Routledge
Taylor & Francis Group

🔓 OPEN ACCESS | Check for updates

Interaction and collaboration in robot-assisted language learning for adults

Olov Engwall[a] and José Lopes[b]

[a]Department of Speech, Music and Hearing, KTH Royal Institute of Technology, Stockholm, Sweden; [b]Interaction Lab, Heriot-Watt University, Edinburgh, UK

ABSTRACT

This article analyses how robot–learner interaction in robot-assisted language learning (RALL) is influenced by the inter-action behaviour of the robot. Since the robot behaviour is to a large extent determined by the combination of teaching strategy, robot role and robot type, previous studies in RALL are first summarised with respect to which combinations that have been chosen, the rationale behind the choice and the effects on interaction and learning. The goal of the summary is to determine a suitable pedagogical set-up for RALL with adult learners, since previous RALL studies have almost exclusively been performed with children and youths. A user study in which 33 adult second language learners practice Swedish in three-party conversations with an anthropomorphic robot head is then presented. It is demonstrated how different robot interaction behaviours influence interaction between the robot and the learners and between the two learners. Through an analysis of learner interaction, collaboration and learner ratings for the different robot behaviours, it is observed that the learners were most positive towards the robot behaviour that focused on interviewing one learner at the time (highest average ratings), but that they were the most active in sessions when the robot encouraged learner–learner interaction. Moreover, the preferences and activity differed between learner pairs, depending on, e.g., their proficiency level and how well they knew the peer. It is therefore concluded that the robot behaviour needs to adapt to such factors. In addition, collaboration with the peer played an important part in conversation practice sessions to deal with linguistic difficulties or communication problems with the robot.

CONTACT O. Engwall ✉ engwall@kth.se 🏛 KTH Royal Institute of Technology, Stockholm, Sweden

Introduction

Robot-assisted language learning (RALL) was proposed as early as 1986, when Harwin, Ginige, and Jackson (1986) argued for the advantage of using robots compared to software-based computer-assisted language learning (CALL), since robots allow for physical interaction. Later, van den Berghe, Verhagen, Oudgenoeg-Paz, van der Ven, & Leseman (2019) concurred that interaction in and with the learner's physical environment is particularly important for language learning, and that this is a key feature of RALL. Aidinlou, Alemi, Farjami, and Makhdoumi (2014) advocated other advantages of RALL, such as repeatability, flexibility, humanoid appearance and expression of emotions. The argument that robots may improve learning of a second language (L2) has been corroborated by e.g., Han, Jo, Jones, and Jo (2008) and Wedenborn, Wik, Engwall, and Beskow (2016). Han et al. investigated the educational benefits of using a robot compared to other means of presentation. Korean children with similar language abilities were divided into three groups to be taught the same material using, respectively, book and audio tape; a personal computer with web-based instruction; and the semi-humanoid robot IROBI, equipped with speech recognition and synthesis. The group that interacted with the robot was more engaged in learning and performed significantly better on English language problems in the post-test. The study by Wedenborn et al. (2016) similarly demonstrated a higher level of Russian vocabulary retention in post-tests for adult Swedish learners after practicing with the robot head Furhat (which is also used in the present study). The robot group remembered significantly more words than the group practicing with a computer-animated screen-based face, and the one practicing with an impersonal interface that only presented the words audio-visually.

The above studies hence suggest that robots may contribute to more effective language learning compared with other practice methods. Han (2012) and Aidinlou et al. (2014) summarised early RALL, studies as well as general theoretical and practical considerations for RALL. In two recent extensive surveys of earlier RALL studies, van den Berghe et al. (2019) and Randall (2020) made extensive reviews of, respectively, 33 and 79 studies, discussing the methodology, results and limitations of using robots for first and second language learning of vocabulary, reading and writing skills, grammar and sign language. van den Berghe et al. (2019) pointed out that since learners tend to anthropomorphise robots based on their humanoid or animal-like appearance and verbal and non-verbal behaviour, the robots can be given specific roles as, for example, teacher or peer, and interact as such when the learning material is presented. However, the type of robot used influences the role that it may

be given and how the learning material should be presented (i.e., the teaching strategy), and this is further related to the learners' age. Some previous studies have considered different types of educational robot hardware and software (Han, 2012; Mubin, Shahid, & Bartneck, 2013a; Randall, 2020), teaching strategies and underlying pedagogical theories (Mubin, Stevens, Shahid, Al Mahmud, & Dong, 2013b; Wu, Wang, & Chen, 2015), the role of the robot in the practice (Aidinlou et al., 2014; Han, 2012; Mubin et al., 2013b) and learner age groups (van den Berghe et al., 2019), but this article expands on the question how different robot types should be combined with different teaching strategies and robot roles, in general, and for adults in particular. Belpaeme et al. (2018) described similar considerations for preschool children, but since there are few previous RALL studies with adult learners, the goal of the literature survey is to answer the question: *What conclusions for RALL for adults can be drawn from how robot type, teaching strategy and robot role have been combined for different learner age groups in previous studies?*

Based on the summary of how robot types, robot roles and teaching strategies have been used in previous studies, a collaborative robot-assisted language learning practice set-up with adults is presented. In this set-up, pairs of adult L2 learners practice social conversation led by a robot. The pedagogical and technological benefits and limitations of such a setting are then explored in a user study, with the explicit research question: *How does the robot's interaction behaviour in conversational practice with two adult language learners influence the learners' interaction, collaboration and perception of the robot?*

Literature survey: Combinations of robot types, teaching strategies and robot roles in RALL

RALL interaction can to large extent be described by the teaching strategy and the role of the robot, which are both influenced by the type of the robot (regarding its hardware and software features). These factors are summarised in Figure 1, which shows the appearance of the robots in previous RALL studies together with their roles, teaching strategy, learner age groups, target language and duration of the study. Table 1 defines different teaching strategies (following, but expanding from, e.g., Wu et al., 2015) and relates them to different robot roles. The rationale for these different robot roles is briefly summarised in the heading of Table 1.

Wu et al. (2015) stressed the importance of first selecting the pedagogical strategy and only then the robot technology. However, unless one creates an in-house robot as Wu *et al.* did, one is restricted by the

	low		Anthropomorphism		higher
	Toy-like	Face or belly screens		Humanoid	Head
low	**Mindstorm** (Mubin, 2013) Tool TPR, CLL, TBLT 10-12 years English-ROILA 2 hours	**IROBI** (Han, 2005) Tutor MBE, ALM 12-13 years English 40 minutes	**Robosem** (Park, 2011) Assistant MBE, say-after-me 10-13 years Korean-English 4 weeks x. 2 hours/week	**Robosapien** (You, 2006) Assistant Cheering, TPR, storytelling 10-11 years Mandarin-English 2 weeks x 2 times x 40 minutes	**Mero** (Lee, 2011) Tutor ALM, pronunciation feedback 9-13 years Korean-English 8 weeks x 2 days x 2 hours
Expressiveness	**Tega** (Gordon, 2016) Peer CLL game 4 years English-Spanish 3-7 sessions	**EngKey** (Lee, 2011) Tutor Role-play 9-13 years Korean-English 8 weeks x 2 days x 2 h	**Robovie** (Kanda, 2004) Companion CLT 10-11 years Japanese-English 2 weeks, 31 days during 9 weeks	**Mec Willy** (Mazzoni, 2015) Peer CLT, CLL 4-6 years Italian-English 45 minutes	**Furhat** (Wedenborn, 2016; Lopes, 2017) Tutor, companion ALM, CLT, CLL Adults Varied-{Russian, Swedish} 1–4 times x 15 minutes
higher	**iCat** (Mubin, 2013) Partner CLL game 10-12 years English-ROILA 30 minutes	**PET** (Wu, 2015) Tutor TPR, CLT, ALM 10-11 years Mandarin-English 4 times x 50 minutes	**Nao** (Kennedy, 2016; Tanaka, 2012; Alemi, 2015; Balkibekov, 2016; Khalifa, 2017; Schodde, 2017) Tutor, assistant, learner, opponent TPR, CLL, ALM 3-6, 9-10, 9-10, 12-13, 18-24, μ=24 years English-French, {Japanese, Persian, Kazakh, Japanese}-English, German-Vimmi 6 days, 5 weeks x 2 x 1 hour, 10 min, short lesson, 10 min, 15 min		

Notes: Robots are grouped by similarity in appearance sorted in terms of human-likeness (from left to right) and expressiveness (top to bottom).They are represented by simplified drawings, their role, interaction strategies, learner age, L1-L2 language combination and interaction duration in previous studies (references abbreviated to first author).

Figure 1. Robots in previous RALL studies.

available robots and needs to find a suitable combination of teaching strategy, robot role, robot type and learner age. Many combinations of robot types, teaching strategies and robot roles have already been tested with middle-school students (Alemi, Meghdari, & Ghazisaedy, 2015; Balkibekov, Meiirbekov, Tazhigaliyeva, & Sandygulova, 2016; Han et al., 2008; Kanda, Hirano, Eaton, & Ishiguro, 2004; Lee et al., 2011; Park, Han, Kang, & Shin, 2011; Tanaka & Matsuzoe, 2012; et al., 2013; You et al., 2006), and some in addition with preschool children (Gordon et al., 2016; Kennedy, Baxter, Senft, & Belpaeme, 2016; Mazzoni & Benvenuti, 2015), demonstrating the potentials of RALL for these age groups. RALL studies with adult learners are on the other hand still scarce – the authors are only aware of the work by Khalifa, Kato, and Yamamoto (2017) and Schodde, Bergmann, and Kopp (2017), in addition to the own studies preceding this one (Lopes, Engwall, & Skantze, 2017; Wedenborn et al., 2016). There may be large differences between children and adults regarding learner preferences, since adult learners demand realistic interaction building on their previous experience and learning that is directly relevant to them, according to andragogy principles. How robot type, teaching strategy and robot role should be chosen and combined for adult learners hence requires additional attention. The goal of the summary of previous studies, which were mostly performed with children, is therefore to determine how these previous studies may guide the setup of RALL for adults. The following sections summarise the *Robot type* categories introduced in Figure 1, the *Teaching strategies*

Table 1. Overview of teaching strategies and robot roles used in previous RALL studies.

