STATE OF THE ART – ROBOTICS + ESL

TITLE
Belpaeme et al (2018) "GUIDELINES FOR DESIGNING SOCIAL ROBOTS AS SECOND LANGUAGE TUTORS"

KEYWORDS:
Social robot – Second Language learning – Robot Tutor – Human-robot interaction

SUMMARY

1- Introduction

They develop the L2TOR ('El Tutor') for 5 year preschool children.
One-to-one interaction between children and SoftBank NAO robot for teaching English to native speakers of Dutch, German and Turkish.
It is important to take into account the *Zone of Proximal Development*.
Human-Robot Interaction (HRI) is important.

```
Aspects of HRI:
- Human-computer interaction.
- Artificial intelligence.
- Robotics.
- Natural Language understanding.
- Design.
- Social sciences.
```

Automatic Speech Recognition (ASR)

```
Aspects of ASR:
- Automatic speech recognition (ASR)
- Computer speech recognition or speech to text (STT)
```

2- Second Language Learning

EU 2002 → All EU citizens have to learn at least two languages aside from their L1.
Preschool years are vital for L2 learning.
Patterns of L2 learning largely mirror those of L1 learning.
Interactive experience in learning a language is vital.
The amount of input predicts the vocabulary size in each language.
Adults must consider a balance between two languages for bilingual children.
Bilingual children have heightened sensitivity to those non-linguistic cues, rely more on eye gaze when learning novel words, they are better at understanding and using gestures and gaze direction to infer referential intent.
The usage of additional teaching opportunities becomes more important to increase the quantity of L2 input. The consideration of embodied technologies seems reasonable because they invite intuitive interactions.

3- Robots for Language Tutoring

3.1- Learning from Robots

a) <u>A social robot as an assistant to a teacher in a 5 week period</u>. A class with this robot learned significantly more than with the human teacher. It showed improved retention of acquired vocabulary.
b) <u>10 words to be taught by a robot left for 12 days</u>. Children showed a significant increase in the number of acquired words.
c) Robot tutoring can lead not just to vocabulary gains, but also improved speaking ability. Start with a lesson delivered by a computer, proceed to pronunciation training with a robot, which sould detect words with an expanded lexicon based on commonly confused phonemes.

3.2- Embodiment

Language allignment, i.e., the use of similar verbal patterns between interacting parting, when using an L2 appears to be not so affected when using a virtual robot, as opposed to a real one. Children between the ages of 4 and 5 were taught vocabulary showing a human and a robot issuing words. While children aged 5 were able to perform almost as well when taught by a robot, those aged 4 didn't seem to learn from the robot at all.

There is evidence for the physical presence of a robot having a positive impact on various interaction outcomes, including learning. Effectiveness of robot tutors lies in the social behaviour that a robot can exhibit and the motivational benefits this carries.

3.3- Social Behaviour

```
Components of behavior:
- biology
- environment
- cognition
- emotion
```

Studies:

A) Use an iCat robot as a tutor, which included verbal and non-verbal manipulations which aimed to influence feedback provision, attention guiding, empathy, and communicativeness. The tutor with these socially supportive behaviours significantly increased the child's learning potential when compared to a neutral tutor. This study used a variety of measures such as vocabulary acquisition, pronunciation and grammar.

B) Use a NAO robot to teach French and Latin verb conjugations to children aged 10-12 years old. In one condition robot would look towards the student whilst completing worksheets, but in the other, the robot would look away.

C) Verbal immediacy on the effect of learning. NAO was used to teach French with gender and articles. Children showed significant improvement in both conditions when comparing pre- and post-test scores and were able to retain their acquired knowledge as measured by means of retention test. Particularities of robot behavior could be potentially observed in the long term.

D) A robot was used to teach English as L2. A survey found that students who were taught by the robot were significantly less anxious about their lessons than those that were not, due to various factors, such as the fact that the robot was programmed to make intentional mistakes which the students should correct.

4- Designing Robot Tutoring Interactions for Children

4.1- Pedagogical issues

Although the process of language learning does not drastically differ between L1 and L2, there are a few differences.

L2TOR had some pedagogical guidelines based on existing literature and pilot data collected.

Guidelines: a) Age differences | b) Target word selection | c) use of meaningful context and interactions to actively involve the child | d) the dosage of intervention.

