. Only two children commented on the RALL situation negatively by saying that it was

weird because the robot did not understand what the children were saying. One child commented on the RALL situation as *unusual* and another pointed out that *it was fun but they had to repeat the same things many times*. The two negative comments came from participants who did not experience an extensive amount of communication breakdowns with the robot (5 and 1 respectively). Nor did their pair face exceptionally many breakdowns in the RALL situation (5 and 5 respectively). Also, both children who were present in the situation where there were altogether 24 dialogue breakdowns were positive about RALL, and thought that it was fun to talk with the robot. This shows that the number of dialogue breakdowns did not seem to affect participants' first impressions on RALL.

5. Discussion and conclusion

The results of the present study show that RALL situations may be interrupted by a relatively high number of dialogue breakdowns. In our data, these breakdowns were due mainly to the robot's technical dysfunctions and especially to problems in automatic speech recognition. This finding is not surprising, because ASR systems are known to be less accurate with children (Potamianos & Narayanan, 2003) and because the L2 learners have more variation in their speech patterns than L1 speakers (cf. Gurunath Shivakumar & Georgiou, 2018). Learners' limited linguistic resources were another reason for dialogue breakdowns. Most importantly, the learners did not know how to pronounce the L2 words (e.g. differences between UK and US pronunciations) or their pronunciation was influenced by their L1. This finding is in line with studies showing that ASR systems are not yet very flexible in understanding accented speech (cf. Moussalli & Cardoso, 2016). Our results confirm that even minor problems in spoken utterances can cause total unintelligibility in CRI (cf. Derwing & Munro, 2009). Interestingly, one of the participants of our study commented on this same dilemma after the RALL situation, saying that a human participant would have been able to continue the conversation whereas the robot was not. This comment shows that dialogue breakdowns can shift learners' perspective to regard robots more as artefacts than as conversational partners. This shows that if the robot's pre-programmed interactional model is rigid and does not allow

for adaptations, dialogue breakdowns are likely to occur. Therefore, although social robots have been developed from being assistants to being partners, if they are not based on highly adaptive algorithms, social robots in RALL cannot yet be considered as fully autonomous agents (cf. Bartneck & Forlizzi, 2004).

Our main objective was to find out how the children use gaze to resolve dialogue breakdowns which interrupt the progressivity of the interaction. We hypothesised that in cases not following the typical IRE model (cf. Sinclair & Coulthard, 1975; Lyle, 2008; Walsh, 2015), the learners would use gaze to invite other participants of the RALL space to complete the activity. Our results showed that children typically kept eye-contact with the robot even during dialogue breakdowns until the robot accepted their response and closed the IRE sequence. This was a common behaviour probably because they were instructed to look at the robot so that the robot could understand their answers and so that they could continue the activity. If they did not look at the robot, they looked most often at the teacher inviting her to confirm the positive feedback and to close the IRE sequence. The participants also used gaze to seek for help from their peers (cf. Honkalammi et al., 2022). These findings confirm our hypothesis: when problems occurred and when the robot did not close the IRE sequence by approval, children used their gaze to seek support from other human participants in the interactional space. These shifts of attention from the robot to the human participants can also be interpreted to reveal the ambiguity of the robot's role in the interaction: during moments of mutual gaze it is seen as an interactional partner but when dialogue breakdowns occur, gaze shifts to other participants may indicate that it is seen more as an artefact (cf. Maijala & Mutta, 2022). This interpretation, however, needs to be studied more in detail in the future.

Despite the relatively high amount of dialogue breakdowns during the RALL situations, most of our participants commented on their first encounter with a social robot positively at the end of the robot lesson. This may, however, reflect a novelty effect which is often observed for social robots (e.g. Kanda et al., 2004). To overrule this possibility, there is a need for

longitudinal studies on the influences of dialogue breakdowns in RALL. Also, we limited our analysis to the participants' gaze behaviour during dialogue breakdowns, but participants' non-verbal interaction at these moments could be examined more holistically together with verbal interaction. This type of multimodal micro-analysis would be useful to study more subtle ways of maintaining progressivity during the RALL interactions. Finally, the influence of dialogue breakdowns on short-term and long-term learning outcomes of RALL was outside the scope of the present study, but it is also an important matter of future research.