Teaching strategy	Robot roles				
	Teacher assistant	Tutor	Peer, partner, tool, opponent	Learner	Social companion
	Enhance classroom lessons (motivation, illustration) led by a human teacher	*Robot-led lessons to teach or practice a curriculum*	*Learners practice with equally proficient robot*	*Robot peer tutored*	*Communication and socialising*
Practice on specific learning material *Multimedia-based education (MBE): pictures and words presented on screen. Audio-Lingual Method (ALM): learners mimic robot utterances (Richard & Rodgers, 2014).*	• MBE with IROBI (Han, 2005) or Robosem (Park et al., 2011) • Cheering by Robosapien (You et al., 2006) or Nao (Alemi, 2015) • Pronunciation model (Robosapien in You, 2006; Robosem in Park et al., 2011) • Storytelling (Robosapien in You, 2006; Nao in Alemi, 2014)	• MBE vocabulary or grammar learning with IROBI (Han, 2005), Nao (Kennedy, 2016) or Furhat (Wedenborn et al, 2016) • Pronunciation practice with Meroo (Lee et al, 2011), Storytelling by PET (Wu et al., 2015)	Competitive game on vocabulary learning against Nao (Balkibekov et al., 2016)		
Physical interaction *Enhance interaction by robot body gestures connected to verbal instructions, e.g., Total Physical Response (TPR) (Asher, 1969).*	Robosapien demonstrating body gestures (raising its hand etc) when asked by the students (You, 2006)	Learning body names with PET (Wu et al., 2015)	Controlling Mindstorm robot (Mubin et al., 2013a)	Body gesture response by Nao (Tanaka & Matsuzoe, 2012)	Socialising physical interaction (hugging etc) with Robovie (Kanda et al., 2004)
Communication *Encourage information transfer, fluency and confidence, rather than form, based on Communicative Language Teaching (CLT) (Spada, 2007).*	Asking and replying to questions (Robosapien in You, 2006; Nao in Alemi, 2014)	Conversation with PET (Wu et al., 2015)	Nao deliberately making mistakes for new material (Alemi et al., 2015)		Conversation with Robovie (Kanda et al., 2004) Social relation with Nao (Kennedy et al., 2016)

Table 1. Continued.

			Robot roles		
	Teacher assistant	Tutor	Peer, partner, tool, opponent	Learner	Social companion
Role play *The robot presents a scenario practicing meaningful real-life verbal skills, using the task-based language teaching (TBLT) approach (Ellis, 2003).*		Role-play with EngKey as salesperson and learners as customers (Lee et al., 2011)	Learning vocabulary by helping an avatar together with Tega (Gordon et al., 2016)		
Collaborative language learning (CLL) *Robot or peer support in the practice (Nunan, 1992).*	Collaborations on tasks with Nao (Alemi et al., 2014)	Interactive alignment with Nao (Khalifa et al., 2017)	Collaboration on tasks with iCat (Mubin et al., 2013a) or Tega (Gordon et al., 2016) or vocabulary with Mec Willy (Mazzoni & Benvenuti, 2015).		Social L2 interaction with Furhat (Lopes et al., 2017; present study)

Legend: Examples of teaching strategy and role combinations employed in previous RALL studies (references abbreviated to first author), broadly ordered by increasing freedom in the interaction with the robot (from left to right) and complexity of task (from top to bottom).

listed in the left-most column of Table 1 and the *Robot roles* listed in the first row of Table 1.

Robot types

Based on common features of the different robots used in previous RALL studies, they may be grouped as:

Toy-like robots (*Lego Mindstorm, Tega, iCat*) have appearances and behaviours that are familiar to children and thus non-threatening. This may benefit engagement, reduce anxiety and facilitate simplified interaction, especially with younger learners. On the other hand, they lack human likeness in visual interaction signals and expected behaviour, which may affect what roles and teaching strategies are appropriate.

Face or belly screen robots (*PET, IROBI, EngKey, Robosem*) extend multimedia-based practice on screen by establishing a situated interaction, with limited face signals and body gestures, in addition to the practice material presented on screen. The primary learning interaction is in most cases with the screen, but as shown by Han et al. (2008), the robot may nevertheless contribute to more effective learning than screen-only practice.

Humanoid robots (*Robosapien, Robovie, Mec Willy, Nao*) take additional advantage of the robot's physical body to incorporate more elaborate arm and leg movements in the practice (c.f. Physical interaction below) and/or to build on humanlike gestures for the interaction. The anthropomorphic effect is strong for the humanoid robots, which could in part explain why *Nao* stands out in Figure 1, having been used with different teaching strategies in several different roles and for a larger range of target learner ages, whereas other robots have primarily been used in one combination.

Robotic heads (*Mero, Furhat*) instead focus on the importance of facial signals in communication (e.g., lip, eye and eyebrow movements to signal attention and emotions) and language learning (e.g., lip movements for pronunciation training). The *Furhat* robot, which is used in the present study, is uniquely anthropomorphic and expressive in comparison with other robots and has previously acted as e.g., host in competitive or collaborative quiz games (Skantze, Johansson, & Beskow, 2015), companion to autism children, detector of early signs of dementia (Jonell, Mendelson, & Storskog, 2017), or simulated Alzheimer patient (Kanov, 2017).

As the features of the robot will influence the learner–robot interaction, the first key question is if all robot types are suitable for all roles, teaching strategies and learner ages. The second is if human–human interaction strategies should be used for the robot, or if a different

strategy, which instead builds on features of the robot, should be used. As noted above, adult learners may have higher demands on the interaction being transferable to real-life settings, and this imposes higher requirements on the robot regarding realism in appearance and behaviour. It can be noted that the robots that have this far been used in RALL studies with adults, *Nao* and *Furhat*, are the ones that have the highest degree of combined anthropomorphism and expressiveness, as illustrated in Figure 1. The following section focuses on how previous RALL studies have taken advantage of the hardware and software features of the different robot types by employing them with different teaching strategies.

Teaching strategy

This section describes and exemplifies teaching strategies that may be used with robots and discusses their combination with different robot types, whereas the next section considers the interaction with different robot roles. As all studies use both a teaching strategy and a robot role, different aspects of previous studies will inevitably be covered in either of the two sections, depending on if they are primarily related to teaching strategy or robot role. Brief overviews of the individual studies are instead provided by Figure 1 and Table 1.

Practice on specific learning material is often based on multimedia-based education (MBE) or the audio-lingual method (ALM). MBE is readily combined with face- or belly–tablet robots, such as *IROBI* (Han et al., 2008; as described in the Introduction) or *Robosem*, which Park et al. (2011) employed in exercises where the robot could, e.g., recognise which letter a student wrote, display study material on screen and pronounce words presented on cards. Alternatively, another robot type may be combined with a separate tablet screen, as illustrated by the vocabulary learning studies employing *Nao* (Schodde et al., 2017) or *Furhat* (Wedenborn et al., 2016; as described in the Introduction). Robots can also be combined with ALM for listening or pronunciation practice, as illustrated by You, Shen, Chang, Liu, and Chen (2006) and Lee et al. (2011), who respectively used *Robosapien* and *Mero*. For such practice it is, however, important that the robot software (speech synthesis) and hardware (e.g., lip movements) support learning, as illustrated by the fact that the post-test in Lee et al. (2011) showed that the practice led to an important improvement in speaking skills, but not in listening skills, which the authors attributed to the robot's poor text-to-speech synthesis. Another use of robots in focused practice is to increase engagement, as illustrated by the studies by Balkibekov et al. (2016), Kennedy et al. (2016) and

Alemi, Meghdari, and Ghazisaedy (2014), which all used *Nao* robots in the practice of, respectively, vocabulary, grammar and listening skills. Increased engagement is the result of learners forming a relationship with the robot, often based on anthropomorphising it, which signifies that the robot appearance and verbal as well as non-verbal behaviour (e.g., body movements, facial and emotional signals) need to be suitable for the target learner age group. Practice on specific learning material often utilises the robot in the role of teaching assistant or tutor to present the material.

Physical interaction, based on total physical response (TPR), is best combined with humanoid, full-body robots, which can be used to demonstrate body gestures to students (*Robosapien* in You et al., 2006; *PET* in Wu et al., 2015), or which the students can verbally control (*LEGO Mindstorm* in Mubin et al., 2013a) or instruct (*Nao* in Tanaka & Matsuzoe, 2012). The paradigm used by Tanaka and Matsuzoe (2012) was that a teacher first taught the child the meaning of a verb. In the second part of the lesson, first the teacher and then the child showed the robot how to enact the verb. The post-tests indicated that the children had learned more verbs with the robot than with conventional teaching and that retention was better, thus illustrating the benefit of tutoring someone else, even if that is a robot. The main consideration for physical interaction with robots is the degrees of freedom in the body movements, as more degrees of freedom allow to illustrate or enact a larger variety of and more complex body gestures (with *Nao* currently being the most advanced robot in this respect).

Communication practice, based on communicative language teaching (CLT), can be combined with several different robot types, as the combination depends more on how structured the interaction is. Interaction in previous studies ranges from the robot asking and replying to questions (*Robosapien* in You et al., 2006; *Nao* in Alemi et al., 2014), over structured conversation practice (*PET* in Wu et al., 2015), to freer conversation (*Robovie* in Kanda et al., 2004; *Mec Willy* in Mazzoni & Benvenuti, 2015; *Furhat* in Lopes et al., 2017). It can be noted that the studies with freer conversation employ robots that can e.g., detect learner identities (Kanda et al., 2004) or emotions (Mazzoni & Benvenuti, 2015) and express emotions through facial signals (Lopes et al., 2017; Mazzoni & Benvenuti, 2015). This can contribute to establishing social relationships in the more realistic communication practice. Kanda et al. (2004) used RFID tags to allow *Robovie* to recognise the elementary school students that it interacted with and the robot further expressed social behaviours such as greeting and hugging, in addition to producing simple every-day English phrases and recognising 50 different English words. It was found that the popularity of the robot faltered after one week,

firstly because the learners had expected that it should be able to engage in more advanced interaction and secondly because the learners' linguistic level came to surpass that of the interaction with the robot. In a 2-month follow-up study, Kanda, Sato, Saiwaki, and Ishiguro (2007); therefore, introduced improved social interaction strategies. In addition to calling learners by name, the robot adapted its behaviour to each learner and became more personal over time. These additions led to better maintained motivation, which learners attributed to their interest in making friends with the robot, illustrating the importance of forming personal relationships in longer learning interactions, as further discussed when describing the Social companion role below.

Role play is used with task-based language teaching (TBLT) and builds upon establishing a social relationship between the robot and the learner in the scenarios. This means that robots able to display and/or detect non-verbal social signals, such as facial and bodily displays of emotion, are the most suitable for this type of interaction. One example is the shop scenario created by Lee et al. (2011), which let students act as customers to *EngKey,* who was a shop keeper that displayed expressive facial signals for, e.g., pleasure, dislike, surprise, embarrassment and pride, and face and body motions such as winking, cheering and sulking. Another example is the personalised version of *Tega* who collaborated as a peer with children to help an on-screen animated toucan during a trip in Spain, using Spanish vocabulary (Gordon et al., 2016). Personalisation was achieved by including software for affective analysis of the learners' facial expressions and a cognitive model that defined the verbal and non-verbal behaviour of the robot based on affective analysis and educational content. Gordon *et al.* found that the children learned new Spanish words regardless of if *Tega* was personalised or not, but that the affective personalisation made the learners' emotions towards the robot more positive in the long-term perspective.