<u>4.1.1- Age Effects</u>

It is desirable to start L2 tutoring as early as possible, especially for children whose school language is an L2, because this could bridge the gap in language proficiency.

Preschool-aged children (3-5 years old) undergo major cognitive, emotional and social developments, such as the expansion of their social competence.

Targeted children aged between 3 and 4 years with a NAO robot. There was a first introduction to get familiarized with the NAO robot.

4.1.2- Target Words

Previous researchs recommend that vocabulary items should be taught in semantic clusters and embedded in a conceptual domain. For L2TOR, three domains were chosen: (a) number domain: language about basic number and pre-mathematical concepts; (b) space domain: language about basic spatial relations, and (c) mental states domain: language about mental representations such as ‘being happy’ and propositional attitudes such as ‘believe’ or ‘like’.
To select appropriate target words a number of frequently used curricular, standard texts and language corpora were used.
Target words should be based both on semantic coherence and relevance to the content domain and on children’s L1 vocabulary knowledge.

4.1.3- Meaningful Interaction

Explicit instruction on target words in meaningful dialogues involving defining and embedding words in a meaningful context yields higher word learning rates than implicit instruction through fast mapping or extracting meaning from multiple uses of a word in context as the basic word learning mechanisms.
Overall theme for L2TOR project is a virtual town that the child and the robot explore together, and that contains various shops, buildings and areas, which will be discovered one-by-one as the lesson series progresses. All locations are familiar to young children.

4.1.4- Dosage of Language Input

Previous research has shown that vocabulary interventions covering a period of 10 to 15 weeks with one to four short 15 to 20 minutes sessions per week are the most effective.
As for the number of novel words presented per session, the common practice is to offer 5 to 10 words per session, at least in L1 vocabulary interventions.
To determine the number of target words to be presented in the L2TOR project, a pilot study was conducted. The results showed that for children to learn any of these words at all, the maximum number of L2 words that could be presented in one session was six. The results also showed that a high number of repeated presentations of each word was necessary for word learning:

4.2- Child-Robot Interaction Issues

4.2.1- Introducing the Robot

The first encounter between a robot and a child plays a large role. A group introduction in the Kindergarten prior to one-on-one interactions influenced the subsequente interaction positively. Introducing the robot in a one-to-many setting was more appreciated than in a one-on-one setting, because the familiarity with their peers can reduce possible anxiety in children.
A first session is developed in which the robot is introduced to children in small interactive groups. The teacher first tells a short story about the robot using a picture book, explaining certain similarities and differences between the robot and humans in order to establish some initial common ground. During the story, the robot is brought into the room while in an animated mode to familiarise the children with the robot’s physical behaviour. The children and the robot then jointly engage in a meet-and-greet session, shaking hands and dancing together. Results show that almost all children were happy to engage with the robot during the group sessions, including those who were a big anxious at first.

4.2.2- Framing the Robot
It is beneficial to frame the robot as a peer, because children are attracted to various attributes of a robot and tend to treat a robot as a peer in long-term interactions.
Framing a robot as a peer could make it more acceptable, when the flow of the interaction is suboptimal due to technical limitations of the robot.
Framing the robot as a peer who learns a new language together with the child sets the stage for learning by teaching.
While the robot is framed as a peer and behaves like a friend of the child, the tutoring interactions will be designed based on adult-like strategies to provide the high quality input children need to acquire an L2, such as providing timely and sensible non-verbal cues or feedback.

4.2.3- Interaction Context
To facilitate language learning, it is important to create a contextual setting that provides references to the target words to be learned.
The embodied cognition approach, states that language is grounded in real-life sensorimotor interactions, and consequently predicts that childrens interactions with real-life objects will benefit vocabulary learning.
It could be expected that children learn new words better if they manipulate physical objects rather than virtual objects on a tablet, as the former allows children to experience sensiorimotor interactions with the objects. Findings indicate that it doesn't matter much whether the context is presented through physical objects or a tablet computer.
Displaying the context on a tablet doesn't seem to hamper learning, which is convenient, since using a tablet makes designing contexts more flexible and reduces the need to rely on complex object recognition and tracking. Tablets display the target objects but also allows children to perform actions on these objects.
Since at present ASR for children is not performing reliably, the roboc cannot monitor children's pronunciation or other verbal responses.
There is a focus on language comprehension rather than on language production. The use of a tablet in the ineraction allows to monitor the child's understanding of language and to control the interaction between child and robot.