Our study shows that dialogue breakdowns occur frequently in RALL and that children are very patient in trying to resolve these situations in their L2 with the robot. They use their gaze to include other participants in the interaction and thereby to seek assistance and positive feedback. This study contributes to the RALL literature in showing the importance of gaze and non-verbal communication in CRI and by revealing how learners address the robot, the teacher or peers. It underlines the importance of multimodal approaches in studying and in assessing language learning when it is mediated by non-human interactional partners.

The authors report there are no competing interests to declare.

References

Ahtinen A., & Kaipainen K. (2020). Learning and Teaching Experiences with a Persuasive Social Robot in Primary School – Findings and Implications from a 4-Month Field Study. In S. Gram-Hansen, T. Jonasen, & C. Midden (eds.). *Persuasive Technology. Designing for Future Change*. 15th International Conference on Persuasive Technology, PERSUASIVE 2020, Proceeding (12064). Cham: Springer, 73–84. https://doi.org/10.1007/978-3-030-45712-9_6

Alemi, M., Meghdari, A., & Ghazisaedy, M. (2014). Employing humanoid robots for teaching English language in Iranian junior high-schools. *International Journal of Humanoid Robotics, 11*(03), 1450022. https://doi.org/10.1142/S0219843614500224

Alemi, M., Meghdari, A., & Ghazisaedy, M. (2015). The impact of social robotics on L2 learners' anxiety and attitude in English vocabulary acquisition. *International Journal of Social Robotics,* 7(4): 523–535. https://doi.org/10.1007/s12369-015-0286-y

Alemi, M., Meghdari, A., & Haeri, N. S. (2017). Young EFL learners' attitude towards RALL: An observational study focusing on motivation, anxiety, and interaction. In *International Conference on Social Robotics.* Cham: Springer, 252–261. https://doi.org/10.1007/978-3-319-70022-9_25

Arend, B., Sunnen, P., & Caire, P. (2017). Investigating Breakdowns in Human Robot Interaction: A Conversation Analysis Guided Single Case Study of a Human-NAO Communication in a Museum Environment. *International Journal of Mechanical, Aerospace, Industrial, Mechatronic and Manufacturing Engineering*, *11*(5), 839-845. https://doi.org/10.5281/zenodo.1130169

Bartneck, C., & Forlizzi, J. (2004). A design-centred framework for social human-robot interaction. In *RO-MAN 2004. 13th IEEE international workshop on robot and human interactive communication (IEEE Catalog No. 04TH8759).* IEEE, 591–594. https://doi.org/10.1109/ROMAN.2004.1374827

Bartneck, C., Belpaeme, T., Eyssel, F., Kanda, T., Keijsers, M., & Šabanović, S. (2020). *Human-robot interaction: An Introduction.* Cambridge: Cambridge University Press. https://doi.org/10.1017/9781108676649.001

Belpaeme, T., Kennedy, J., Ramachandran, A., Scassellati, B., & Tanaka, F. (2018). Social Robots for Education: A Review. *Science Robotics,* 3(21): 1–9, https://doi.org/10.1126/scirobotics.aat5954

Bonarini, A. (2020). Communication in Human-Robot Interaction. *Current Robotics Reports*, 1(4): 279–285. https://doi.org/10.1007/s43154-020-00026-1

Bødker, S. (1996). Applying activity theory to video analysis: How to make sense of video data. *Human-Computer Interaction. Context and Consciousness-Activity Theory and Human-Computer Interface. Nardi, BA (Ed.) The MIT Press*, 147-174.