Collaborative language learning (CLL) builds on collaboration between learners (Mazzoni & Benvenuti, 2015; Mubin et al., 2013a) or between robot and a learner (Khalifa et al., 2017; Mazzoni & Benvenuti, 2015). Collaboration between learners imposes less constraints on the robot type, but for collaboration between learner and robot, expressiveness and human-likeness are important. Mubin et al. (2013a) studied collaboration between 2 and 3 human learners controlling a *LEGO Mindstorm* robot using voice commands in the artificial language ROILA (robot interaction language). The task was to get the robot to a specific position in the room where it should hit a target with balls. The effect of human learner collaboration was not investigated separately, but the fact that the collaborative scenario was better appreciated

by the learners than the corresponding vocabulary game when learners interacted individually with *iCat* suggests that collaboration with peers may improve practice. Mazzoni and Benvenuti (2015) investigated collaboration between either two children or between one child and the humanoid robot *Mec Willy*. The robot used the socio-cognitive conflict paradigm, in which the learners become aware of differences in their respective points of view and restructure their perception of the situation through discussion. Rather than giving the correct answer when the children matched English words to pictures, the robot initiated discussions about the solutions (e.g., '*Your suggestion is interesting... but are we sure that it is correct? Could there be an alternative or do we think that this is the correct answer?*'), hence encouraging learners to reason about their ideas. Using pre- and post-tests, the setting with the robot was shown to result in more words learned than that with another child (mean of 3 new word–picture associations learned during a 12-min session, compared to 1.5 for the child–child interaction). Khalifa et al. (2017) studied interactive alignment of one human learner (Japanese university students) interacting with two *Nao* robots. The set-up was that one of the robots alternately asked the other robot and the learner questions, and the replying robot answered as an advanced level learner. The amount of interactive alignment, i.e., how much the human learner used similar new English expressions as the answering robot when asked a similar question, depended on the learner's level. More advanced learners were more prone to learn unfamiliar syntactic structures from the robot, but learners at all levels improved their proficiency in English.

Having covered the teaching strategies used with robots, the next section summarises the interaction that different robot roles give rise to and outcomes of this interaction.

Robot roles

As shown in Table 1, five different groups of robot roles may be identified. These are presented below in order of increasing learner initiative.

Teaching assistant robots are used together with a human teacher as a motivational enhancement by introducing new elements in traditional classroom teaching. This role is exemplified by *Robosapien, Robosem* and *Nao* in, respectively, You et al. (2006), Park et al. (2011) and Alemi et al. (2014, 2015). As shown in Table 1, the robot had similar interaction patterns, such as introducing exercises, telling stories, asking the learners questions or answering theirs. The robot in Alemi et al. (2015) in addition deliberately made mistakes and hence acted as peer in combination

with the assistant role. Through questionnaires, the three studies all showed that the learners were more motivated for learning English (either compared to before the study or to a group learning without robots) and Alemi et al. (2015) further demonstrated that the learner groups that had been practicing with the robot felt less anxious about learning English than the non-RALL group. Since the robot as teaching assistant primarily employs that motivation may increase because interaction with the robot is different from practice with the human teacher it may be an advantage that the robot's appearance and interaction behaviour are robot- or toy-like.

Tutor robots interact with learners without a human teacher present. Examples include *Mero* scoring and giving feedback on the learners' pronunciation (Lee S, 2011), *PET* teaching body part vocabulary, telling the students about itself and practicing conversation (Wu et al., 2015), Nao teaching adult German learners the artificial language Vimmi using either random or adaptive difficulty levels (Schodde et al., 2017), and, as already described in the introduction, *IROBI* and *Furhat* respectively teaching English and Russian vocabulary (Han, 2012: Wedenborn et al., 2016). User surveys showed increased motivation for language learning compared to before practicing with robots (Lee et al., 2011) or to a non-RALL control group (Wu et al., 2015). Post-tests showed learning of vocabulary (Schodde et al., 2017; Wedenborn et al., 2016; Wu et al., 2015) and speaking skills (Lee et al., 2011; Wu et al., 2015). In the tutor role, the robot must convince learners that it itself has a good mastery of the practice material. This introduces – learner age dependent – requirements on credibility in appearance and interaction: The least expressive and anthropomorphic robots *PET* and *IROBI* were shown to be appropriate for middle-school learners, and the most, *Furhat*, for adult learners. *Nao* has been used as tutor with both children (Kennedy et al., 2016) and adults (Schodde et al., 2017). The study by Schodde et al. (2017) was performed with adults, but their assumption was that the same set-up and adaptive teaching strategy should also be applicable to children, which was their main target group.

Peer, partner, opponent, or tool robots are instead presented as having a similar L2 level as the learners. The peer robot is 'learning' the language together with the learner; the partner robot interacts with the learner to solve a task using the target language; the opponent robot competes with the learner to solve a task in the target language; and the tool robot is controlled by the learners in the target language. Examples include:

Peer: *Tega* was presented as a peer or teammate when working with children around the tablet game created by Gordon et al. (2016), and

provided instructions, hints and encouragement in this role (c.f. Role play).

Partner: *iCat* interacted with learners who had previously been taught relevant words and phrases in ROILA to match words forming a logical pair (Mubin et al., 2013a). *Mec Willy* collaborated with Italian children, without any previous knowledge of English, in matching English words of fruits and vegetables to the correct pictures (Mazzoni & Benvenuti, 2015) as described in Collaborative language learning above.

Opponent: *Nao* played a competitive vocabulary game in which half of the children met a robot aiming to win, and the other half to lose (Balkibekov et al., 2016). The post-game-test not only clearly showed that the learners had improved their vocabulary, but also that girls competing with the winner robot learned more words than those playing with the losing, whereas the effect was the opposite for boys, illustrating the importance of adaptation of the robot's behaviour to factors, such as learner level, age and gender.

Tool: A *LEGO Mindstorms* robot was controlled by voice commands in ROILA (Mubin et al., 2013a), as described in collaborative language learning above.

As indicated by Table 1, robot peers, partners and opponents have been used together with all teaching strategies, and the combination to be considered is rather that between robot type, interaction and learner level and age, in that it needs to be aligned so that the more realistic the task and the more advanced the learners, the more anthropomorphic the robot needs to be in appearance and interaction behaviour in order for the learning to be transferable to real-world situations. As the tool role is not building on human–human interaction, these requirements do not apply.

Learner robots transfer the pedagogic initiative to the student to teach the robot, as exemplified by the verb learning study with *Nao* by Tanaka and Matsuzoe (2012), described under Physical interaction above. As the interaction is based on that with peers, the learner robot needs to have similar interaction behaviour as a less advanced peer. With increasing student age and level and learning task complexity, a more humanoid robot type and behaviour would therefore be required, in order for the student to anthropomorphise the robot in the learner role.

Social companion robots practice the second language by social interaction, rather than through explicit exercises. The robot–learner conversations with *Nao* (Khalifa et al., 2017) and the present study with *Furhat* below are examples of robot–learner interaction modelled on human social exchange. Kennedy et al. (2016) explored the social aspect further, by letting Nao give information about itself, ask how the learner felt

about the teaching material (the gender of French nouns), give more praise, match gaze, and refer to itself and the learner jointly with 'we' and 'our' in a 6-day study. The hypothesis was that interaction with a personal robot would result in better learning than with one that was impersonal. This hypothesis was not corroborated. However, in a follow-up study by Baxter et al. (2017), which was longer (two weeks) and focused on tasks in the native language, learning was indeed better with the personal robot. As the material was different, the two studies are not directly comparable, but they may nevertheless highlight the greater importance of forming a social relationship in long-term interaction. Since the social companion role relies on human–human interaction, a higher degree of anthropomorphic interaction behaviour is required.

Implications for RALL applied to adults

From the above analysis of previous work, it may be concluded that robot types, teaching strategies and robot roles have been combined differently, to take advantage of the robot type and to suit different learner age groups. RALL for adult learners may benefit from combining a teaching strategy directly related to human–human situations (hence using primarily TBLT, CLT or CLL to increase realism of communication), a robot role allowing learners to be involved in defining the practice (i.e., mainly using peer, learner, opponent or social companion roles), a robot type that allows for the use of human-like verbal and non-verbal interaction signals and an interaction promoting the learners' own intrinsic motivation (focusing on social exchange) rather than extrinsic (focusing on explicit rewards in the practice).

The line of research in the present work, which is described further below, consists in exploring technological, interactional and educational aspects of using a **robotic head** (*Furhat*) in the role as **social companion** in **collaborative language learning** focused on **communication practice** (social conversations) with pairs of **adult** learners. Based on the literature review above, it can be concluded that the use of the robot as a social companion, using communication practice as teaching strategy and adult learners as target group, signifies that a robot should have an anthropomorphic appearance, behaviour and expressiveness that promote social relationship and facilitate L2 understanding using visual information. The Furhat robot meets the requirements on human-like appearance, as it consists of a computer-animated face that is back-projected on a 3D facial mask (Al Moubayed, Beskow, Skantze, & Granström, 2012), and the face is therefore substantially more realistic and expressive than other robots' in terms of linguistic signals, such as lip and eye movements.

Furthermore, the motor-servo-controlled neck that allows Furhat to turn his head to address either person in a multi-party setting, is beneficial for communication practice, as it allows to replicate human–human interaction signals for turn-taking. Several studies using robots as social companions and/or in communication practice have further illustrated the importance of creating a personal relationship between learners and the robot, which signifies that the robot should both show an interest in getting to know the learners and have an own personality and personal background to convey during the conversation. The final aspect to consider concerns how to design the interaction in order to make the best use of the CLL teaching strategy. This aspect is covered in the next section.

Motivation and methodology for the user study on interaction in collaborative RALL

The well-established teaching paradigm of collaborative language learning, in which peers take advantage of joint knowledge and skills for clarifications, formulations and interaction management, has additional benefits for RALL, due to two shortcomings of robots compared to human language educators.

The first is pedagogical, i.e., that collaborative learning may to some extent compensate for the robots' teaching skills being vastly inferior to that of human instructors. Instead of relying solely on the learning effectiveness of the robot–learner interaction, parts of the practice take place between the peers, with the robot as catalyst.

