# Social Robots as Language Tutors

de Wit, Jan; Krahmer, Emiel; Vogt, Paul

*Published in:*
Proceedings of the Workshop on the Challenges of Working on Social Robots that Collaborate with People, ACMCHI Conference on Human Factors in Computing Systems (CHI2019 SIRCHI Workshop).

*Publication date:*
2019

*Document Version*
Publisher's PDF, also known as Version of record

# Social Robots as Language Tutors: Challenges and Opportunities

**Jan de Wit**
Tilburg University
Tilburg, the Netherlands
j.m.s.dewit@uvt.nl

**Emiel Krahmer**
Tilburg University
Tilburg, the Netherlands
e.j.krahmer@uvt.nl

**Paul Vogt**
Tilburg University
Tilburg, the Netherlands
p.a.vogt@uvt.nl

## ABSTRACT

In this paper we highlight several challenges we encountered while developing an Intelligent Tutoring System. Most importantly, technical limitations are currently standing in the way of the robot's ability to behave fully autonomously, and there is a need for methods and best practices from the field of human-computer interaction to ensure that user experience goals related to the quality of the holistic experience of interacting with a robot are set, and subsequently met. We also identify opportunities in the form of a modular (technical) architecture, and the implementation of a human-centered design process by including this discipline as one of the core components when setting up a project in the field of human-robot interaction.

This research is conducted as part of the L2TOR project, which has received funding from the European Union's Horizon 2020 research and innovation programme under the Grant Agreement No. 688014.



*CHI2019 SIRCHI Workshop, May 2019, Glasgow, UK*

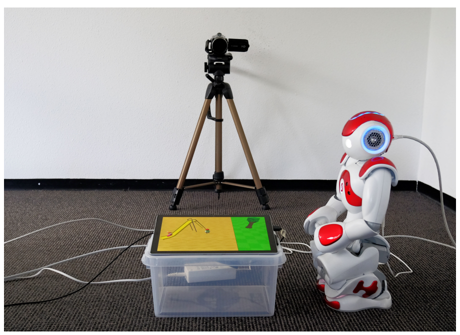

**Figure 1: The setup of the Intelligent Tutoring System (published with permission from [10]).**

## CCS CONCEPTS

• **Human-centered computing** → **Interaction design process and methods**; • **Applied computing** → *Interactive learning environments*; • **Computer systems organization** → **Robotics**; **Robotic autonomy**.

## KEYWORDS

Social robots; user centered design; design methodology; human-robot interaction

**ACM Reference Format:**

Jan de Wit, Emiel Krahmer, and Paul Vogt. 2019. Social Robots as Language Tutors: Challenges and Opportunities. In *Proceedings of the Workshop on the Challenges of Working on Social Robots that Collaborate with People, ACM CHI Conference on Human Factors in Computing Systems (CHI2019 SIRCHI Workshop)*. ACM, New York, NY, USA, 6 pages.

## INTRODUCTION

We have recently developed an Intelligent Tutoring System (ITS), which consists of a robot and a tablet on which educational content is presented. It was designed to teach children (aged five to six) English as a second language. Using this ITS, we conducted a longitudinal field study, which consisted of seven sessions with the robot, at several primary schools in the Netherlands. A total of 194 children participated in one of four experimental conditions: 1) a control condition where children had brief interactions with the robot, but no educational content was presented; 2) a tablet only condition where children interacted with the tablet while the robot was not physically present — its speech output was routed through the tablet's speakers; 3) a condition where children interacted with the robot and tablet, and the robot used deictic gestures to guide the child's attention; 4) the same as condition 3, with the addition that the robot also used iconic gestures whenever it mentioned an English target word. Figure 1 shows the experimental setup of the robot conditions.