Chang, C. W., Lee, J. H., Wang, C. Y., & Chen, G. D. (2010). Improving the authentic learning experience by integrating robots into the mixed-reality environment. *Computers & Education*, 55(4): 1572–1578. https://doi.org/10.1016/j.compedu.2010.06.023

Derwing, T. M., Munro, M. J., & Carbonaro, M. (2000). Does popular speech recognition software work with ESL speech? TESOL Quarterly, 34(3): 592–603. https://doi.org/10.2307/3587748

Derwing, T. M., & Munro, M. J. (2009). Putting accent in its place: Rethinking obstacles to communication. *Language teaching*, *42*(4): 476–490. https://doi.org/10.1017/S026144480800551X

Gurunath Shivakumar, P., & Georgiou, P. (2018). Transfer Learning from Adult to Children for Speech Recognition: Evaluation, Analysis and Recommendations. *arXiv e-prints*, arXiv-1805.03322

Honig, S., & Oron-Gilad, T. (2018). Understanding and Resolving Failures in Human-Robot Interaction: Literature Review and Model Development. *Frontiers in psychology*, 9(861). https://doi.org/10.3389/fpsyg.2018.00861

Honkalammi, H-M., Veivo, O. & Johansson, M. (2022). Advice-giving between Young learners in robot-assisted language learning. *FRIAS Junior Researcher Conference: Human Perspectives on Spoken Human-Machine Interaction (SpoHuMa21)*, University of Freiburg (online), 15–17 November 2021. https://doi.org/10.6094/UNIFR/223816

Jakonen, T., & Jauni, H. (2021). Mediated learning materials: visibility checks in telepresence robot mediated classroom interaction. *Classroom Discourse,* 12(1–2): 121–145. https://doi.org/10.1080/19463014.2020.1808496

Jakonen, T., Veivo, O., Mutta, M., Maijala, M., Honkalammi, H.-M. & Johansson, M. (2022). ‘Am I saying it wrong?’ Progressivity-related troubles and instructional opportunities in

child-robot L2 interaction. [Manuscript submitted for publication]. School of Languages, University of Turku.

Jefferson, G. (2004). Glossary of Transcript Symbols with an Introduction. In G. H. Lerner (ed.). Conversation Analysis: Studies from the First Generation. Amsterdam: John Benjamins, 13–31. https://doi.org/10.1075/pbns.125.02jef

Jokinen, K., Nishida, M., & Yamamoto, S. (2010, February). On eye-gaze and turn-taking. In *Proceedings of the 2010 workshop on eye gaze in intelligent human machine interaction.* 118–123. https://doi.org/10.1145/2002333.2002352

Kanda, T., Hirano, T., Eaton, D., & Ishiguro, H. (2004). Interactive robots as social partners and peer tutors for children: A field trial. *Human–Computer Interaction*, 19(1-2): 61–84. https://doi.org/10.1207/s15327051hci1901%262_4

Kääntä, L. (2012). Teachers' embodied allocations in instructional interaction. *Classroom Discourse*, 3(2): 166–186. https://doi.org/10.1080/19463014.2012.716624

Lee, S., Noh, H., Lee, J., Lee, K., Lee, G. G., Sagong, S., & Kim, M. (2011). On the effectiveness of Robot-Assisted Language Learning. *ReCALL,* 23(1): 25–58. https://doi.org/10.1017/S095834401000273

Lyle, S. (2008). Dialogic teaching: Discussing theoretical contexts and reviewing evidence from classroom practice. *Language and Education*, 22(3): 222–240. DOI: 10.1080/09500780802152499

Mondada, L. (2019). Contemporary issues in conversation analysis: Embodiment and materiality, multimodality and multisensoriality in social interaction. *Journal of Pragmatics,* 145, 47–62. https://doi.org/10.1016/j.pragma.2019.01.016

Maijala, M. & Mutta. M. (2022). Teachers' Role in RALL Classroom Ecology. [Manuscript submitted for publication]. School of Languages, University of Turku.