The second is the technological, i.e., that current state-of-the-art of automatic speech recognition (ASR) and text-to-speech synthesis (TTS) may be problematic for non-native speakers:

ASR may fail either because the input from a learner is linguistically incorrect in terms of vocabulary or grammar; or because the ASR has not been trained on the learner's accent and hence misrecognises a correctly constructed utterance. Unless the scope is to achieve native pronunciation, learners expect a language learning system to respond appropriately to linguistically correct utterances, and justifiably also to a flawed one that could be properly understood by a human listener. In collaborative settings, the peer can both help with utterance formulation and confirm that a misrecognised utterance was in fact correct, thus avoiding that the learner gets an incorrect impression from the robot's non-understanding. Such erroneous feedback (explicit or implicit) could otherwise be detrimental for learning (Neri, Cucchiarini, Strik, & Boves, 2002).

TTS can rather adequately convey utterances with a natural standard intonation, emphasis and speed. However, in a RALL situation, when learners may have difficulties understanding parts of the utterance, it is far less obvious that a dialogue system can adjust as human teachers do: detect which parts need clarification and lower the speaking rate and put additional emphasis on that part or rephrase the utterance. In this situation, the peers may firstly collaborate to understand the robot's utterances and secondly, they may calibrate their views on the robot's speech. Where a solo learner may lose confidence by not understanding the robot, peers may together conclude that it is in fact the robot's TTS that is at fault.

Technical limitations may hence still influence the RALL interaction, and as a consequence two choices were made for the present study. Firstly, it was explored how two language learners interacting simultaneously with the robot can support each other. Secondly, a semi-automated wizard-of-Oz setup was used, i.e., unknown to the participants, a human operator used shortcut keys to select the most appropriate pre-generated utterance from the robot out of a small set presented to the wizard depending on the preceding robot utterance. This strategy was used to avoid that ASR misrecognitions or dialogue management failures (i.e., that an inappropriate response is selected) influenced the interaction in an uncontrolled manner. As the learners believed that they interacted with an autonomous robot and as the wizard could only select robot utterances determined by the system, the interaction should nevertheless be representative of that with an autonomous robot.

Design for the user study on robot interaction behaviours for conversation practice

We performed a user study in which Furhat conducted simplified conversations with two adult learners of Swedish. The topics were personal matters (hobbies, family, work, background), languages (e.g., similarities and differences between Swedish and the learners' first languages), Sweden or the news, all topics that occur frequently in language cafés (Engwall, Lopes, & Åhlund, 2020). The aim of the present study was to investigate how the robot's behaviour influences interaction and collaboration between the learners and between the learners and the robot. The four robot behaviours were:

As *Interviewer*, employing a Q&A strategy in sequential one-to-one interaction, Furhat addressed one learner at the time, asking a set of short, well-defined questions and then turned to the other learner with

similar questions. As *Interviewer*, he did not give information about himself, even if asked.

As *Narrator,* exploring *storytelling, quiz* and *egocentric small talk* in predominantly unidirectional robot-to-learner interaction, Furhat told the learners about a topic (e.g., robots or himself), asked them trivia questions about Sweden combined with telling them facts about the country, or held a conversation where he asked questions, but his objective was to convey his own story or opinions.

As *Facilitator,* promoting two-party learner–learner interaction, Furhat attempted to get the learners to talk with each other as much as possible, by encouraging them to suggest topics to discuss, only asking open questions to both learners and requesting that they comment on what the peer said.

As *Interlocutor*, using a personalised communication approach, Furhat was targeting a three-party interaction, in which both learners and the robot contribute with their personal stories and opinions and in which they comment on each other. In addition, Furhat was calling the learners, their home countries and mother tongues by name. This setting is hence a more personal combination of the other three: the robot asked direct questions, but then asked the peer for input and also gave its own views.

These four robot behaviours were based on a subset of interaction strategies of human moderators, determined through questionnaire responses by 105 moderators and observations of small language café conversations with three participants (Engwall et al., 2020). The selected behaviours were designed to differ along the two dimensions robot–learner initiative (from robot-led in *Interviewer* to learner-driven in *Facilitator*) and robot–learner focus (from robot-focused in *Narrator* to learner-focused in *Interviewer/Facilitator*). Figure 2 exemplifies typic robot utterances in the four behaviours, together with a schematic graphical overview of how the robot behaviour influences the interaction between the robot and the learners, with respect to if it mainly consists of spoken exchanges between robot and learners or between the two learners.

A user study was performed with 33 second language learners (18 women, 15 men, in the age range 20–54 years, average 32 years) at low intermediate level (B1 to B2, according to the Common European Framework of References for Languages) recruited from Swedish for immigrant courses. Their L1s were Arabic (10), Spanish (3), Italian (2), Polish (2), Russian (2), Ukrainian (2), Chinese (1), Croatian (1), Dari (1), Filipino (1), French (1), Greek (1), Kurdish (1), Persian (1), Portuguese (1), Punjabi (1), Somali (1) and Tigrin (1) and the set of country of origin and educational level was heterogeneous. The subjects were informed that they would be having four conversations with a social robot and a peer in order to assist the robot in practicing how to engage in social small talk but were
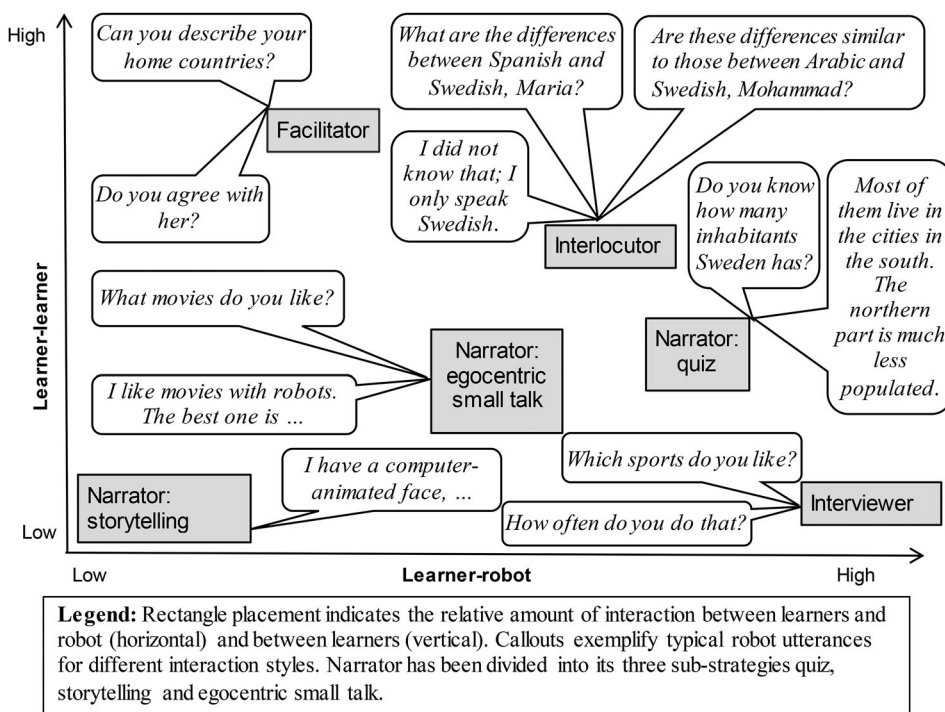
Figure 2. Schematic overview of amount and type of interaction in the user study.

not otherwise informed about the different robot behaviours or instructed how to interact with the robot. The set-up was that each pair of learners were seated side-by-side at a small round table with Furhat placed on the table, at the opposite side. Furhat could hence turn his head to address either one of the learners or both at the same time, and they could similarly turn to either Furhat or the peer.

The original study design, shown schematically in Figure 3, was that each learner should experience all four robot behaviours, during short conversations of 10–15 min each on two consecutive days (conversations 1–2 were performed on day 1 and conversations 3–4 on day 2). The subjects were paired with one peer on day 1 and another on day 2, in order to vary how well the subjects in the pairs knew one another beforehand, so that some knew each other well or superficially, while others had never met. In the original design, the order in which different pairs interacted with different behaviours was rotated, to ensure that the same number of subjects interacted with each robot behaviour in their first, second, third or fourth session and overall. However, as 8 students did not come to class on day 2, a modified study design was required, with 5 replacement subjects recruited to fill in the gaps. This resulted in 19 subjects experiencing all four robot behaviours, with an additional 14 experiencing 2, or in one case 3,

behaviours, and also differences in the total number of sessions with each robot behaviour, globally and between conversations 1–4. Quantitative and qualitative interaction analyses of the robot–learner interaction were made, and learner preferences were collected through a post-session questionnaire, as described in the following sections.

Methodology for quantitative and qualitative interaction analysis

Every session was a) video recorded with two webcams, each filming one participant's upper body, b) audio recorded with one head-mounted microphone per participant, and c) automatically logged with respect to the robot's utterances and head turns. The audio recordings were later transcribed manually, and the transcriptions were used as basis for quantitative analysis of the interaction. The qualitative analysis was performed using a grounded theory analysis of the 48 audio and video recordings and text of the transcriptions. The grounded theory analysis focused on collecting data on the amount of and examples of *interaction* in the three different categories {robot and individual learner}, {robot and both learners}, and {between the learners}, and the amount and type of *collaboration* between the learners. This data was then analysed to identify common properties of the conversations with the four different robot behaviours, in order to characterise them and contrast them against each other.

Methodology for collecting and analysing learner preferences

After each conversation with one robot behaviour, the two learners were given a tablet to fill in an electronic questionnaire, in which they assessed the conversation using a 6-point Likert scale from 0 to 5. For the present study, the ratings of learning effectiveness (*'How would you rate the session from a learning perspective'*) and the robot's interaction behaviour (*'How would you rate the robot's behaviour as a conversation participant?'*) were analysed to determine how preference was influenced by robot behaviour, session order, and learner gender, age, L2 level, experience of language cafés and familiarity. Data for the three latter properties was provided by the learners on a self-rating scale 0–5. The relationship between learner properties and preferences was investigated in order to find if any of the robot behaviours were more suitable for any particular category of learners, since the group of learners was substantially more heterogeneous (L1, age, educational level, as detailed above) than in previous RALL studies.

