”Am I saying it wrong?”: Progressivity-related troubles and instructional opportunities in child-robot L2 interaction

Teppo Jakonen, Outi Veivo, Maarit Mutta, Minna Maijala, Hilla-Marja Honkalammi & Marjut Johansson

Viittausohje:

Jakonen, T., Veivo, O., Mutta, M., Maijala, M., Honkalammi, H-M., & Johansson, M. (2023). ”Am I saying it wrong?”: Progressivity-related troubles and instructional opportunities in child-robot L2 interaction [“Sanonks mä sen väärin?”: Toiminnan progressiivisuuden ongelmat ja ohjauksen mahdollisuudet lasten ja robotin välisessä kakkoskielisessä vuorovaikutuksessa]. *Prologi – Viestinnän ja vuorovaikutuksen tieteellinen aikakauslehti*, *20*(1), tulossa. https://doi.org/10.33352/prlg.120961

To cite this article:

Jakonen, T., Veivo, O., Mutta, M., Maijala, M., Honkalammi, H-M., & Johansson, M. (2023). ”Am I saying it wrong?”: Progressivity-related troubles and instructional opportunities in child-robot L2 interaction. *Prologi – Journal of Communication and Social Interaction*, *20*(1), forthcoming. https://doi.org/10.33352/prlg.120961

Prologi

– Viestinnän ja vuorovaikutuksen tieteellinen aikakauslehti

journal.fi/prologi/

ruotsiksi: Prologi – Tidskrift för Kommunikation och Social Interaktion
englanniksi: Prologi – Journal of Communication and Social Interaction

Julkaisija: Prologos ry.

Prologos

Avoin julkaisu / Open Access
ISSN 2342-3684 / verkko

Article

Prologi, 20(1)
forthcoming
https://doi.org/10.33352/prlg.120961

"Am I saying it wrong?": Progressivity-related troubles and instructional opportunities in child-robot L2 interaction

Teppo Jakonen
PhD, Academy Research Fellow
School of Languages and Translation Studies
University of Turku
teppo.jakonen@utu.fi

Outi Veivo
University Lecturer
School of Languages and Translation Studies
University of Turku
outqvi@utu.fi

Maarit Mutta
Associate professor (PhD)
School of Languages and Translation Studies
University of Turku
maarit.mutta@utu.fi

Minna Maijala
Professor
School of Languages and Translation Studies
University of Turku
minna.maijala@utu.fi

Hilla-Marja Honkalammi
Research Assistant
School of Languages and Translation Studies
University of Turku
himhon@utu.fi

Marjut Johansson
Professor
School of Languages and Translation Studies
University of Turku
marjut.johansson@utu.fi

received 17.8.2022 / accepted 23.10.2023 / published 30.11.2023

Abstract

Participants' orientation to progressivity (i.e., the smooth and non-delayed progress of sequences of action) is a fundamental feature of human social interaction. We explore how such progressivity is maintained in human-robot interaction (HRI) by drawing on c. 14 hours of video recordings showing small groups of primary school children interacting with Nao, a programmable humanoid robot. Facilitated by a teacher, the children in our data are completing a short robot-assisted language learning lesson aimed at training English vocabulary and oral skills at a Swedish-speaking school in Finland. We investigate how the teacher and the students address emerging troubles in a word repetition sequence, which the robot is programmed to carry out with a student. Our analysis focuses on two kinds of troubles related to sequence closure: the robot's so-called "third" turns that either do not ratify the student's just-prior word repetition as "correct" or are (treated as) incongruent within the sequential context. We show how the human participants make sense of such troubles, recruit the teacher's assistance to secure the progress and eventual closure of the action sequence, and orient to pronunciation instruction in situated ways. The results shed light on how children accommodate to, and are socialised into, human-robot interaction.

KEYWORDS: child-robot interaction; language learning in interaction; Nao robot; progressivity; social robots

Introduction

During the past couple of decades, digital technology has fundamentally expanded the means and contexts of social interaction and communication. Besides the nearly ubiquitous presence of video calls and various forms of online communication, we are increasingly more likely to encounter - and interact with - communicative technology that is based on artificial intelligence (AI), natural language processing and generation (NLP/NLG), and voice recognition. These developments are not without problems. Hepp (2020) argues that communicative bots such as Apple's Siri or Amazon's Alexa and work-related bots that can be used, among other things, to generate automated news content are making way for a "quasi-communication" of sorts. At the same time, communicative bots and social robots such as Nao and Pepper are challenging the fields of social interaction and communication studies by presenting a need to rethink the relationship and agency distribution between humans and technology. In this rapidly changing technological landscape, understanding how human-machine communication (HMC), i.e., "people's interactions with technologies designed as communicative subjects, instead of mere interactive objects" (Guzman & Lewis, 2020, p. 71), is interactionally organised is key to assessing the potential future implications of AI in the society. Among other things, this requires paying attention to how humans orient to social and communicative robots in interaction (see Laaksonen et al., 2020), and considering when and how humans need to facilitate interaction and adapt their routinised social practices to make technology "work".

In this article, we explore such practices of human sense-making of technology by investigating how small groups involving a teacher and two to four students interact with a social robot (Nao) that uses a language learning application

(Elias) in a school context. To date, social robots have been extensively researched in the multidisciplinary and growing field of human-robot interaction (HRI). In recent years, studies have begun to problematise and reconceptualise the notion of interaction in HRI from a variety of human and social perspectives (Dautenhahn, 2007). One such perspective is offered by an emerging set of studies such as ours that draw from the microanalytical and qualitative tradition of conversation analysis (CA) and explore HRI as cooperative, sequentially organised, and accountable interaction (for other recent examples, see Pelikan, 2023; Pelikan & Broth, 2016; Pitsch, 2020; Rollet & Clavel, 2020; Skantze, 2021). One potential contribution that CA can offer to the fields of HMC and HRI is a fine-grained picture of how humans interact with technological devices that produce and make sense of (human) action with the help of pre-scripted conversation models and algorithms (Johansson, 2021). As Pelikan (2023) has recently argued, such an empirically grounded understanding can shed light on how HRI is similar or different to human-human interaction and, ultimately, help design AI-based robots that can better coordinate action with human participants.

The interactional and institutional context of our study can be seen as a technology-rich classroom in which the social robot Nao is used in combination with a language learning application to instruct a second language (L2) to children. Research in robot-assisted language learning (RALL) is still very much emerging, but existing studies seem to suggest that social robots can take on a dual identity in such settings, being sometimes treated as a participant-in-interaction - such as a teacher, tutor, or peer - and at other times as a technological object/tool (Kanda et al., 2004; Peura & Johansson, 2023; Randall, 2019). However, much less is known about L2 interactional practices between children and social robots (but see Honkalammi et al., 2022; Peura & Johansson, 2023), including how robot features such as natural language processing, its embodied conduct such as gestures and movements, and its non-lexical sounds (Pelikan, 2023) offer interactional and instructional resources. What seems to emerge from pedagogical studies is that instructional gains are not likely to emerge by just bringing a robot into a classroom, but - like any other pedagogical tool - learning outcomes depend on how it is used and accommodated to as part of classroom instruction (see Randall, 2019; van den Berghe et al., 2019). This means that teachers and students need a form of *robot literacy* (Peura & Johansson, 2023) – and perhaps its relevance is not limited to the educational domain in a world that is arguably becoming ever more technology-intensive.