Matsui, T., Tani, I., Sasai, K., & Gunji, Y. P. (2021). Dialogue Breakdown and Confusion between Elements and Category. In *Companion of the 2021 ACM/IEEE International Conference on Human-Robot Interaction.* ACM/IEEE. 276–279. https://doi.org/10.1145/3434074.3447175.

Moussalli, S., & Cardoso, W. (2016). Are commercial ‘personal robots’ ready for language learning? Focus on second language speech. In S. Papadima-Sophocleous, L. Bradley, & S. Thouësny (eds.), CALL communities and culture – short papers from EUROCALL 2016, 325–329. https://doi.org/10.14705/rpnet.2016.eurocall2016.583

Mutlu, B., Forlizzi, J., & Hodgins, J. (2006). A storytelling robot: Modeling and evaluation of human-like gaze behavior. In *2006 6th IEEE-RAS International Conference on Humanoid Robots.* IEEE. 518–523. https://doi.org/10.1109/ICHR.2006.321322

Okita, S. Y., Ng-Thow-Hing, V., & Sarvadevabhatla, R. K. (2011). Multimodal approach to affective human-robot interaction design with children. *ACM Transactions on Interactive Intelligent Systems,* 1(1), 1–29. https://doi.org/10.1145/2030365.2030370

Peura, L., Mutta, M., & Johansson, M. (2021). Playing with pronunciation. A study on a robot-assisted French pronunciation in a learning game. [Manuscript submitted for publication]. School of Languages, University of Turku.

Potamianos, A., & Narayanan, S. (2003). Robust recognition of children's speech. *IEEE Transactions on speech and audio processing*, 11(6): 603–616. https://doi.org/10.1109/TSA.2003.818026

Poulisse, N. (2019). *The Use of Compensatory Strategies by Dutch Learners of English.* Berlin, Boston: De Gruyter Mouton. https://doi.org/10.1515/9783110868975

Randall, N. (2019). A Survey of Robot-Assisted Language Learning (RALL). *ACM Transactions on Human-Robot Interaction,* 9(1): 1–36. https://doi.org/10.1145/334550.

Saerbeck, M., Schut, T., Bartneck, C., & Janse, M. D. (2010). Expressive robots in education: varying the degree of social supportive behavior of a robotic tutor. In *Proceedings of the SIGCHI conference on human factors in computing systems.* New York: ACM Press, 1613–1622. https://doi.org/10.1145/1753326.1753567

Serholt, S. (2018). Breakdowns in children's interactions with a robotic tutor: A longitudinal study. *Computers in Human Behavior*, *81*, 250-264. https://doi.org/10.1016/j.chb.2017.12.030

Sinclair, J. M., & Coulthard, R. M. (1975). Towards an Analysis of Discourse. Oxford: Oxford University Press.

Uchida, T., Minato, T., Koyama, T., & Ishiguro, H. (2019). Who Is Responsible for a Dialogue Breakdown? An Error Recovery Strategy That Promotes Cooperative Intentions From Humans by Mutual Attribution of Responsibility in Human-Robot Dialogues. *Frontiers in Robotics and AI*, *6*. https://doi.org/10.3389/frobt.2019.00029

van den Berghe, R., Verhagen, J., Oudgenoeg-Paz, O., van der Ven, S., & Leseman, P. (2019). Social robots for language learning: A review. *Review of Educational Research*, 89(2), 259–295. https://doi.org/10.3102/0034654318821286

Vertegaal, R., Slagter, R., van der Veer, G., & Nijholt, A. (2000). Why conversational agents should catch the eye. In *CHI '00: Extended Abstracts on Human Factors in Computing Systems*. New York: ACM Press, 257–258. https://doi.org/10.1145/633292.633442

Walsh, S. (2015). *Classroom Interaction for Language Teachers*. English Language Teacher Development Series. Alexandria, Virginia: TESOL International Association.

Webbink, P. (1986). *The power of the eyes*. New York: Springer Publishing Co.