4.2.4- Non-verbal Behaviour
The use of gestures facilitates L2 learning in various ways. Gestures could take the form of deictic gestures, such as pointing to refer to physical objects near the child, or of iconic gestures used to emphasize physical features of objects or actions in a more representational manner. Such iconic gestures help to build congruent links between target words and perceptual or motor information, so learners may benefit not only from observing gestures, but also by way of execution, such as enactment and imitation.
A robot tutor has the ability to use its physical embodiment to its advantage when interacting with the child, for example, through the manipulation of objects in the real world, or simply through the use of gestures for various communicative purposes. The robot's ability to use gestures is one of the primary advantages of using a robot tutor compared to a tablet computer, since it can enrich the language learning environment of the child considerably by exploiting the embodiment and situatedness of the robot to facilitate the child's grounding of the second language.
Despite the fact that non-verbal cues such as gestures aid learning, translating human's non-verbal behaviour to a robot remains a challenge, mainly due to hardware constrains, such as degrees of freedom or physical reach, making it unable to perform certain gestures. Motions may sometimes seem rigid, causing the robot's movements to appear artificial rather than human-like. Especially when certain subtleties are required, such shortcomings are not desirable.
There is a need to design a robot carefully and test appropriate referential gestures.

4.2.5- Verbal Behaviour

Using digital technology in robots means that they can be programmed to speak multiple languages without an accent, such as Text-to-Speech engines which can generate synthetic voices with few prosodic capacities. Children also rely on prosodic cues to comprehend spoken language. Moreover, adults typically use prosodic cues to highlight important parts of their speech when addressing children.

In a research, a robot-tablet concept was introduced, with the robot describing content displayed on the tablet screen, and the children were trained on how and when to provide answers by means of touching images on said screen. The children then proceeded with the main task, which involved the counting of animals in two languages. The interaction was managed by using multiple utterances from a control panel in order to prompt the children to give the answer ony after they were asked to. The operator triggered appropriate help and feedback from the robot to the child when required. Finally, at the end of the task, the robot asked the children to count up to five again with the robots help and then without any help at all.

Voice and video recordings were used to record the interactions with five children aged 4 to 5 years old. The first and final repetitions of the children pronouncing words and rated for accuracy on a 5-point Likert scale., indicating “excellent” agreement. Based on these ratings imply that repetitions generally improve pronunciation. Several children initially find it hard to pronounce L2 words without help. Task needs updating to improve the children’s recall.

Children can learn the pronunciation of the L2 from the robot’s synthetic voice, but it has to be compared to performance ratings of children that have learned the L2 from native speakers. It is worth noting that they seem to have some reservation speaking a foreign language.

4.2.6- Feedback

A typical adult-like strategy known to support language learning is the use of appropriate feedback. Adult caregivers tend to provide feedback explicitly and negative feedback implicitly by recasting the correct information. However, evidence suggests that a peer does not generally provide positive feedback and that they provide negative feedback explicitly without any correction.

A research has been approached by giving six children who stopped with the experiement before it was finished. The children were randomly assigned to one of the three conditiones, varying the type of feedback: adult-like feedback, peer-like feedback, and no feedback. The adult-like feedback of the robot used reformulations to correct the children in case they made a mistake and positive feedback when children responded correctly. In the peer-like condition, only explicit negative feedback without correction was provided whenever children make a mistake and no feedback was provided when they responded correctly. In the no feedback condition, the robot simply continued with the next task without providing any feedback.

During the experiment, the robot taught the native speaking children pronouncing L2 words. The interaction consisted of an introductory phase followed by the tutoring phase. During the introductory phase, the target words were described and associated with their concept in sentences with the vocabulary put into context. In the tutoring phase, the robot asked the child to pick up a certain number of blocks that had been placed in front of them. All intructions were provided in L1 and only the target words were provided in L2. After the child collected the blocks, the robot provided either adult-like feedback, peer-like feedback, or no feedback, depending on the experimental condition assigned to the child.

Results from a repeated measures indicated that on average, the children maintained their gaze significantly longer at the blocks and the robot than at the experimenter, regardless of their assigned condition.