As a thorough statistical analysis of the full questionnaire is presented in Engwall et al. (2020), a binary analysis was instead performed to
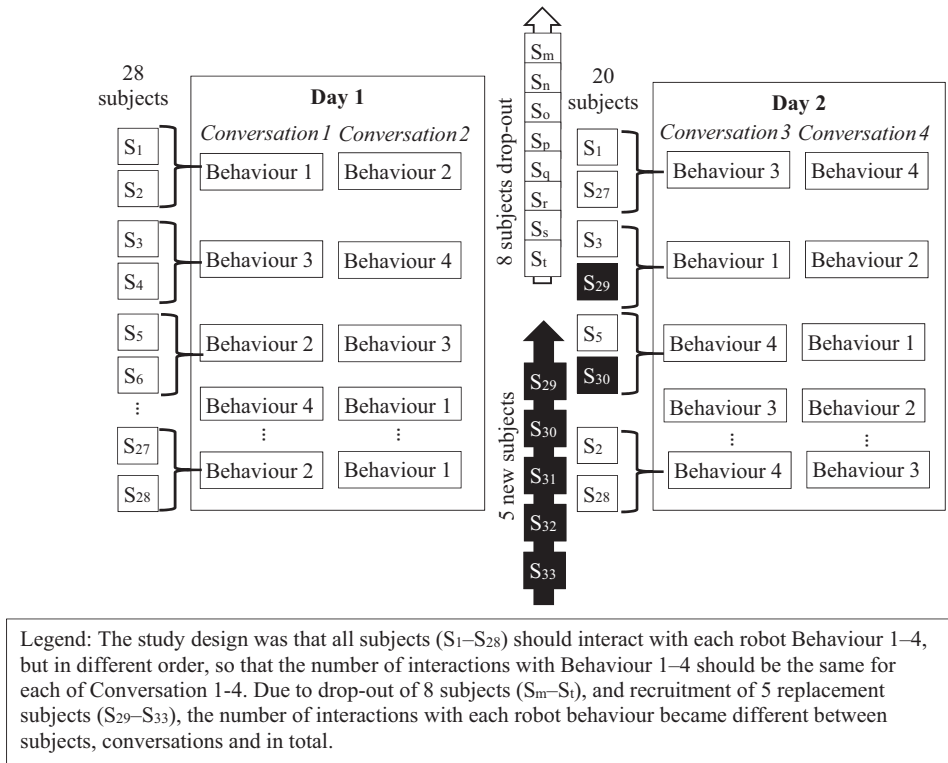
Legend: The study design was that all subjects (S$_1$–S$_{28}$) should interact with each robot Behaviour 1–4, but in different order, so that the number of interactions with Behaviour 1–4 should be the same for each of Conversation 1-4. Due to drop-out of 8 subjects (S$_m$–S$_t$), and recruitment of 5 replacement subjects (S$_{29}$–S$_{33}$), the number of interactions with each robot behaviour became different between subjects, conversations and in total.

Figure 3. Study design to assign robot behaviours to different conversations and learner pairs.

identify the most preferred robot behaviour, in general, and by different categories of learners. This analysis started by identifying conversations that were rated higher than the mean over all sessions by the 19 subjects who experienced all behaviours ($m_{learning}$=3.20 and $m_{dialogue}$=2.96). For these conversations, common characteristics for the subjects giving above-mean ratings were identified. That is, if these more positive subjects were younger or older than the mean over all subjects (m_{age}=32.4 - years), if they had more or less language café experience than the mean (m_{lc}=1.16), higher or lower proficiency than the mean (m_{prof}=2.37), if they were more or less familiar with the peer than the mean over all sessions (m_{pf}=1.99), and if the conversation was one of the first two or last two. In addition, the conversation (or conversations, as a subject could give the same rating to different conversations) that each participant rated the highest was identified, hence corresponding to an implicit intra-subject ranking of the robot behaviours. This rank analysis was performed since the comparison relative the mean over all subjects can be biased by inter-subject differences in rating (e.g., positive/negative subjects may rate all four behaviours higher/lower than the mean).

Results

Quantitative analysis of robot–learner interaction

Using the manually transcribed dialogues, the different robot behaviours were first quantified, in terms of:

a. average numbers of robot and learner statements per conversation (μ_{SR} and μ_{SL});
b. average numbers of robot and learner questions per conversation (μ_{QR} and μ_{QL});
c. the robot's and the learners' average share of the utterances during the conversation (σ_R and σ_L) and;
d. average lengths, measured as the number of characters, of robot and learner utterances (λ_R and λ_L);
e. the average total number of utterances during the conversation (Σ);
f. the total length of all earner utterances during the conversation (Λ_L), calculated as $\Lambda_L = \sigma_L \times \Sigma \times \lambda_L$, which hence takes into account both the total number of learner utterances and their length.

In the quantitative analysis (as opposed to the qualitative), the robot's statements and questions were not sub-divided to indicate if they were addressing one learner or both, nor were the learner utterances divided to indicate if they addressed the robot only, the peer only or both robot and peer, or if an utterance addressing the peer was intended for interaction or collaboration. This choice was made since the number of utterances per category would be small and the variation between conversations large. However, in general, the following patterns apply for the different behaviours: For *Interviewer,* robot questions and statements addressed one of the learners at the time and learner statements addressed the robot. For *Narrator*, robot questions and statements addressed both learners and learner statements addressed the robot. For *Facilitator*, robot statements addressed both learners, as did questions, unless the question was a request directed to one learner to comment on what the peer had said. The learner statements addressed the robot, the peer or both, depending on the interaction in the session, as described in *Qualitative analysis of robot–learner and learner–learner interaction.* For *Interlocutor,* the addressee of both robot and learner utterances varied more, but the robot predominantly addressed both learners, whereas the learners addressed the robot rather than the peer. Learner questions were in general directed to the robot. The large majority of peer-addressed learner utterances targeted interaction, and different types of collaboration occurred only in a few examples each.

The above measures can be used to compare the different robot behaviours quantitatively, since

a.　the balance between the number of robot statements μ_{SR} and robot questions μ_{QR} indicates if the focus of the conversation is that the robot or the learners should provide information.

b.　the balance between the number of learner statements μ_{LR} and learner questions μ_{LQ} indicates the extent to which the learners took initiatives in the conversation.

c.　the average shares of the utterances, σ_R and σ_L, together with the utterance lengths, λ_R and λ_L, measures if the robot or the learners were dominating the conversation.

d.　the total number of utterances (Σ) and their lengths (λ_R, λ_L) indicate how verbose the conversation was.

e.　the total length of all learner utterances (Λ_L) indicates the how much speaking practice the learners got during the conversation.

From these measurements, shown in Figure 4, it can be observed that:

Interviewer dialogues had the highest average number of questions per session from the robot ($\mu_{QR}=35$) of the four behaviours and the second highest number of statements from the robot ($\mu_{SR}=46$, mainly consisting of responses to the learners' answers). The average lengths of utterances were, respectively, $\lambda_R=58$ and $\lambda_L=42$. The learner statements were almost exclusively answers to robot questions, and their number ($\mu_{SL}=42$) was very similar between sessions (shown by the low standard deviation in Figure 4a). *Interviewer* had the largest average share of robot utterances ($\sigma_R=63\%$) and the average total length of learner utterances, $\Lambda_L=1989$ characters, was the second shortest.

Narrator dialogues contained the most robot statements ($\mu_{SR}=54$) and fewer robot questions ($\mu_{QR}=18$). The robot utterances were much longer ($\lambda_R=115$) than in the other behaviours and the learner utterances the shortest ($\lambda_L=24$), since the learners mainly acknowledged the information provided by the robot. *Narrator* further contained the highest number of long robot utterances ($\sigma_R=3\%$ of session utterances). The large difference in utterance length signifies that the robot was dominating the conversation. Even if the learners had almost as large share of the utterances ($\sigma_L=45\%$) and almost as many statements ($\mu_{SL}=42$), the estimated total length of learner utterances, $\Lambda_L=1415$ characters, was the shortest.

Facilitator dialogues had the fewest robot statements ($\mu_{SR}=23$) and questions ($\mu_{QR}=15$). The robot utterances were shorter ($\lambda_R=34$) than for other behaviours and the learner utterances longer ($\lambda_L=56$). In consequence, the learners dominated the conversation, with longer utterances
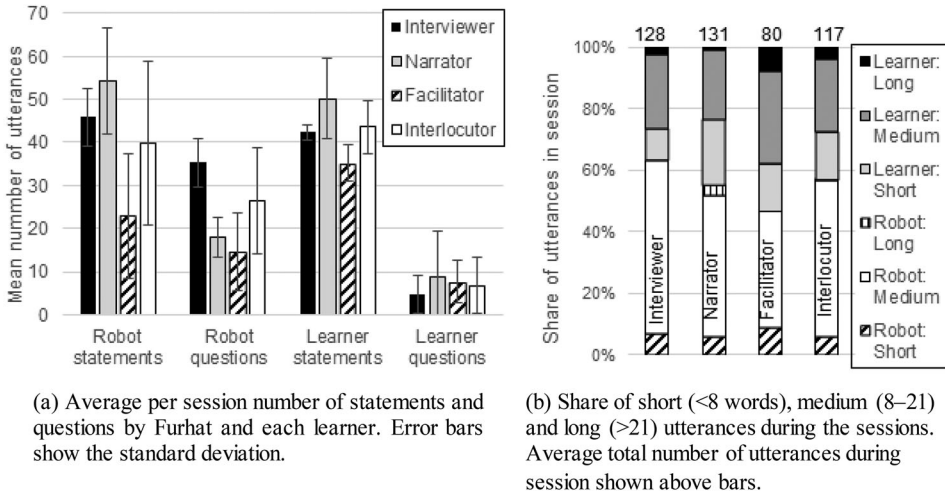
(a) Average per session number of statements and questions by Furhat and each learner. Error bars show the standard deviation.

(b) Share of short (<8 words), medium (8–21) and long (>21) utterances during the sessions. Average total number of utterances during session shown above bars.

Figure 4. Quantitative summary of the amount and type of interaction with different robot behaviours.

(σ_L=8% of the session utterances were long) and a higher share of the session (σ_L=53%). The total number of utterances during the session were the lowest ($\Sigma = 80$), but the estimated total length of learner utterances, Λ_L=2374 characters, was nevertheless the longest.

Interlocutor dialogues were similar to those with *Interviewer,* with the main differences being the lower number of robot questions (μ_{QR}=26) and statements (μ_{SR}=40), the lower total number of utterance per session ($\Sigma = 117$) and the slightly larger share of learner utterances (σ_L=43%), as the formulations of the robot utterances encouraged more verbose learner utterances (c.f. Figure 2). The length of robot and learner utterances (λ_R=56 and λ_L=44) and number of learner statements (μ_{SL}=44) were similar to that with the *Interviewer* behaviour, but the larger share of learner utterances means that the estimated total length of learner utterances, Λ_L=2213 characters, was the second longest.

This interactional data hence demonstrates that the different robot behaviours indeed resulted in quantitative differences, using the measures described above, in focus of the interaction (*Interviewer* was the most learner-focused, with most robot questions, and *Narrator* the least), in learner initiative (*Interviewer* had the fewest questions by learners and *Narrator* the most), in dominating conversational partner (*Facilitator* had the largest learner share and longest learner utterances, while *Interviewer* had the lowest learner share of utterances and *Narrator* the shortest total length of learner utterances), in total amount of robot–learner exchange (*Narrator* had the most utterances per session and *Facilitator* the fewest), and in total learner activity (*Facilitator* had

the longest total learner utterance length and *Narrator* the shortest). As a consequence, different robot behaviours may be more or less suitable for different categories of learners, which was explored using the qualitative interaction analysis and the learner preference responses.