Our contribution to studies of HRI and educational robotics is an investigation of how participants deal with troubles of progressivity of instructional interaction with the robot – in broad terms, moments when interaction “gets stuck”. One motivation for focusing on progressivity is that, as we shall argue in the next section, orientation to progressivity is a ubiquitous feature of human-human interaction; yet it is something that stretches at least the current capabilities of AI-driven robots and bots, which struggle with interactional repair (see Stommel et al., 2022). Furthermore, a large part of instruction in a broad range of educational contexts is accomplished through three-turn activity sequences (IRE), which account for up to 60% of instructional talk according to Lyle (2008, p. 225). Teachers’ and students’ participation in such instructional sequences is configured by principles of sequential relevance and progressivity, which is why it is important to understand the ways in which - and the extent to which - artificial intelligence and social robots can operate in such interactional and pedagogical activity environments. More specifically, we examine the following research questions:

1) How do participants make sense of and resolve troubles in the progressivity of action sequences in task-related L2 interaction with a social robot?

2) How do such troubles occasion teaching and learning activities?

By addressing these questions, we thus aim to contribute a close multimodal analysis of how human participants (a) facilitate human-robot interaction and orient to the robot as a particular kind of participant, and (b) turn interactional problems into language learning opportunities. In what follows, we will first describe how progressivity is conceptualised in the CA tradition and sketch how we approach it in the empirical study by zooming in on progressivity troubles related to so-called sequence-closing third actions by the robot (section 2). We will then describe our data, method and analytical collection (section 3) and proceed to the empirical analysis (section 4) and the discussion of findings (section 5).

Progressivity in human-human and human-robot interaction

In the conversation analytic framework, the notion of progressivity is a fundamental feature of the incremental and temporal nature of social interaction. In Schegloff's (2007) view, "[m]oving from some element to a hearably-next-one with nothing intervening is the embodiment of, and the measure of, progressivity" (p. 15). The "elements" between which progress matters for participants can be of very different granularity, ranging from individual sounds, words and clauses to turns that make up a sequence of actions. Disruptions of progressivity often emerge in moments when participants conduct repair by addressing emergent "problems in speaking, hearing, and understanding" (Schegloff et al., 1977, p. 361; see also Dingemanse et al., 2014). When the current speaker self-initiates repair by, for example, cutting off a word or beginning a word search, they at the same time halt the progressivity of their ongoing turn (Kitzinger, 2012). The syntactic progressivity of a turn may also be interrupted when the speaker needs to coordinate talk with a simultaneous physical activity (Hofstetter et al., 2021).

Particularly relevant for our purposes is the notion of progressivity of adjacency pair-based sequences in which some first action makes conditionally relevant a particular kind of second action – such as a question and an answer, reciprocal greetings or a request and the granting of the request. Stivers and Robinson (2006) have argued that participants orient to a social and structural pressure to bring ongoing question-answer sequences to completion to further the activity at hand: in other words, they display a preference for sequence progressivity. One way in which the preference becomes visible is in moments when for some reason the second action is not forthcoming. Analysing question-answer sequences in everyday and institutional (doctor-patient) multi-party interactions, Stivers and Robinson (2006) showed that participants who had not been selected to answer a question nevertheless intervened to provide an answer if the selected participant was observably having difficulty responding, claimed inability to answer, or was somehow temporarily unavailable. Stivers and Robinson (2006) suggested that this kind of preference for sequence progressivity is not only structural but "very much a social issue" (p. 387), being tied to participants' interactional rights and obligations in the situation.

Digital and robotic technologies can introduce new kinds of progressivity-related troubles. Interaction with a robot may for instance be disturbed due to the robot's software or hardware problems or the robot's failure to interact with its environment (Honig & Oron-Gilad, 2018). Interaction with a social robot or a conversational agent can also lead to situations where the human participant simply gets frustrated and loses motivation to further communicate with artificial intelligence (Matsui et al., 2021). Sometimes, the human participants disengage from such situations in abrupt ways, displaying less interactional tact than in typical closings of human-human encounters (Licoppe & Rollet, 2020). On the other hand, in robot-assisted learning contexts, children have been observed to display interactional perseverance when interacting with a social robot, and when problems occur, other human participants can provide an important interactional resource to resolve them (Honkalammi et al., 2022; Veivo & Mutta, 2022).

Our interest in how participants deal with moments of trouble in the progressivity of sequences in HRI parallels a recent study in a slightly different context by Fischer et al. (2019), exploring interaction with the virtual assistant Alexa in households. The authors analysed questions and commands presented to Alexa, and showed how the human participants worked to ensure sequence progressivity past troubles introduced by Alexa's non-answer responses such as "I'm having trouble understanding". Similarly, Stommel et al. (2022) found that human participants completing a health survey with a Pepper robot adapted their conduct in miscommunication situations that needed repair, for example, by repeating their turns with clearer or louder articulation. Occasionally, the human participants settled for "second best answers" and thereby displayed an orientation to the progressivity of the survey activity as opposed to repairing interactional trouble.

In this article, we zoom in on progressivity troubles in sequences of a word repetition activity, which the robot is programmed to carry out with a student as a series of three-part extended sequences consisting of (a) a robot-uttered word while an image of that word is shown on an accompanying laptop screen, (b) a student repetition of the word, and (c) the robot's ratification of the repeated word as 'correct', which enables the student to move on to the next word and sequence. The programmed repetition sequence is similar to the initiation-response-evaluation/feedback sequence (IRE/F, Lyle 2008; Mehan, 1979) that is commonly found in various kinds of instructional interactions. However, as we will argue in this article, the situated accomplishment of the repetition sequence is an interactional achievement that requires adaptation and facilitation from the human participants. For purposes of contrast to the analysis, Extract 1 (on the next page) illustrates a typical trajectory of successful completion of the repetition sequence, showing how S2 repeats the word "brother" uttered by the robot (R).

Extract 1. Completing a repetition sequence

```
      01       (1.6) ((S1 selects a word on laptop))
->    02   R   brother (.) beep
      03       (0.8)
      04   T   mhm ((nods to S2))
      05       (3.3) ((T moves robot closer to S2))
->    06   S2  brother
      07       (1.1)
->    08   R   brother (.) beep
      09   T   mhm? (0.4) bra? ((T nods; S1 reaches for laptop))
                          'good'
      10       (0.7)
      11   S1  °till följande?° ((to T))
                'to the next one?'
```

As S1 selects a word on a laptop, the robot initiates the repetition sequence by uttering the selected word ("brother", l. 2). This makes relevant a student's response in the form of saying the target word aloud (l. 6), which the teacher here facilitates by allocating the turn to S2 (l. 4) and by positioning the robot for optimal audio recognition (l. 5). The student response is followed by the robot's third turn (l. 8), which, akin to the IRE/F sequence, is hearable as a turn that addresses the correctness or adequateness of the student response. Upon recognizing a student saying the target word in the second position, the robot is programmed to repeat the word once more or utter "okay", and make a longer "OK" beep. Often (but not here), the light in the robot's eyes will also change colour. In addition to these robot conduct, the Elias laptop application will also provide written feedback of correctness (such as displaying the text "Awesome") as well as enable movement forward in the activity. The robot's verbal third turn thus parallels how teachers' verbatim repeats of a preceding student response in the IRE sequence are treated as a signal of positive evaluation (Hellermann, 2003; Margutti & Drew, 2014; Park, 2014), and in addition to that, the technological system is designed to provide signals that have a distinctively positive valence for an adequate student response.