The way the robot provides feedback does not seem to affect the engagement of the child with the robot. This would suggest that, as far as the child's engagement with the robot and task is concerned, it does not matter how the robot provides feedback or whether the robot provides feedback at all. Hence, the choice for the type of feedback that the robot should give can, thus, solely be based on the effect feedback has on learning gain.

4.3- Interaction Management

4.3.1- Objective

The objective is to realise robot-child tutoring interactions that provide a pleasant and challenging environment for the child, while at the same time being effective for L2 learning, when interaction management plays a crucial role.

4.3.2- Proposed model

A heuristic is employed that maximises the beliefs of all skills while balancing the single skill-beliefs with one another. This strategy is comparable to the vocabulary learning technique of *spaced repetition* as implemented. For the choice of actions, the model enables simulation of the impact each action has on a particular skill.

4.3.3- Results

The model was implemented and tested with a robot language tutor duing a game-like vocabulary tutoring interaction with adults.

We adopted the game 'I spy with my little eye'. In this game, the robot describes an object which is diplayed on a tablet along with some distractors, by referring to its descriptive features in an artificial L2. The sutdent then has to guess which object the robot refers to. The overal interaction structure, consisting of five phases (i.e., opening, game setup, test-run, game, closing), as well as the robot's feedback strategies were based on our observations og language learning. After the tutoring interaction, a post-test of the learned words was conducted.

The results revealed that learners' performance improved significantly during training with the personalised robot tutor. A mixed-design ANOVA with training phase as a within-subjects factor and training type as between-subject factor demonstrated a significant main effect of training phase, such that learners' performance was significantly better in the final phase as compared to the initial phase.

4.3.4- Outlook

This basic adaptive model can be extended by further integrating skills intedependencies as well as affective user states. Both have already been shown to improve learning. In addition, the model can, and is meant to, provide a basis for exploiting the full potential of an embodied tutoring agent, and will therefore be advanced to the extent that the robot's verbal and non-verbal behaviour will adapt to the learner's state of knowledge and progress. Specifically, it aims to enable dynamic adaptation of (a) embodied behavior such as iconic gesture use, which is known to support vocabulary acquisition as a function of dindividual differences across children; (b) the robot synthetic voice to enhance comprehensibility and prosodic focusing of content when needed; and (c) the robot's socio-emotional behaviour depending on the learners' current level of motivation and engagement.

5- Evaluation Framework for Robot L2 Tutoring

Some future guidelines can be described.

The first step in an evaluation is the development of pre- and post-tests designed to assess children's learning of the targeted vocabulary through comprehension and translation tasks, as well as tasks assessing deep vocabulary knowledge. Not directly targeted but semantically-related vocabulary will also be assessed, as well as general vocabulary and other skills related to word learning (e.g. phonological memory). This is important as children learn not only the words directly used, but can also use these words to bootstrap their further vocabulary learning in the same as well as related domains.

In addition to assessing children's L2 word learning, word learning process will be also evaluated during the interactive sessions between children and the robot by observing, transcribing, and coding video-taped interactions. Measures will include children's and the robot's participation and turn-taking, the type of questions, recasts and expansions, the semantic contingency of responses and expansions, and hte coherence and length of episodes within the sessions. All these aspects are known to promote language learning. Therefore, it is important to evaluate how these processes are taking place within the context of language learning with a social robot.

Finally, given the importance of motivation, an observation will be devoted to how children comply with the robot's initiatives and instructions, how involved they are in the intervention, and to what extent they express positive emotions and well-being duing the lessons. The intervention will consist of multiple sessions, such that children's learning, motivation, and interaction with a social robot can be judged over time.

The design of the evaluation study will be based on a comparison between an experimental and a control group. The experimental group will receive a placebo training (e.g., non language activity with the robot). This design is very common in educational research as it enables testing whether children who participate in an educational programme learn more or just as much as children who follow the normal curriculum. Additionally, learning gains with the robot will be compared to learnign gains using an intelligent tutoring system on a tablet, to test the additional value of a social robot above existing technology used in education. In evaluating the robot-supported program, the aim is not only to assess the effectiveness of the specific tutoring by the L2TOR robot, but also to provide recommendations for further technological development and guidelines for future use of social robots in (L2) language tutoring situations.