Qualitative analysis of robot–learner and learner–learner interaction

The different robot behaviours were observed to lead to different interaction patterns between the robot and the learners and between the two learners.

Interviewer sessions led to very little content-based interaction between the two learners, as the learners directed their responses to the interviewing robot, rather than to both robot and peer. This behaviour is hence taking little advantage of the collaborative setting, but the clearly defined robot-led Q&A structure may nevertheless suit some learners, as discussed in the summary of this section below.

Narrator sessions had different interaction patterns, depending on if *storytelling*, *quiz* or *egocentric small-talk* was dominating the robot's utterances. For *storytelling*, interaction between learners was scarce and the learner–robot interaction mainly consisted of learner backchannels or short confirmations to the robot's narrative. For *quiz,* more peer interaction was observed, as the learners would discuss possible answers before replying. For *egocentric small-talk*, learners responded differently, as some were engaged by the personal information that the robot provided, which increased the amount of interaction, while others became more passive, thus decreasing the amount of two-way interaction. The reason for the learners responding differently may have been a combination of language proficiency, personality, relation to the peer in the session and previous conversations, but such an analysis is beyond the scope of this study.

Facilitator sessions differed substantially between learner pairs and contained – longer or shorter – intervals of *high*, *moderate* or *low* peer interaction.

In *high* peer interaction intervals, the participants were so engaged in discussing between themselves that they did not even include the robot in the conversation. This was a planned outcome of the *Facilitator* behaviour.

In *moderate* peer interaction intervals, both learners were involved in the conversation, but expected the robot to lead it, even though the robot tried to promote peer interaction, e.g., requesting them to compare their experiences or views and only addressing questions to the two learners together.

Low peer interaction intervals (which could last the entire session) led to inefficient interaction, or even temporary communication breakdown.

Inefficiency occurred, e.g., if the robot encouraged the learners to suggest a topic for discussion and they did not, resulting in the robot having to resort to its predefined topics, with questions addressed at both learners. Breakdown occurred if the learners did volunteer a topic, but then expected the robot to lead the conversation and ask questions, which it could not if that topic did not happen to be included in its utterance database. The robot then tried to transfer the initiative ('*Good topic, but I do not know that much about it. Could you start telling us something?*'), but this strategy was not always successful.

Moderate peer interaction was most common, followed by inefficient *low* and then *high*, illustrating that *Facilitator* was often not a successful interaction strategy. However, it was successful with some, more proficient, learner pairs.

Interlocutor sessions differed in a similar way, with *high*, *moderate* or *low* peer interaction intervals.

In *High* peer interaction intervals, learners mainly interacted with each other. This is not undesired per se, but since the experiment was aimed at differentiating between robot behaviours, the robot would interrupt in this setting after some time, to become part of the conversation.

In *moderate* peer interaction intervals, the conversation included both learners and the robot in a three-party conversation, just as intended for the *Interlocutor* behaviour.

Low peer interaction intervals became very similar to conversations with *Interviewer,* since the robot asked questions and follow-up questions that should involve both learners, but some of them nevertheless addressed their answers to the robot only.

Moderate and *low* peer interaction was the most common and in general, as already illustrated by the quantitative interaction measures in Figure 4, the interaction with *Interlocutor* was not very different from that with *Interviewer*, with the exception of the robot providing more personal information, on own initiative or after learner questions.

The qualitative analysis showed that different robot behaviours were the most appropriate for different learner pairs. *Interviewer* was most successful with learner pairs who did not already know each other or the robot (i.e., it was most appropriate in conversation 1 or 2) and with learners of lower (but not the lowest) self-rated proficiency, since the structured interaction with focused questions to an individual learner clearly guides the learner regarding her expected contribution. *Narrator* was, as described above, received very differently by different pairs, but this seems to be linked more to personality factors, such as if the learners were introvert or extrovert, and their interest in technology (when the robot talked about itself), which were not captured by the questionnaire information. *Facilitator* resulted in the

most active conversations with learner pairs who knew each other and had a sufficient L2 proficiency level to interact with each other in this less structured setting. As learners with the highest proficiency levels tended to ask the robot questions – which it could not answer – the behaviour was less successful with learners at both the lowest and highest levels. As *Interlocutor* included a combination of interaction elements from the three other behaviours, its success in engaging learners in conversation in a given conversation depended more on, on the one hand, how the interaction elements were combined in the conversation and, on the other, individual learners' preferences regarding these elements. Since the effectiveness of the robot behaviour depended on the learner pair, the robot should normally be ready to adapt or change strategy during the session. In this study, only one was used throughout, in order to contrast the different behaviours.

Qualitative analysis of peer collaboration

The *amount* of peer collaboration was to some extent influenced by how much peer interaction the robot behaviour led to (more collaboration with more peer interaction, i.e., for *Interlocutor* and *Facilitator*) and the activity of the learners (more collaboration when learners were more active, i.e., for *Interlocutor* and *Facilitator*). The *type* of peer collaboration did not primarily depend on the robot behaviour, but rather on general situations that could occur in conversations or dealing with shortcomings of the robot (TTS, perceived misunderstandings) regardless of robot behaviour. The types of collaboration identified were *Robot utterance interpretation* and *Learner utterance formulation*.

Robot utterance interpretation occurred through *peer assistance* or *peer collaboration*:

Peer assistance signifies that one learner clarified the robot's utterance in Swedish when the peer had difficulties understanding it (primarily by lowering the speaking rate of the entire utterance, altering the emphasis or clarifying a word that was badly pronounced by the robot). The assisting learner was thus acting as an intermediate, allowing the peer to attempt answering questions that would otherwise lead to communication breakdown. This is similar to strategies in multi-party human–human L2 interaction.

Peer collaboration signifies that the learners reasoned among themselves, in Swedish or in another language they had in common, to interpret what the robot said, rather than asking the robot for clarifications. The first reason may be that the robot could repeat its previous utterance, but not rephrase it or lower the speaking rate. Some pairs may therefore have concluded that it was not a fruitful strategy to ask the

robot for clarification. A second reason may have been that the two peers were physically (placement around the table) and mentally at one side and the robot on the opposite (some peers knew each other beforehand, all were from the same school, whereas the robot was an outside, native-speaking, visitor), and it was hence quite natural for the peers to collaborate. This reasoning was remarkably often whispered or accompanied by covering the microphone with the hand, to avoid that the robot 'heard' it (e.g., 'Did you understand what he said?" "I think it was…').

Learner utterance formulation collaboration occurred in two types. Firstly, that one learner explicitly asked the peer, in Swedish or another common language, for help finding a word (e.g., 'How do you say [word in English]?'). Secondly that a learner by own initiative offered the correct word when the peer hesitated, or recasting a problematic utterance (i.e., repeating the utterance but removing linguistical errors), to allow the peer to align her utterance. Both cases are clear examples of collaborative language learning, in which peers rely on joint linguistic resources.

Analysis of learner preferences for robot behaviour

The results of the analysis of the questionnaire are summarised in Table 2, in which the factors that are over-represented for higher than mean ratings and highest ranking are presented for each robot setting. The following observations can be made:

Interviewer was most often rated above the mean for both learning and dialogue behaviour and was ranked highest by most subjects for learning and second-most for dialogue behaviour (row 1). It was rated and ranked higher by male and older subjects and when the peer was not previously known (row 2–7).

Narrator had the second-lowest proportion of above mean ratings for learning and was ranked highest by the fewest subjects for both learning and dialogue behaviour. On the other hand, it had the second-highest proportion of above mean ratings for dialogue behaviour (row 1). *Narrator* was rated and ranked higher by female subjects and subjects with higher proficiency (row 2–7).

Facilitator had the lowest proportion of above-mean ratings for both learning and dialogue behaviour but was more often ranked as a preferred behaviour than *Narrator* on both aspects (row 1). It received the highest ratings from older subjects with higher proficiency (row 2–7).

Interlocutor received the second-highest proportion of above mean ratings for learning and was also ranked highest by the second-most subjects. For dialogue behaviour it was third in terms of above-mean

ratings, but was ranked first by most subjects (row 1). This behaviour received higher ratings from subjects with higher L2 level and to some extent by younger learners (row 2–7).

Session order influenced the ratings, so that *Interviewer* was more often rated above the mean when experienced as one of the first two behaviours, whereas the other three were more often rated above the mean and ranked highest when they were among the last two. It hence appears that previous interaction with the robot influences user preferences for robot behaviour, in terms of desired increased variation and complexity for subsequent conversations.

Language café experience also influenced the ratings so that more above-mean ratings for all behaviours except *Interlocutor* were given by subjects with a lower than average previous experience of language cafés.

It should be noted that the ratings for the *Narrator, Facilitator* and *Interlocutor* behaviours varied substantially between sessions. This is a natural consequence of the interaction depending on within-behaviour differences for the robot between sessions for *Narrator (quiz, egocentric small talk or storytelling)* and on how the peers interacted with each other (cf. *Qualitative analysis of learner interaction* above).

Summary and implications of results of the user study

Related to the research question, the following observations can be made, with pedagogical implications for CALL instructors summarised in bold (in the present context, the implications are stated for RALL and robots, but most are general and should apply also to CALL with computer-animated intelligent tutors):

Age – Adult learners in RALL: The quantitative amount of robot–learner interaction (on average 21–30 utterances per learner and 19–41 robot utterances per session), the qualitative analysis of the robot–learner and learner–learner interaction (a variety of different robot–learner exchanges, such as asking and answering questions, social greetings and acknowledgement, and learner–learner exchanges initiated by the robot) and reasonably favourable ratings of learning with the robot as a partner in conversational practice (averages of 3.5, 3.4, 3.3 and 3.2, for, in that order, *Interviewer, Narrator, Interlocutor* and *Facilitator*, when counted over all subjects), indicate that the practice was suitable for the adult learners in this study. As one of the first studies with this target age group, it hence shown that **RALL may be successfully used with adult learners**.

Teaching strategy – Communicative language teaching: *Interviewer* was the most preferred robot behaviour overall, followed by *Interlocutor*,

Table 2. Factors influencing the learners' rating of learning effectiveness (L) and the robot's dialogue interaction (D) for different robot behaviours.