IRE is not only part of the programming of the activity and robot conduct, but it is also the interactional organisation that the human participants in Extract 1 orient to and strive to accomplish. This can be seen in how here the robot does not seem to operate as it is programmed to do: its eyes do not change colour and it does not give the distinctive "OK" beep at line 8 but instead provides a similar "regular" beep as when introducing the target word at line 2. Nevertheless, the participants treat the robot's third-turn repetition of "brother" turn at line 8 as a ratification of the response, which is visible in how the teacher affirms it with an evaluative token ("bra", *good*) at line 9. These affirmations are in our data mostly given in Swedish, the language of instruction in the classroom (see Maijala &

Mutta, in press). Knowing when a repetition sequence reaches closure is also needed by the participant whose task it is to click the next word on the laptop. Here, S1 orients towards the laptop immediately after the robot's turn in line 8, displaying an orientation to its ratifying nature even before the teacher has uttered "bra".

In contrast to line 8, our empirical analysis focuses on the robot's actions that are *not* treated as a progressivity-securing third turn that would complete the ongoing repetition sequence. This includes the robot's silences or stand-alone embodied conduct such as nods, which participants treat as an indication of trouble. Contrary to human interaction, the robot's stand-alone nod does not mean acceptance of the student's answer: it is a reaction to detecting a human voice but not the target word. For members, such unexpected (Gehle et al., 2015) or inapposite robot conduct presents a need for interpretative work regarding possible causes of the trouble. As we aim to show, in our data, the human participants' sense-making of the robot's actions - and the robot as a particular kind of participant - largely takes place in and through insert sequences to the three-part repetition sequence. These moments can also become a site for pedagogical work in the form of guidance, suggestions and interpretations of the "meaning" behind the robot actions that the teacher provides to the children to solve the emergent troubles (see also Maijala & Mutta, in press).

Data and method

Our video-recorded data corpus (13 hrs 42 min) includes robot-assisted language teaching situations, collected in the spring of 2019 in four Swedish-speaking primary schools in southern Finland. The participants, a teacher and two to four students at a time, are completing a series of English language activities with a NAO6 robot that is being operated with the Elias language learning application. NAO is a small humanoid robot (height: 57 cm, weight: 5.4 kg) with voice recognition, speech synthesizer, gaze recognition and the ability to move. Elias is a Finnish commercial application that has pre-programmed short (c. 15-20 mins) lessons on fixed themes in a range of different languages, but users can also program interactive tasks with it. The application is operated with a computer or a tablet, which is connected to the robot. Many of the pre-programmed language lessons focus on vocabulary learning, with Figure 1 showing the activities in a typical lesson.

Altogether, we filmed 42 teacher-student constellations (a total number of 111 participants) completing one 15–20-minute lesson focusing on either family words or colours. At the time of the data collection, the students, aged 10–13, had been learning English as a foreign language at their school for 1.5–3.5 years (having started it in Year 3). This was the first occasion that the students encountered and interacted with a social robot in the school. Before data collection, permission for data collection was granted by the local school board and rectors, and informed consent was provided by the parents on behalf of the minor participants. During the activity, the teacher did not interact directly with the robot but instead facilitated child-robot interaction, for instance, by moving the robot toward the student whose turn was to speak to the robot and by advising students in moments of (technical) problems.

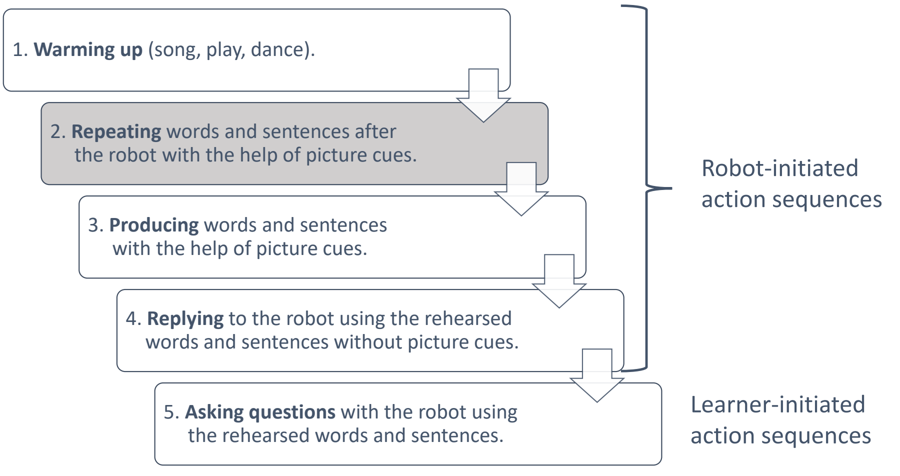

Figure 1. The course of a typical lesson in the robot application used in the corpus (cf. Veivo & Mutta, 2022).

In the present study, we focus on data from the repetition activity (highlighted in Figure 1 and described in Extract 1). We began our work by examining a large collection showing moments of interactional trouble in the dataset, which was analysed in an earlier pilot study (Honkalammi et al., 2022). Within that collection, our interest was directed to how participants made sense of the robot's feedback to students. Based on that, we created a collection of 100+ repetition sequences in which the progressivity of the sequence was suspended between a student's repetition and the robot's subsequent sequence-closing third turn. This involves situations in which the participants orient to the third turn as delayed, missing, sequentially incongruent or otherwise problematic. In this article, we present our observations on these kinds of third-turn progressivity troubles by analysing a selection of data extracts that illustrate participants' recurring practices in addressing the troubles.

The extracts are transcribed using Jefferson's (2004) and Mondada's (2019) transcription conventions. Besides talk, the NAO robot produces both electronic (beeps) and mechanical sounds when it moves (nods, readjusts its body, puts its hands on hips, etc.), which can be cumbersome to transcribe. Given that the deliberately designed beeps are a central sense-making resource for human participants (Pelikan, 2023), in our setting to assess the progress of the repetition sequence, we have transcribed them in the verbal transcript (as in Extract 1). Similarly, we have annotated the robot's embodied conduct in the multimodal transcription, but for reasons of readability have not included the non-communicative mechanical sounds that accompany the robot's movements within the transcript of talk.

Analysis

In this section, we show how participants conduct interactional work to bring the pre-designed three-part repetition sequence to closure in moments where its progressivity is jeopardised. We begin by discussing the role of technological and material resources for progressivity in the situation, and the recruitment of (teacher) assistance (Extract 2). We then analyse how word repetitions are modified when no third turn seems forthcoming (Extract 3) and how the three-part repetition sequence can occasion pronunciation instruction (Extracts 4a and 4b).

Extract 2 (on the next page) demonstrates how two types of robot conduct, a missing response (l. 4) and a nod (l. 6), are treated as conduct that suspends sequence progressivity when they occur in the third position of the three-part repetition sequence in our data. It also illustrates how the students in our data orient to the teacher as the facilitator of the activity by recruiting the teacher's help when obstacles to progressivity emerge. The extract begins as the robot initiates the repetition sequence by offering the (alternative) words "mother" and "mum" (l. 1). Altogether, S2 needs to utter the word "mother" three times in the extract before the sequence is brought to closure.

S2 leans closer to the robot (fig. 2.1) and repeats the word "mother" (l. 3). In contrast to the robot's utterance (and many varieties of English), S2 utters the word using what approximates an alveolar */d/* sound instead of the voiced dental fricative */ð/* in the international phonetic alphabet (IPA). Instead of signalling acceptance by way of a beep or a verbal token, the robot remains unresponsive during the ensuing long silence, which halts the progression of the sequence. Approximately 2.5 seconds into the silence, S2 shifts their gaze away from the robot towards the teacher. The shifting gaze orients to the robot's conduct as trouble and recruits the teacher's assistance to resolve the emerging trouble. Similar to an open repair initiation, it does not provide the recipient an analysis of the nature of the trouble. The teacher's response to S2 is similarly embodied, a nod (fig. 2.2) that serves as a request for S2 to retry and repeat the target word (l. 4).