The goal of this study was to investigate whether the implemented tutoring system was successful at teaching children English vocabulary, whether children learned more when interacting with a robot and a tablet compared to only having a tablet present, and whether the robot's use of iconic gestures resulted in higher learning gains than when the robot did not use iconic gestures. To ensure repeatability of the study, the experimental design and the proposed analyses were preregistered and the source code of the implemented system is made publicly available. Although research suggests that the learning process can be supported by a robot's physical presence and its ability to exhibit socially intelligent behavior [4], the results of our study do not provide evidence to support these claims [10]. A possible explanation for this is that the robot's social supportive behavior was limited due to various technical limitations (outlined in the following sections) and subsequent design decisions. In the following sections, we discuss the two major challenges we encountered while developing the aforementioned ITS, followed by opportunities we have identified for applying a human-centered process to the design of social robots. We aim to share our experiences, to underline the importance of including methodologies and guidelines from the field of human-computer interaction when designing similar systems, and to provide input for further discussion during the workshop.

## CHALLENGES

### Technical Implementation

An important aspect of a social robot is its ability to operate (semi-)autonomously [2]. This allows the robot to play an active role when interacting with humans, rather than merely facilitating human-human communication (i.e., telepresence). Based on a literature review, Beer et al. define robot autonomy as [3]:

> "The extent to which a robot can sense its environment, plan based on that environment, and act upon that environment with the intent of reaching some task-specific goal (either given to or created by the robot) without external control."

In the case of social robots, the task-specific goal — to teach English vocabulary — will include or be supported by the robot's socially intelligent behavior, which in turn relies on its ability to sense and understand its complex and dynamic environment, including any humans that are active therein. Not all technologies that power the sensing, planning and acting abilities needed for socially intelligent behavior are currently reliable enough to be used autonomously in a complex physical environment. As a concrete example, although Automatic Speech Recognition (ASR) is relatively reliable with adults, our project involved social interactions with children, speaking in their second language, and we could not achieve a usable level of ASR [1]. This greatly affected the design of the interaction, because we had to rely on a combination of Wizard of Oz (with the risk of introducing researcher bias) and different modes of getting input from children (e.g., manipulating objects in a virtual environment instead of verbalizing answers). Because of technical challenges regarding the robot's sensing capabilities, the subsequent planning and acting steps are based on abstract or incomplete information. In the case of lacking input from ASR, the robot is unable to maintain a personalized dialogue, which resulted in responses that were to a large extent scripted and impersonal. Furthermore, the robot's speech can contain imperfections and might lack emotion [1] and, due to its limited degrees of freedom, the robot is not able to gesture as fluently and with as much detail as humans. We have further discussed these specific challenges related to the production of natural language (and co-speech non-verbal behavior) at a workshop on Natural Language Generation (NLG) for HRI [12].

The decision to include a virtual environment for the human-robot interaction was not only based on the technical limitations related to the understanding and generation of natural language. Another

[1] We have recently become aware of a commercial solution, KidSense.ai, which might support our ASR needs.

challenge involved the robot's ability to sense and track physical objects in a dynamic environment [11], as well as to manipulate these objects [6], without either augmenting the objects or restricting the physical context. One final challenge we encountered was to reliably estimate the state of human interlocutors. Social interactions are complicated, even when studied between humans, therefore it is not a straight-forward process to implement socially intelligent behavior in robots [13]. This again relies on complex sensing functionalities, such as detecting whether someone is still engaged in the interaction [9], or identifying when miscommunication occurs [8], and then adapting the robot's behavior in real time based on these observations.

### HCI Practices

As the previous section shows, implementing human-robot interactions with socially intelligent behavior is inherently multi-disciplinary. It requires a number of different features for the robot to sense, plan, act and reflect (e.g., monitoring the successful completion of its goals). On top of that, these interactions need to be well-designed so that they allow the robot to achieve its goals, while providing a good experience. Therefore, the design of human-robot interactions stands to benefit from methods and processes from the human-computer interaction field. However, these methods will have to be adapted to accommodate the specific application domain of HRI. Lindblom [7] identifies three major challenges when trying to incorporate user experience (UX) into the field of HRI. The first challenge is the need for an iterative design process, which is relatively challenging to achieve due to the high costs of rapid prototyping with robots, the complexity of an engineering process that includes many hardware and software components, and variations in interactions when a robot performs autonomously. Besides the increased costs of adding such variations in favor of a more personalized interaction, we also found it challenging to ensure that our studies remained ecologically valid. Ideally we would want to provide a unique and optimal experience for each child, while still being able to make a fair comparison between children's learning outcomes.