		Interviewer		Narrator		Facilitator		Interlocutor	
		sessions	subjects	sessions	subjects	sessions	subjects	sessions	subjects
Above mean-ratings (Highest ranking)	L	53%	58%	32%	20%	21%	21%	42%	37%
	D	74%	42%	68%	16%	58%	32%	61%	53%
Proficiency (low-high)	L	**high** +8%	+12%	**high** +50%	+36%	**high** +25%	+17%	**high** +13%	+25%
	D			+19%	+25%	+25%	+17%	+8%	+28%
Lang. café experience (low-high)	L	**low** +30%	+14%	**low** +38%	+21%	**low** +50%	+33%		
	D	+14%	+19%	+17%	+25%	+4%	+50%		
Gender (male-female)	L	**male** +28%	+23%	**female** +22%*	+22%*				
	D	+14%	+16%						
Age (young-old)	L	**old** +8%	+19%	**young** +7%	+17%	**old** +25%	+17%	young?	+28%
	D			+13%	+13%	+14%	+23%		
Peer familiarity (low-high)	L	**low** +30%	+14%					**high** +17%	+6%
	D							+14%	+8%
Sessions (first-last)	L	**first** +20%	+5%	**last** +17%	+21%	**last** +14%	+50%	**last** +13%	+38%
	D	+14%	+19%					+25%	+6%

Legend: First row: (left) share of sessions for which the rating was higher than the mean over all subjects and sessions, and (right) share of participants (implicitly) ranking the setting highest. The greyscale heatmap illustrates the relative learner preferences for the different robot behaviours, from most often preferred (black) to least (white). Following rows: percentages refer to the over-representation of the stated factor compared to equal distribution, for, respectively, the session ratings and subjects' rankings. Factors in bold have at least +20% over-representation; empty cells correspond to balanced distribution for the factor.
*Equals 100% women.

which resulted in similar interaction. However, the interaction differed between learner pairs, and more proficient learners and pairs knowing each other better rated robot behaviours that allowed them to engage in more active learner–learner interaction higher, whereas less proficient learners and those who did not know the peer had a higher preference for interacting with the robot. In consequence, the most appropriate default robot behaviour in conversation practice appears to be that **the robot leads the conversation and prioritises robot–learner interaction,** but that the behaviour should be adapted, based on the learners' proficiency and familiarity, and possibly also gender and age.

Teaching strategy – Collaborative language learning: The benefit of using a collaborative setting, rather than one-to-one interaction with the robot, differed between sessions and learner pairs, as described in the analyses of interaction and collaboration above. Learner–learner interaction was nevertheless an important part of a sub-set of the *Narrator*, *Facilitator* and *Interlocutor* conversations. Collaboration on understanding the robot or formulating learner utterances played a constructive role in several conversations, in particular to alleviate problems related to the appropriateness of TTS for L2 learners. This demonstrates the general potential of **CLL for RALL to deal with both the learners' linguistic problems and technological shortcomings**.

Robot role – social companion: As *Interviewer* was preferred by more learners than *Interlocutor*, which in turn was preferred over *Narrator*, little evidence was found that making the robot more personal improved the interaction or the learners' perception of it during the short interactions in this study (in line with the findings of the short-term study by Kennedy et al., 2016, but contrary to long-term studies by Baxter, Ashurst, Read, Kennedy, & Belpaeme, 2017; Gordon et al., 2016; Kanda et al., 2007). However, learners asked Furhat personal questions (which the robot only answered in the *Narrator* and *Interlocutor* behaviours) and comments in the free form field of the post-session survey indicated that it was appreciated when the robot answered such questions. This suggest that efforts on creating a robot background story are worthwhile for social robot–learner interactions. Two other observations could be made with respect to robot behaviour in longer robot–learner interactions. Firstly, *Interviewer,* which has a clear structure of interaction focused on one learner at the time, was most preferred in the first sessions, whereas the other behaviours, which allow for more variability and more learner–learner interaction, were preferred in the last two. This indicates that, over a series of robot–learner conversations, **the robot's behaviour in RALL should develop to take development of the robot–learner and learner–learner relationships into account**. Secondly,

learners in this study commented when Furhat had asked a similar question in a previous session, and they rated the robot lower if they felt that the conversations became repetitions of previous sessions. Consequently, a robot memory, which keeps track of topics that have been covered with any particular learner and possibly the learner's answers, is required. Avoiding already covered topics will **reduce the repetitiveness of the conversations**, and previous learner answers may be employed to personalize the conversations.

Limitations of the study

The present study has a number of limitations affecting how representative the results can be considered to be for RALL in general.

Duration: Being a short-term study (40–50 minutes per learner over two days), it is not possible to draw conclusions on the long-term use of robots as conversational partners in L2 language practice. However, this is a limitation that is shared with most previous studies on RALL, which are often based on short practice duration, with a small number of interactions (1–4) during a restricted time (1–6 days). This signifies that the novelty effect must be considered to be an important factor and future work is in general needed to investigate long-term effects of RALL, in particular for the less restricted roles of peer, partner and social companion.

Heterogeneous learner group: Contrary to most previous RALL studies, the subjects differ substantially in terms of L1, L2 proficiency and age. This signifies that it is more difficult to conclude if observed differences between conversations are due to factors related to the practice (robot behaviour, peer familiarity, session order) or to the learners, and if so, which of the learner factors that were the most important for the observed differences. Rather than being focused on benefits for a particular learner background, this study investigates the general suitability of using a robot for conversational L2 practice with adult learners. However, the subjects are representative for the intended migrant learner group, which is indeed as heterogenous as in the study.

Subject drop-out: The fact that not all of the initially recruited subjects experienced all four robot behaviours (cf. *Study design*) limits the available amount of data that can be used to compare the different robot behaviours. The 19 subjects who did experience all four settings should nevertheless be a sufficient population size for the analysis made.

Semi-automated wizard-of-Oz study: As a human operator interpreted the learner utterances and selected the most appropriate predefined robot response, the performance of the robot was better than if a fully automatic ASR-based system had been used. However, as already discussed

above, the interaction is nevertheless to be considered as representative for RALL, since the learners believed that they were interacting with an autonomous robot and as all other parts of the robot's interaction were system-generated.

Discussion on future work in RALL

In order to further develop the aspects listed in *Summary and implications of results of the user study* above, the following further efforts are required:

Age – Adult learners in RALL: As earlier RALL studies have to a large extent focused on children, there is a need for further research on the application to adult learners. Our analysis of previous work indicated that a teaching strategy building on human–human communication should be used, that the robot should take a role that allows learners to contribute to define the practice, that the robot type should be able to use human-like verbal and non-verbal interaction signals, and that the interaction should focus more on social exchange than within-practice rewards. The interactional observations in this study indicate that adult learners accepted Furhat as a conversational partner and rated its behaviour reasonably high in particular in the *Interviewer* setting. Further work is required to extend the practice to, e.g., task-based learning conversations relevant for specific professions.

Teaching strategy – Communicative language teaching: When the robot leads conversations in a limited number of adequate domains (as *Interviewer*), a seemingly intelligent robot behaviour can be achieved with a predefined state-chart-based approach, but further research on approaches to real-time utterance generation is required for interactions with more learner initiative to allow the robot to respond to a larger variety of learner utterances. In addition, since the results of this study demonstrate that the robot behaviours had varied success with different learner pairs, and that learners' preference for robot behaviours was influenced by several different factors, the robot must adaptively combine several types of interaction and switch between these, depending on knowledge about the learners and/or the progression of the session. To determine when a change of strategy is needed, automatic tracking of learner engagement, frustration and anxiety is required, using e.g., computer vision analysis of facial features and acoustic tracking of prosodic or extra-linguistic vocal features (Ekman & Hartmanis, 2019; Ljung & Månsson, 2019; Mannerstråle & Hansson Svan, 2019; Melander & Wänlund, 2019 perform preliminary investigations on the material recorded in this study).

Teaching strategy – Collaborative language learning: CLL is used in the current setting to handle both learners' linguistic problems and technological shortcomings of TTS and ASR for use in L2 learning. Even if it has been argued above that the collaborative setting alleviates the problems caused by speech technology components, the goal is to allow the learners to focus collaboration on their own learning, rather than on communication problems with the robot. In order to achieve this, work on TTS adaptation for L2 learning and tailored ASR for L2 conversations is required. TTS adaptation may be needed for longer and more complex synthesised robot utterances. Adaptation to lower the speaking rate or emphasis of such utterances would in principle need to be made automatically in real time, after detecting learner difficulties understanding an utterance, and determining which part caused the problem. A short-cut is however available, by hypothesising where problems of understanding are likely to occur, using complexity analysis of the utterances and either lower the overall speaking rate for these utterances (as demonstrated by Hardin & Wellenstam, 2019 for the robot utterances in this setting), or pregenerate repair utterances with altered speaking rate, emphasis and/or replacement of specific words. Work on tailoring ASR for L2 conversations should focus on recognition of intended message rather than exact words (through keyword spotting and/or word vector distance to a knowledge database) so that utterances containing linguistic errors or deviant pronunciation are correctly interpreted. For this study, the recorded audio files were submitted to a state-of-the-art cloud ASR post-session, to investigate how well the learner utterances could be recognised. The amount of recognition errors, due to the learners' linguistic level, accent or hesitations and the relatively unconstrained dialogue in the language café setting demonstrate the need for implementing a more robust handling of L2 utterances in conversation.

Robot role – social companion: Longer-term RALL studies (Baxter et al., 2017; Gordon et al., 2016; Kanda et al., 2007) have found that it is important to establish a personal relationship between the robot and the learners in order to maintain interest, learner feelings towards the robot and motivation for the task. Whereas no such effects were found in this short-term study, previous work hence suggests that further work is required to strengthen the robot's social abilities, both verbal and non-verbal, for long-term use. Verbal ability improvement includes expanding personal background stories for the robot and the ability to present this story in a socially adequate manner, i.e., being able to retrieve the information if asked by the learners, or to reciprocate a learner utterance. Non-verbal social abilities that should be strengthened are e.g., matching learner gaze and displaying more facial emotional signals. Another aspect of social competence relates to transitioning from one robot behaviour to another in order to make

conversations more robust. As described above, audiovisual tracking may be used to detect *that* a change is required, but such tracking does not determine *what* change should be made. Simpler cases include that a more robot-led interaction may be required if interaction between peers is halting, or that a more learner–learner focused interaction may be fruitful for more proficient pairs, but how different robot interaction behaviours should be blended in general requires information on both the ongoing conversation and the learners active in it. A database storing information about the interaction between the robot and each learner can, together with sentence similarity measures, also be used to keep track of questions that have already been covered with a particular learner, even if the formulation of the questions varies, as demonstrated by Olsson and Södergren (2019) in a preliminary study on the robot utterances in this setting. However, substantially more work on the robot's dialogue generation module is required to allow previous learner responses to influence the dialogue. Finally, further pedagogical research is required on how to combine long-term use of RALL with classroom teaching. In this study, the level and topics of discussion were adapted to the general content of the courses that the students were attending, but if RALL should be useful as a long-term complement for L2 classes, it must be possible for teachers to select dialogue exercises that are relevant for the students at a particular time in the course.