A similar suspension of sequential progress emerges as S2 utters "mother" for the second time (l. 5), pronouncing it similarly as on the first time but with a slightly rising intonation. This time, the robot first nods and repositions itself during the ensuing long silence (l. 6) but does not say anything. This indicates that it has registered human talk but that it is not equivalent to the target word, i.e., that the robot has not "heard" S2's turn as the correct word. Again, S2 recruits assistance by shifting their gaze to the teacher, who verbally requests yet another retry (l. 7). Finally, the student's third attempt (l. 8) is met with a third turn that completes the repetition sequence, as the robot ratifies the answer by way of nodding, a verbal token ("okay") and a sound signal (beep) at line 10. Notice that the teacher echoes the ratification and orients to it as a signal that participants can move on in the activity (l. 11–12), which is demonstrated by the way the teacher turns the robot and points towards the student (S1) whose turn it will be to repeat the next word.

Besides showing a typical way of recruiting teacher assistance in our data, Extract 2 illustrates a routinised embodied and material configuration of child-robot interaction. When addressing a turn-at-talk to the robot, S2 leans closer to the robot. Conversely, S2 marks disengagement from interaction with the robot by reclining into a more typical seating position (fig. 2.3). Such a practice helps to provide maximally clear and audible verbal turns to ensure the robot's voice recognition can process them. At the same time, these different embodied configurations provide other participants a way to interpret each other's engagements and addressees of a turn-at-talk. In Extract 2, S2 reclines back 0.2 seconds after uttering the target

Extract 2. Recruiting help for troubles in sequence progressivity

```
01 ROB   mother (.) mum¤ (.) beep¤
   rob                 ¤nods-----¤
02       (0.5)*(0.9)¤*(0.2)
   s2         *......*leans forward->
   rob              ¤hands on hips->
03 S2    mother¤# ((IPA: mʌdər))
   fig          #fig2.1
   rob       ->¤
04       (0.2)*(1.0)   *(1.3)*(0.3)^(0.3)#*(0.4)^*(0.2)
   s2       ->*reclines*     *gz-teacher--*      *...->
   tea                             ^nods--------^
   fig                                   #fig2.2
05 S2    mother,* ((IPA: mʌdər))
   s2    .......*leans fwd->
06       (1.0)¤(0.2)*(0.4)¤(0.8)*(0.6)¤(0.2)*(0.4) ¤
   rob        ¤nods-------¤           ¤lowers hands¤
   s2             ->*reclines---*           *gz-teacher->
07 TEA   >en ^gång till< hh*
         'one more time'
   tea       ^smiles, nods-->
   s2                     ->*...->
08 S2    .hh* mot^her, ((IPA: mʌdər))
   s2    ...*smiles, leans forward->
   tea        -->^
09       (0.3)*(0.7)¤(0.1)*
   s2       ->*reclines---*
   rob              ¤nods->
10 ROB   okay¤& (0.5) [beep]
11 TEA                [  °o]kej?°¤^
   rob     ->¤adjusts hands------¤
   s1         &gaze to laptop->
   tea                            ^turns robot twd s1->
12       (1.0)^*#(0.5)*&^(0.2)
   tea     -->^         ^points at s1->
   s1                 ->&gaze to robot->
   s2          *smiles*
   fig          #fig2.3
```

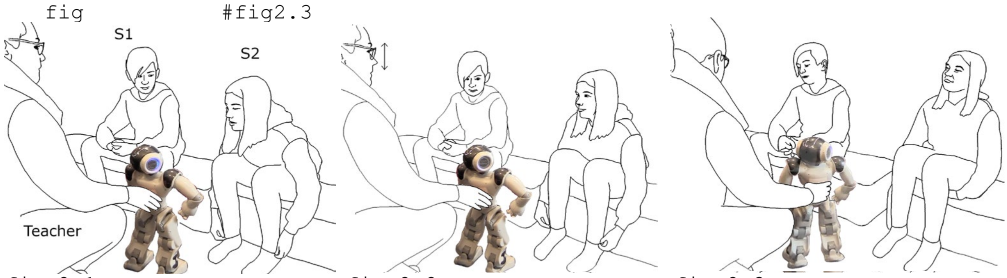

fig 2.1 fig 2.2 fig 2.3

```
13 TEA   °sen° &vidare?^
         'then forward'
   tea                ->^
   s1        ->&gaze to laptop>>
14 S1    ((clicks forward on laptop))
```

word for the first time (l. 3), indexing an orientation to the task as complete. Conversely, after the second attempt (l. 5), S2 waits 1.2 seconds before reclining, thus projecting possible further trouble and a need to repeat the word once more. After the third attempt to utter the word (l. 8), S2 leans back soon after the turn, and here the non-expectedness of trouble potentially may orient to the earlier teacher's request to say the word *one more time* (l. 7). In addition to these embodied actions, the extract also showcases the role of the laptop with which the human participants concretely close one repetition sequence and transition to another one. Thus, an orientation to laptop use and computer commands – as in S1's gaze shift to it when the robot says "okay" (l. 10) and the teacher's turn at line 13 ("vidare", *forward*) – can become a measure of (restored) activity progression.

When a participant faces a need to repeat their utterance upon a missing third turn, their way of modifying (or not) the repeated utterance shows an analysis of the possible cause of the trouble. In Extract 2, both repeated words ("mother") were uttered with similar volume and pronunciation but with slightly rising intonation. In contrast, Extract 3 (on the next page) illustrates how repetition of the target word with a similar pronunciation but in a louder voice treats the missing third turn as a problem related to the robot's voice recognition capabilities, which may potentially be resolved through a clearer articulation. Here, a group of three students and a teacher are repeating the same word as in Extract 2, but in contrast, S1 repeats both alternative options ("mother" and "mum") offered by the robot. In addition to this, differences in accent may impede the robot's voice recognition in Extract 3. Whereas the robot utters the short form of "mother" more or less as "mum" is pronounced in certain dialects of British English (as */mʌm/* in IPA), S1 utters it throughout the extract approximately as the short form of "mother" is said in some dialects of American English as "mom" (*/ma:m/*). These have been marked in the extract.

S1 responds to the robot's sequence initiation[1] (l. 1) altogether four times before the robot finally acknowledges and ratifies the response (l. 14). The first attempt (l. 2) is delivered early, taking place in overlap with the robot sounds, and partially speeded up (">mother<"). During the following silence, the robot nods but does not utter anything, which indicates that the robot has recognised human talk but not the target word. S1's subsequent retries are modified versions of the initial attempt, in ways that orient to clarity of the verbal utterance. S1 first reduces speech rate and increases stress on the first sounds of both words (l. 4). As this does not mobilise a ratifying third turn from the robot, S1 emphasises the first syllable of "mother" even further (l. 6). Despite these modifications, the robot does not say anything but merely nods during the ensuing silence.

Unlike in Extract 2, here S1 initially addresses the lack of progressivity individually for the first and second occasions when the sequence does not proceed to the robot's third turn (l. 3 and 5). It is only after the third attempt at uttering the word that S1 recruits the teacher's assistance by way of a gaze shift (fig. 3.1) and a verbal help request (l. 8). The turn attributes the progressivity trouble to the pronunciation of the target word(s). Notice that the teacher first responds to S1's gaze shift by offering a nod to encourage a retry (as in Extract 2). The teacher then verbally refutes (l. 11) S1's doubts about having made pronunciation errors (l. 10).

[1] The robot unusually makes three beeps at line 1, possibly due to a technical glitch.

Extract 3. Modifying the delivery of a word

```
01 ROB    beep ¤mother (.) mum (.) beep ¤beep [beep    ] ((IPA: /mʌm/))
02 S1                                        *[>mother<] (.) mom, ((/mɑːm/))
   tea         ¤moves robot closer to s1¤
   s1                                        *gaze to robot->
03        ^ (0.4) &(1.1)& (0.3) ^
   s1     ^smiles, tightens lips^
   rob            &nods-&
04 S1     &mother& (.) mom.& ((IPA: /mɑːm/))
   rob    &turns-&         &adjust hands, nods->
05        (1.0)&(1.3)
   rob       ->&
06 S1     MOther mom. ((IPA: /mɑːm/))
07        ^(0.7)&(0.8)&§(0.7)*(0.7)^
   s1     ^smiles, tightens lips---^
   s1                      ->*gaze to teacher->
   rob          &nods-&
   s2                  §smiles->
08 S1     #+>&säger+ jag *fel?<&
            'am I saying it wrong?'
   s1                  ->*gaze to robot->
   tea     +nods---+
   rob       &adjusts hands----&
   fig    #fig3.1
09        (0.6)
10 S1     [>eller va<]
          'or what'
11 TEA    [°( )°     ] (0.4) &°nej°&
                              'no'
   rob                       &nods-&
12 S1     mo^ther (.) <mom.># ((IPA: /mɑːm/))
   s1       ^leans closer to robot->
   fig                      #fig3.2
```

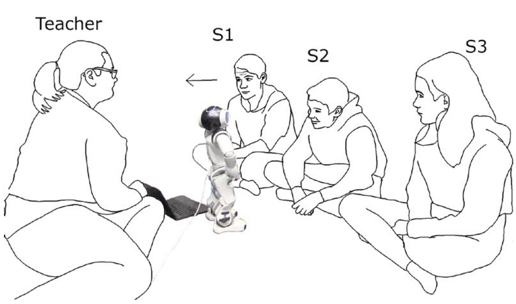

fig 3.1

fig 3.2

```
13        (1.0)
14 ROB    &mother& (.) mum (.) ^[↑psh ↓psh beep] ((IPA: /mʌm/))
15 S1                           [°( )°         ]
   rob    &nods--&
   s1                         ->^
16 S1     man måste säg(a) [mum ]§ ((IPA: /mʌm/))
          'one must say mum'
17 ROB                     [beep] beep
   s2                          ->§
18 TEA    eller ↓mother
          'or mother'
```

When repeating the response for the third time (l. 12), S1 also leans closer to the robot (fig. 3.2). Compared to the earlier repetitions, the utterance contains less word-initial stress but is produced in clearer voice quality, partly achieved through the elongation of "mom". The robot finally nods and ratifies the attempt by uttering the target words and by providing the longer "OK" beep (l. 14). The reduced distance may aid the robot's voice recognition.

It is interesting that S1 orients increasingly to the possibility of pronunciation trouble, as the three-part repetition sequence drags on. Besides the help request, this becomes observable in S1's post-sequence account (l. 16), an inference of how the target word *must* ("måste") be said. At line 16, S1 seems to orient to this difference in vowel sound quality and length between their earlier utterances and the robot's model. In this sense, the account orients to pronunciation learning as an institutional purpose of the activity, and demonstrates an ability to distinguish between the phonemes that distinguish "mom" from "mum".

In Extract 3, an orientation to pronunciation differences was made visible by the student doing the repetition activity. In contrast, Extracts 4a and 4b exemplify how the activity can pave the way for pronunciation-related instructional actions by the teacher. In Extract 4a, possibly because of some kind of technical glitch, the robot recognises and repeats S2's response as a different word ("butter", l. 7) from the initial prompt ("father", l. 1). However, the teacher treats this as a relevant occasion for pedagogical correction and suspends progressivity by treating the robot's incorrect "hearing" (l. 9) as a third action that does not close the repetition sequence (for teacher roles in RALL activities, see also Maijala & Mutta, in press).

Extract 4a. Suspending progressivity for pedagogical correction

```
01 ROB    father (.) dad (.) beep uncle beep
02        (0.2)*(0.8)^(0.6)
   s2          *gaze TEA->
   tea               ^gaze S2->
03 TEA      *^bara (.) father* eller dad.
              'just father or dad'
   tea     ->^
   s2     ->*leans twd robot-*
04        *(0.8)*(0.3)
   s2     *leans*
05 S2     father#* ((IPA: fʌdər))
   s2              *straightens posture, gaze S1/laptop->
   fig            #fig4.1

06        (0.8)#(0.4)*
   s2                ->*gaze ROB->
   fig          #fig4.2
```

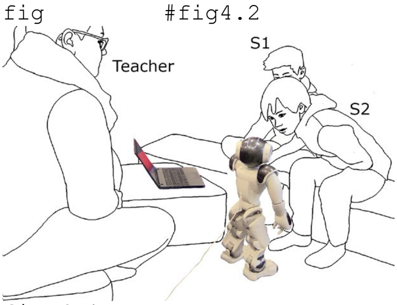

fig 4.1


fig 4.2

```
07 ROB    butter (.) ¤beep
   s1               ¤gaze to laptop->
08         (0.7)
09 TEA    hh ^nu *¤hörde han-#
              'now he heard'
   tea        ^gaze S2->
   s2           ->*gaze TEA>>
   s1            ->¤gaze TEA>>
   fig                         #fig4.3
```

```
   fig 4.3
10         (0.9)
```

The robot's sequence initiation (l. 1) is unusual in that it also provides the word "uncle" in addition to the word pair "father" and "dad", which could be a software malfunction. It interrupts the progressivity of the interaction, and during the following silence, S2 looks at the teacher, who in turn instructs them to ignore the word uncle (l. 3). S2 leans closer to the robot (fig. 4.1), suspending the movement and resuming it after the teacher's turn has reached completion. Similar to Extract 2, S2's response is pronounced more or less as "fudder" (IPA: /*fʌdər*/), with a short first vowel and the alveolar plosive /*d*/ as opposed to the voiced dental fricative /*ð*/. After uttering the response, S2 leans back in their seat and turns their gaze towards S1 and the laptop (fig. 4.2). Such an embodied disengagement from interaction with the robot and orientation to the laptop projects moving on to the next repetition cycle (involving S1) as the next relevant activity.

However, S2 returns their gaze to the robot at the end of the silence (l. 6), just as the robot begins the third turn. The turn consists of the utterance "butter" and a ratification beep, but the teacher orients to the robot's ratifying response as problematic and suspends the progression of the repetition sequence. The teacher's turn at line 9 is an example of what Koshik (2002) has described as "designedly incomplete utterances" (DIU). DIUs are turns that teachers use to elicit knowledge displays from students by providing an utterance that the student is expected to complete syntactically. As an instructional action, the DIU at line 9 identifies a problem and makes S2's completion relevant at line 10. Instead of explicitly correcting the student, it thus leaves it to the student to figure out how to correct the verbal turn so that the robot could "hear" it correctly. In our dataset, the teacher typically gives these kinds of clues in the students' L1 to mediate understanding problems (see Maijala & Mutta, in press). At the same time, treating the trouble as the robot's hearing problem instead of a student's pronunciation problem avoids implying that the student has made a mistake. Avoiding providing overt critique is an example of what Seedhouse (1997) has described as "the missing 'no'" in classroom interaction.