In this article, a thorough analysis of the combination of robot type, teaching strategy and robot role in previous RALL studies was first presented, demonstrating in greater detail than in previous survey articles how the pedagogical design choice regarding this combination is influenced by features of the robot type and how it influences the robot–learner interaction. The user study on conversational practice with pairs of adult learners and an anthropomorphic robot then provided additional insights on the influence of robot behaviour on the interaction for this, to large extent new, target learner group.

Acknowledgements

Disclosure statement

Funding

References

Aidinlou, N.A., Alemi, M., Farjami, F., & Makhdoumi, M. (2014). Applications of robot assisted language learning (RALL) in language learning and teaching. *International Journal of Language & Linguistics*, 2(3), 12–20. doi:10.11648/j.ijll.s.2014020301.12

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. 1450022-1-25. doi: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. doi:10.1007/s12369-015-0286-y

Al Moubayed, S., Beskow, J., Skantze, G., & Granström, B. (2012). Furhat: A back-projected human-like robot head for multiparty human–machine interaction. In A. Esposito, A. M. Esposito, A. Vinciarelli, R. Hoffmann, V. C. Müller (Eds.), *Cognitive behavioural systems. Lecture Notes in Computer Science* (Vol. 7403, pp. 114–130). Berlin Heidelberg: Springer.

Asher, J. (1969). The total physical response approach to second language learning. *The Modern Language Journal*, 53(1), 3–17. doi:10.2307/322091

Balkibekov, K., Meiirbekov, S., Tazhigaliyeva, N., & Sandygulova, A. (2016). *Should robots win or lose? robot's losing playing strategy positively affects child learning.* Paper presented at IEEE International Symposium of Robot and Human Interactive Communication, New York, USA, pp. 706–711.

Baxter, P., Ashurst, E., Read, R., Kennedy, J., & Belpaeme, T. (2017). Robot education peers in a situated primary school study: Personalisation promotes child learning. *PLoS One*, 12(5), e0178126. doi:10.1371/journal.pone.0178126

Belpaeme, T., Vogt, P., van den Berghe, R., Bergmann, K., Göksun, T., de Haas, M., … Pandey, A. K. (2018). Guidelines for designing social robots as second language tutors. *International Journal of Social Robotics*, 10(3), 325–341. doi:10.1007/s12369-018-0467-6

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. doi:10.3102/0034654318821286

Ekman, P., & Hartmanis, E. (2019). *Facial activity recognition as predictor for learner engagement of robot-lead language cafés* (BSc thesis). KTH Royal Institute of Technology, Stockholm, Sweden.

Ellis, R. (2003). *Task-based language learning and teaching.* Oxford: Oxford University Press.

Engwall, O., Lopes, J., & Åhlund, A. (2020). Robot interaction styles for conversation practice in second language learning. *International Journal of Social Robotics*. doi:10.1007/s12369-020-00635-y

Gordon, G., Spaulding, S., Westlund, J. K., Lee, J. J., Plummer, L., Martinez, M., … Breazeal, C. (2016). *Affective personalization of a social robot tutor for children's*

second language skills. Paper presented at the 30th AAAI Conference on Artificial Intelligence, Phoenix, USA, pp. 3951–3957.

Han, J.-H., Jo, M.-H., Jones, V., & Jo, J.-H. (2008). Comparative study on the educational use of home robots for children. *Journal of Information Processing Systems*, *4*(4), 159–168. doi:10.3745/JIPS.2008.4.4.159

Han, J. (2012). Emerging technologies: Robot assisted language learning. *Language Learning & Technology*, *16*(3), 1–9.

Hardin, M., & Wellenstam, J. (2019). *Komplexitetsanalys av yttranden med hjälp av maskininlärning för att anpassa text-till-talsyntes [In Swedish: Machine-learning based complexity analysis of utterances to adapt text-to-speech synthesis]* (BSc thesis). KTH Royal Institute of Technology, Stockholm, Sweden.

Harwin, W. S., Ginige, A., & Jackson, R. D. (1986). A potential application in early education and a possible role for a vision system in a workstation based robotic aid for physically disabled. In R. Foulds (Ed.), *Interactive robotic aids: One option for independent living: An international perspective* (Vol. 37, pp. 18–23). World Rehabilitation Fund.

Jonell, P., Mendelson, J., & Storskog, T. (2017). Machine learning and social robotics for detecting early signs of dementia. arXiv:1709.01613v1

Kanov, M. (2017). *Sorry, what was your name again? How to use a social robot to simulate Alzheimer's disease and exploring the effects on its interlocutors* (MSc thesis), KTH Royal Institute of Technology, Stockholm, Sweden.

Khalifa, A., Kato, T., & Yamamoto, S. (2017). *Measuring effect of repetitive queries and implicit learning with joining-in type robot assisted language learning system*. Paper presented at ISCA workshop on Speech and Language Technology in Education, Stockholm, Sweden, pp. 13–17.

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), 61–84. doi:10.1207/s15327051hci1901&2_4

Kanda, T., Sato, R., Saiwaki, N., & Ishiguro, H. (2007). A two-month field trial in an elementary school for long-term human–robot interaction. *IEEE Transactions on Robotics*, *23*(5), 962–971. doi:10.1109/TRO.2007.904904

Kennedy, J., Baxter, P., Senft, E., & Belpaeme, T. (2016). *Social robot tutoring for child second language learning*. Paper presented at *ACM/IEEE International Conference on Human–Robot Interaction*, New Zealand, pp. 231–238.

Lee, S., Noh, H., Lee, J., Lee, K., Lee, G., Sagong, S., & Kim, M. (2011). On the effectiveness of robot-assisted language learning. *ReCALL*, *23*(1), 25–58. doi:10.1017/S0958344010000273

Lopes, J., Engwall, O., & Skantze, G. (2017). *A first visit to the robot language café*. Paper presented at ISCA workshop on Speech and Language Technology in Education, Stockholm, Sweden, pp. 7–12.

Ljung, M., & Månsson, L. (2019). *Klassificering av engagemangsnivå hos en samtalsdeltagare med hjälp av maskininlärning [In Swedish: Classification of a conversation participant's engagement level using machine learning]* (BSc thesis). KTH Royal Institute of Technology, Stockholm, Sweden

Mannerstråle, C., & Hansson Svan, A. (2019). *Prediktion av användar-omdömen om språkcafé-samtal baserat på automatisk röstanalys [In Swedish: Prediction of user ratings of language café conversations based on automatic voice analysis]* (BSc thesis). KTH Royal Institute of Technology, Stockholm, Sweden.

Mazzoni, E., & Benvenuti, M. (2015). A robot-partner for preschool children learning English using socio-cognitive conflict. *Educational Technology & Society*, *18* (4), 474–485.

Melander, G., & Wänlund, R. (2019). *Maskininlärning som medel för att betygsätta samtal med språklärande syfte mellan robot och människa In Swedish: A machine learning approach to rating robot–human language learning conversations* (BSc thesis). KTH Royal Institute of Technology, Stockholm, Sweden.

Mubin, O., Shahid, S., & Bartneck, C. (2013a). Robot assisted language learning through games: A comparison of two case studies. *Australian Journal of Intelligent Information Processing Systems*, *13*(3), 9–14.

Mubin, O., Stevens, C., Shahid, S., Al Mahmud, A., & Dong, J.-J. (2013b). A review of the applicability of robots in education. *Journal of Technology in Education & Learning*, *1*, 1–7.

Neri, A., Cucchiarini, C., Strik, H., & Boves, l. (2002). The pedagogy–technology interface in computer assisted pronunciation training. *Computer Assisted Language Learning*, *15*(5), 441–467. doi:10.1076/call.15.5.441.13473

Olsson, J., & Södergren, E. (2019). *Long term memory in conversational robots* (BSc thesis). KTH Royal Institute of Technology, Stockholm, Sweden.

Park, S. J., Han, J. H., Kang, B. H., & Shin, K. C. (2011). Teaching assistant robot, Robosem, in English class and practical issues for its diffusion. *Advanced Robotics and its Social Impacts* (pp. 8–11). Half-Moon Bay, CA. doi:10.1109/ARSO.2011.6301971

Nunan, D. (1992). *Collaborative language learning and teaching*. Cambridge: Cambridge University Press.

Randall, N. (2020). A survey of robot-assisted language learning (RALL). *ACM Transactions on Human–Robot Interaction*, *9*(1), 1–36. doi:10.1145/3345506

Richard, J., & Rodgers, T. (2014). *Approaches and methods in language teaching*. Cambridge: Cambridge University Press.

Schodde, T., Bergmann, K., & Kopp, S. (2017). *Adaptive robot language tutoring based on Bayesian knowledge tracing and predictive decision-making*. In Paper presented at 12th ACM/IEEE International Conference on Human–Robot Interaction, Vienna, Austria, pp. 128–136. doi:10.1145/2909824.3020222

Skantze, G., Johansson, M., & Beskow, J. (2015). A collaborative human–robot game as a test-bed for modelling multi-party, situated interaction. In *Intelligent virtual agents. Lecture Notes in Computer Science* (pp. 348–351). Berlin: Springer.

Spada, N. (2007). Communicative language teaching: Current status and future prospects. In J. Cummins, & C. Davison (Eds.), *International handbook of English language teaching* (pp. 271–288). Berlin: Springer.

Tanaka, F., & Matsuzoe, S. (2012). Children teach a care-receiving robot to promote their learning: Field experiments in a classroom for vocabulary learning. *Journal of Human–Robot Interaction*, *1*(1), 78–95. doi:10.5898/JHRI.1.1.Tanaka

Wedenborn, A., Wik, P., Engwall, O., & Beskow, J. (2016). *The effect of a physical robot on vocabulary learning*. Paper presented at International Workshop on Spoken Dialogue Systems Technology, Saariselkä, Finland (arXiv:1901.10461).

Wu, W.-C., Wang, R.-J., & Chen, N.-S. (2015). Instructional design using an in-house built teaching assistant robot to enhance elementary school English-as-a-foreign-language learning. *Interactive Learning Environments*, *23*(6), 696–714. doi:10.1080/10494820.2013.792844

You, Z.-J., Shen, C.-Y., Chang, C.-W., Liu, B.-J., & Chen, G.-D. (2006). *A robot as a teaching assistant in an English class*. Paper presented at IEEE International Conference on Advanced Learning Technologies, Kerkrade, The Netherlands, pp. 87–91.