Nevertheless, S2 does not complete the DIU during the silence at line 10. Extract 4b shows how the situation continues.

Extract 4b. Instructing pronunciation

```
10          (0.9)
11 TEA      någo(t) annat,
            'something else'
12          (1.0)*(2.2)
   s2           *gaze ROB, leans closer->
13 S2       father ((IPA: fʌdər))
14          (0.3)*(0.9)
   s2         ->*straightens posture->
15 ROB      butter (.) beep*
   s2                   ->*gaze TEA->
16 S1       ¤°hh he ^he° (.) ¤*smör# hörde^ [han* °(  )°
                                'he heard butter'
17 TEA                                      [↑JOO hh #han¤ hörd(h)e *sm(h)ör.
                                              'yeah he heard butter'
18 S2                                       [(hh ja)
   s1       ¤smiles----------¤gaze TEA, smiles-----------¤gaze ROB->
   tea             ^gaze S1, grins-------^smiles->
   s2                      ->*gaze S1----------*gaze TEA, smiles---*gaze ROB->
   fig                               #fig4.4a/b        #fig4.5
```

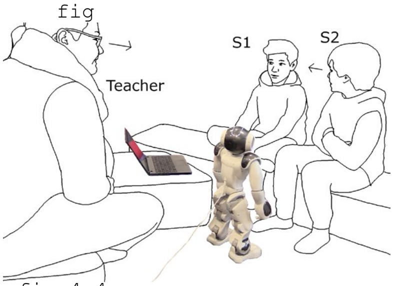

fig 4.4a

fig 4.4b

fig 4.5

```
19 TEA      .hh^ ska du prova säga (.) >en gång till<
               'will you try say one more time'
   tea       ->^
20 S2         *°okej,° (1.0) father ((IPA: fʌtər))
   s2       ->*leans twd robot->
21            *(0.8)   *(0.3)*(0.2)
   s2       ->*reclines*     *gaze-TEA->
22 ROB      butter* (.) beep*^¤
   s2           ->*gaze ROB-*thrusts head back->
   tea                       ^smiles->
   s1                         ¤smiles->
23          (0.7)
24 S2       °↑mhm°#* (  )^
   s2           ->*
   tea                ->^reaches laptop->
   fig           #fig4.6
```

fig 4.6

```
25          ((Teacher plays the word on laptop and asks S2 to listen again))
```

Shifting gaze and leaning closer to the robot, at line 13, S2 pronounces the target word similarly as in Extract 4a, more or less as "fudder" and with no observable increase in loudness. Unlike the previous word repetition, S2 now maintains their gaze on the robot while waiting for its third turn to unfold, and shifts gaze to the teacher immediately afterward (l. 15). Unlike on the first occasion (and in Extract 2), S2's conduct does not anticipate a straightforward sequence completion. S2's gaze shift to the teacher treats the robot's utterance "butter" as something that (again) blocks progression to the next task item.

The participants treat the robot's second "mishearing" as a laughable matter at lines 16–18. S1 chuckles and takes a turn that responds to the teacher's earlier DIU (l. 16) by telling in Swedish what the robot *heard* ("hörde"). The teacher orients to S1's laughter tokens with a gaze shift (fig. 4.4a) and produces what can be characterised as a tense and awkward grin (fig. 4.4b). As the import of S1's turn unfolds, the teacher's facial expression transforms into a more relaxed smile, and she confirms of S1's interpretation of the problem (l. 17). The laughter and the high pitch construct a sense of affiliation and excitement in the teacher's turn, which offers a way to continue to address the progressivity trouble as the robot's problem, not S2's. S2 is the last participant to smile (fig. 4.5) in the sequence as they seem to agree with S1 and the teacher (l. 18). In other words, here smiles and laughter are used as resources for managing interactional trouble in an affiliative manner (see also Sert & Jacknick, 2015).

When requested to utter the word for the third time by the teacher, S2 pronounces it at line 20 differently from line 13 and line 5 (Extract 4a). Instead of using the alveolar plosive */d/*, S2 now pronounces the mid-word consonant as a non-aspirated */t/*. Such a modification of an individual sound orients to a need to revise pronunciation so that the robot can better recognise it as the targeted word. Despite the modification, the robot responds in the same way as before, repeating "butter" in its third turn. While S1 and the teacher react to the robot's response by smiling, S2 thrusts their head backward (fig. 4.6) and provides a grunting sound (l. 24) that together can be seen as a display of frustration in the lack of progress of the sequence. In the end, getting the robot to ratify the target word and reaching sequence completion requires listening to the prompt once more and another repetition attempt by S2. On that occasion, S2 utters the other alternative word, "dad", which does not include the voiced dental fricative */ð/* sound that appeared as problematic in the word "father" (segment not shown here for space considerations).

Concluding discussion

In this article, we have analysed how participants resolve progressivity troubles (RQ1) in instructional, three-part repetition sequences that take place as teacher-student constellations complete pre-designed language learning tasks using a designated language learning application (Elias) coupled with an NAO robot. Focusing on the robot's missing or incongruent turns in the third position that prevent the closure of the IRE-based repetition sequence, we have also explored how participants' interactional work to restore the progressivity of the sequence can generate opportunities for teaching and learning L2 pronunciation (RQ2). Our analysis demonstrates that while the particular learning activity has been pre-designed to be accomplished as a three-part repetition sequence, its successful situated enactment in a real-life instructional setting is an interactional achievement (see also Tuncer et al., 2022). One way in which this becomes visible is through (sometimes lengthy) insert expansions to the sequence and a noticeable degree of human-human interaction between the scripted three turns of the human-robot repetition sequence. The fact that participants strive to reach sequence closure even when the robot might provide an irrelevant or incongruent contribution to the interaction also highlights the accomplished nature of the instructional sequence. As such, the findings illustrate that robot-assisted language learning activities can involve not only a range of intersubjective troubles but also interactional and pedagogical possibilities occasioned by programmed dialogue tasks and the robot's material features. In this sense, the study serves as a reminder that any particular task design can play out in considerably different ways across different participant groups (in human-human task interaction, see Hellermann & Pekarek Doehler, 2010).

The focal repetition activity is designed to enable pronunciation practice and learning, and one way in which it achieves this is through the (sometimes numerous) repetitions of the target word by students. In addition, as we have shown, the participants orient to pronunciation as one possible reason why the robot might not recognise a word or misrecognise it as something different. The ensuing pronunciation modifications (Extract 3) and instructions (Extracts 4a and b) can be seen as interactional work that identifies and takes up learning opportunities catalysed by the specific socio-material ecology of action. In addition to these kinds of momentary orientations to teaching and learning pronunciation, the extracts also illustrate how young children are learning to deal with a situation where they need to make sense of the situated meaning of AI and robotic conduct within the interpretative frame of the unfolding activity. In other words, they need to find a reason why the robot nods, what its different sounds mean and whether and how it "understands" the words used by a human participant. Such sense-making also concerns the accountability of (human) action in that identifying how one's conduct may at times need to be modified to accomplish human-robot interaction (for example by leaning closer to the robot when talking to it) involves an awareness of how one's conduct looks and sounds to other participants and AI agents (i.e., recipient design). The children in our data are thus not only engaged in learning a foreign language, but they are also learning – and being socialised into – ways of upholding intersubjectivity in human-robot interaction, in other words, "doing talking to a robot". In our setting, the teacher plays an important role in this because the students are facing the double task of figuring out how to use their L2 in oral interaction and how to interact with a robot that may not always function as it should.

In today's communicational landscape, such robot literacy skills (Peura & Johansson, 2023) are becoming increasingly needed to uphold interaction and coordinate action with various AI-powered conversational agents. When one's conversational partner is a bot, a social robot, or some other kind of conversational agent, a great deal of the maintenance of intersubjectivity is often left to the human participant(s) (see also Johansson, 2021). In our data, the particular assemblage of an NAO robot and the Elias language learning application is not for instance able to initiate repair, which means that access to a fundamental organisational resource of talk-in-interaction is asymmetrically divided between the participants. This is visible in the way the human participants treat it as their responsibility to repair troubles in the progressivity in the repetition activity by retrying (Extract 2) and modifying their utterances (Extract 3) as well as by providing accounts for the robot's conduct and abilities such as what it "hears" or does not "hear" (Extracts 4a and 4b) in an attempt to make their actions recognisable to the robot. In this way, the human participants orient to an asymmetrical distribution of participants' interactional rights and obligations in human-robot interaction; they treat the robot as less than a fully competent member.

One thing that may assist the children in our data as new users of a particular interactive technology is their familiarity with the institutional nature of classroom-based instruction and its routine interactional practices. The pre-scripted word repetition sequence whereby the robot first utters a target word, the student repeats it, after which the robot is supposed to indicate a student's "correct" repetition and close the sequence with a third-turn word repetition and audio-visual signals (beep and the occasionally flashing eyes) resembles the interactional roles and division of labour between the teacher and students in typical IRE/F sequences (Lyle, 2008; Veivo & Mutta, 2022). At the same time, our analysis has shown that the three-part sequence is locally adapted to features of the technological setting through practices of sequence facilitation. In a different technological context, Gan et al. (2023) have recently described how copresent adults mediate the participation of children in video calls by encouraging and supporting them to produce sequentially relevant actions. Such a phenomenon of sequence facilitation also takes place in our data, even if the repetition sequence is scripted as a two-party child-robot interaction and the teacher is an unaddressed recipient by the robot. Prompting students to produce word repetitions (Extract 2), helping them pronounce the word so that the robot would recognise them (Extract 4b) and explaining the robot's conduct (Extract 4a) are some of the practices through which the teacher facilitates the completion of the three-part sequence. Moreover, the students themselves expand the participation framework beyond the pre-programmed two-party interaction by recruiting the teacher's assistance via gaze shifts (see also Veivo & Mutta, 2022) and questions that initiate insert sequences within the three-part sequence. For these reasons, the word repetitions can be seen as collectively and cooperatively produced. In contrast to the participation of students as a collective cohort in plenary classroom interaction (Payne & Hustler, 1980), in the setting we have described, the teacher is working with the students to sustain their interaction with an (instructional) robot.

We pointed out in the introduction that pedagogical studies in robot-assisted language learning (RALL) suggest that getting social robots to induce instructional gains requires interactional adjustment from teachers and students (see Randall, 2019; van den Berghe et al., 2019). From this perspective, the observations that the repetition sequences require considerable sequence facilitation and involve an orientation to the robot as a less than fully competent member imply that social robots may be more feasibly used as a tool supporting teaching rather than a fully-fledged agentic participant such as the teacher. In our setting, the teacher is much needed and plays an important role in the activity in facilitating the students' participation. It is thus useful to remember that classrooms are complex ecologies of action and instruction (Guerrettaz & Johnston, 2013; van Lier, 2004), and bringing a technological device into the classroom reorganises the ecology with possibly unpredictable changes to interactional patterns. As robots are still a relatively new phenomenon, we know considerably less about how interaction is organised in RALL classrooms than we do about classrooms that involve more "traditional" educational and material resources. The current study has shed some light on this matter, but, given that our study is limited to a particular context and technology (which develops rapidly), more microanalytical research of HRI and RALL is clearly needed. For pedagogical studies, a CA approach can provide useful insights into how participants resolve technical and interactional troubles (Honig & Oron-Gilad, 2018; Maijala & Mutta, in press; Veivo & Mutta, 2022) and the situated competencies that manifest themselves in actual occasions of human-robot interaction. For contexts beyond education, a close look at HRI can thus shed light on questions such as if, when, and how social robots are treated by people as a participant with situated agency, a technological resource for human action, or something in-between. It is interesting that, at least in our data, numerous technical glitches and bugs with very tangible communicative consequences such as extra beeps or non-relevant words are fairly common. Yet, sequences of action can still be successfully completed with the help of interactional work and adjustment by the teacher and students. That itself illustrates the adaptiveness of human interactional competence and the pervasiveness of human sense-making.

Acknowledgements

We are grateful to the anonymous reviewers, the special issue editors and the Prologi editorial board for their constructive suggestions for improving earlier versions of this article. Any remaining argumentative errors and shortcomings are our own. This work has been supported by the Academy of Finland [grant number 343480].

References

Dautenhahn, K. (2007). Socially intelligent robots: Dimensions of human–robot interaction. *Philosophical Transactions of the Royal Society*, *362*, 679–704. https://doi.org/10.1098/rstb.2006.2004

Dingemanse, M., Blythe, J., & Dirksmeyer, S. T. (2014). Formats for other-initiation of repair across languages. An exercise in pragmatic typology. *Studies in language, 38*(1), 5–43. https://doi.org/10.1075/sl.38.1.01din

Fischer, J. E., Reeves, S., Porcheron, M., & Sikveland, R. O. (2019). Progressivity for voice interface design. *Proceedings of the 1st International Conference on Conversational User Interfaces*, 1–8. https://doi.org/10.1145/3342775.3342788

Gan, Y., Greiffenhagen, C., & Kendrick, K. H. (2023). Sequence facilitation: Grandparents engineering parent–child interactions in video calls. *Research on Language and Social Interaction, 56*(1), 65–88. https://doi.org/10.1080/08351813.2023.2170640

Gehle, R., Pitsch, K., Dankert, T., & Wrede, S. (2015). Trouble-based group dynamics in real-world HRI—reactions on unexpected next moves of a museum guide robot. In 2015 *24th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)* (pp. 407-412). IEEE. https://doi.org/10.1109/ROMAN.2015.7333574

Guerrettaz, A. M., & Johnston, B. (2013). Materials in the classroom ecology. *The Modern Language Journal, 97*(3), 779–796. https://doi.org/10.1111/j.1540-4781.2013.12027.x

Guzman, A. L., & Lewis, S. C. (2020). Artificial intelligence and communication: A Human-machine Communication research agenda. *New Media & Society, 22*(1), 70–86. https://doi.org/10.1177/1461444819858691

Hellermann, J. (2003). The interactive work of prosody in the IRF exchange: Teacher repetition in feedback moves. *Language in society, 32*(1), 79–104. https://doi.org/10.1017/s0047404503321049

Hellermann, J., & Pekarek Doehler, S. (2010). On the contingent nature of language-learning tasks. *Classroom Discourse, 1*(1), 25–45. https://doi.org/10.1080/19463011003750657

Hepp, A. (2020). Artificial companions, social bots and work bots: Communicative robots as research objects of media and communication studies. *Media, Culture & Society, 42*(7–8), 1410–1426. https://doi.org/10.1177/0163443720916412

Hofstetter, E., Keevallik, L., & Löfgren, A. (2021). Suspending syntax: Bodily strain and progressivity in talk. *Frontiers in Communication*, 6. https://doi.org/10.3389/fcomm.2021.663307

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. *Proceedings of the Conference Human Perspectives on Spoken Human-Machine Interaction*, 46–51. https://doi.org/10.6094/UNIFR/223816

Jefferson, G. (2004). Glossary of transcript symbols with an introduction. In G. Lerner (ed.), *Conversation Analysis: Studies from the First Generation*, 13–31. John Benjamins. https://doi.org/10.1075/pbns.125.02jef

Johansson, M. (2021). Talking with a chatbot: Simulated understanding of human-chatbot communication? In M. Johansson, S.-K. Tanskanen, & J. Chovanec (Eds.), *Analyzing Digital Discourses. Between Convergence and Controversy* (pp. 105–131). Palgrave MacMillan, Cham. https://doi.org/10.1007/978-3-030-84602-2_5

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&2_4

Kitzinger, C. (2012). Repair. In J. Sidnell & T. Stivers (Eds.), *The Handbook of Conversation Analysis* (pp. 229–256). John Wiley & Sons.

Koshik, I. (2002). Designedly incomplete utterances: A pedagogical practice for eliciting knowledge displays in error correction sequences. *Research on Language and Social Interaction, 35*(3), 277–309. https://doi.org/10.1207/S15327973RLSI3503_2

Laaksonen, S.-M., Laitinen, K., Koivula, M., & Sihvonen, T. (2020). Puhekaverina botti: Viestivä tekoäly inhimillistettynä vuorovaikutuskumppanina. *Lähikuva: Audiovisuaalisen kulttuurin tieteellinen julkaisu, 33*(1), 63–78. https://doi.org/10.23994/lk.91435

Licoppe, C., & Rollet, N. (2020). «Je dois y aller». Analyses de séquences de clôtures entre humains et robot. *Réseaux,* (2), 151–193. https://doi.org/10.3917/res.220.0151

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

Maijala, M., & Mutta, M. (in press). Teachers' role in robot-assisted language learning: the impact on the classroom ecology. *The EuroCALL Review*.

Margutti, P., & Drew, P. (2014). Positive evaluation of student answers in classroom instruction. *Language and Education, 28*(5), 436–458. https://doi.org/10.1080/09500782.2014.898650

Matsui, T., Tani, I., Sasai, K., & Gunji, Y.-P. (2021). Effect of Hidden Vector on the Speech of PRVA. *Frontiers in Psychology, 12*, 627148–627148. https://doi.org/10.3389/fpsyg.2021.627148

Mehan, H. (1979). *Learning lessons: social organization in the classroom*. Harvard University Press. https://doi.org/10.4159/harvard.9780674420106

Mondada, L. (2019). Conventions for multimodal transcription. https://www.lorenzamondada.net/_files/ugd/ba0dbb_986ddd4993a04a57acf20ea06e-2b9a34.pdf

Park, Y. (2014). The roles of third-turn repeats in two L2 classroom interactional contexts. *Applied linguistics, 35*(2), 145–167. https://doi.org/10.1093/applin/amt006

Payne, G., & Hustler, D. (1980). Teaching the class: The practical management of a cohort. *British Journal of Sociology of Education, 1*(1), 49–66. https://doi.org/10.1080/0142569800010104

Pelikan, H. (2023). *Robot sound in interaction: Analyzing and designing sound for human-robot coordination*. Linköping studies in Arts and Sciences 853. Linköping University. https://doi.org/10.3384/9789180751179

Pelikan, H. R. M., & Broth, M. (2016). Why that Nao? How humans adapt to a conventional humanoid robot in taking turns-at-talk. *34TH Annual CHI Conference on Human Factors in Computing systems, CHI 2016*, 4921–4932. https://doi.org/10.1145/2858036.2858478

Peura, L., & Johansson, M. (2023). A friend or a machine? Study on the child–robot relationship in a foreign language class of young learners. In R. Hakli, P. Mäkelä, & J. Seibt (Eds.), *Social Robots in Social Institutions* (pp. 165–173). IOS Press.

Pitsch, K. (2020). Répondre aux questions d'un robot. Dynamique de participation des groupes adultes-enfants dans les rencontres avec un robot guide de musée. *Réseaux, 2*(21), 113–150. https://doi.org/10.3917/res.220.0113

Randall, N. (2019). A Survey of Robot-Assisted Language Learning (RALL). *ACM Transactions on Human-Robot Interaction, 9*(1), Article 7, 1–36. https://doi.org/10.1145/3345506

Rollet, N., & Clavel, C. (2020). "Talk to you later": Doing social robotics with conversation analysis. Towards the development of an automatic system for the prediction of disengagement. *Interaction Studies, 21*(2), 268–292. https://doi.org/10.1075/is.19001.roll

Schegloff, E. A., Jefferson, G., & Sacks, H. (1977). The preference for self-correction in the organization of repair in conversation. *Language, 53*(2), 361–382. https://doi.org/10.2307/413107

Schegloff, E. A. (2007). *Sequence organization in interaction: A primer in conversation analysis*. Cambridge: CUP. https://doi.org/10.1017/CBO9780511791208

Seedhouse, P. (1997). The case of the missing "No": The relationship between pedagogy and interaction. *Language Learning, 47*(3), 547–583. https://doi.org/10.1111/0023-8333.00019

Sert, O., & Jacknick, C. (2015). Student smiles and the negotiation of epistemics in L2 classrooms. *Journal of Pragmatics, 77*, 97–112. https://doi.org/10.1016/j.pragma.2015.01.001

Skantze, G. (2021). Turn-taking in conversational systems and human-robot interaction: A review. *Computer Speech & Language, 67*, 101178. https://doi.org/10.1016/j.csl.2020.101178

Stivers, T., & Robinson, J. D. (2006). A preference for progressivity in interaction. *Language in Society 35*(3), 367–392. https://doi.org/10.1017/S0047404506060179

Stommel, W., de Rijk, L., & Boumans, R. (2022). "Pepper, what do you mean?" Miscommunication and repair in robot-led survey interaction. In *2022 31st IEEE International Conference on Robot and Human Interactive Communication (RO-MAN),* pp. 385–392. IEEE. https://doi.org/10.1109/RO-MAN53752.2022.9900528

Tuncer, S., Gillet, S., & Leite, I. (2022). Robot-mediated inclusive processes in groups of children: From gaze aversion to mutual smiling gaze. *Frontiers in Robotics and AI, 9*, 729146–729146. https://doi.org/10.3389/frobt.2022.729146

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

van Lier, L. (2004). *The Ecology and Semiotics of Language Learning: a Sociocultural Perspective*. Kluwer Academic Publishers. https://doi.org/10.1007/1-4020-7912-5

Veivo, O., & Mutta, M. (2022). Communication breakdowns in child robot interaction in RALL. *Computer-Assisted Language Learning*. https://doi.org/10.1080/09588221.2022.2158203

OTSIKKO JA ASIASANAT SUOMEKSI:

"Sanonks mä sen väärin?": Toiminnan progressiivisuuden ongelmat ja ohjauksen mahdollisuudet lasten ja robotin välisessä kakkoskielisessä vuorovaikutuksessa

ASIASANAT: kielen oppiminen vuorovaikutuksessa, lapsen ja robotin välinen vuorovaikutus, Nao-robotti, progressiivisuus, sosiaaliset robotit