The second challenge is to define user experience goals at the onset of a project, that focus specifically on the quality of the interaction between human and robot, rather than on specific behaviors or features of the robot. This means not only focusing on measuring the robot's task-specific goals (e.g., teaching children a second language) but also on the experience of interacting with the robot, and its ability to maintain a social dialogue with humans. These experience-oriented goals can then be used as a guideline throughout the development and evaluation of the system. Finally, there should be an awareness of the different user experience evaluation methods that exist, and their potential applications within the field of HRI, as well as methods that have been created or adapted specifically for evaluating human-robot interactions. This implies that we introduce a design culture to our projects, where all members are aware of, and share the responsibility for the user experience.

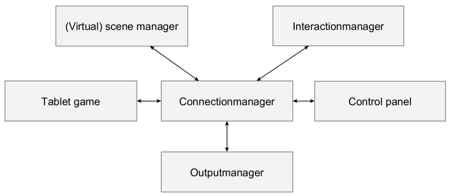


**Figure 2: The technical architecture of the intelligent tutoring system.**

[2]This is a known software architecture pattern, often used to reduce the interdependence between modules of a system.

## OPPORTUNITIES

We believe that the field of HRI is rapidly developing along two parallel paths. On the one hand, researchers in various related fields are investigating ways to improve the performance of individual technical components that drive social human-robot interaction, such as speech recognition. At the same time, all these components are combined into systems, which are then deployed in real-world settings. In order for these integrated systems to take advantage of the rapid technological developments occurring in parallel, we propose to use a modular approach. As an example, the modular architecture of our tutoring system is shown in Figure 2. It is important to highlight the central role of the ConnectionManager, which acted as a broker [2]. This allowed modules to communicate with each other by following a predefined message structure, regardless of differences in their programming language, operating system or the physical context (on the robot, tablet or researcher's control panel). This architecture supports loose coupling, so that individual modules can easily be replaced by different ones, for example when a shift from one robotic platform to another is needed or when new technological developments have resulted in a better performing alternative for a particular feature. Although existing platforms such as the Robot Operating System (ROS) support and stimulate a similar modular approach, we developed our system from scratch because there was a lack of prior knowledge about systems like ROS. In the future we would research existing solutions more thoroughly before creating our own, which would also simplify sharing of code with other researchers.

A modular approach not only speeds up the process of integrating new technological developments, it also facilitates an iterative and human-centered design process by reducing the time and costs involved with rapid prototyping. This would stimulate the inclusion of end-users and other stakeholders — in our case children and teachers — in the design process, as informant or design partner rather than observed user or tester [5]. Furthermore, it allows teams to work on different aspects of the system in parallel, where each module can be in various stages of completion. Temporary placeholder modules or a Wizard of Oz solution can be used for parts of the system that are time-consuming or technically infeasible to develop, so that the system as a whole can already be evaluated during early stages of the design process. This would also allow researchers to start with a simple version of their application where they are still controlling the robot directly, before gradually adding more autonomous behavior. We believe that projects working on the implementation of complex, social human-robot interactions would benefit from having at least one expert on human-centered design as part of the team. This ensures that a process is put in place, user experience goals are defined and end-users are involved. As a result, studies in HRI will become more robust and easier to reproduce.

## SUMMARY

The previous sections list several challenges we have encountered while developing our ITS, and work-arounds that we applied to cope with these challenges. However, these thoughts are based on one case study, so we hope to find out whether other research groups have encountered similar situations, and to discuss how to apply best practices from the HCI field to the context of social robotics.

## REFERENCES


[1] Jaclyn Barnes, Erin Richie, Qiyue Lin, Myounghoon Jeon, and Chung Hyuk Park. 2018. Emotive Voice Acceptance in Human-Robot Interaction. In *Proceedings of the 24th International Conference on Auditory Display*.

[2] Christoph Bartneck and Jodi Forlizzi. 2004. A design-centred framework for social human-robot interaction. In *Proceedings of the 13th IEEE International Workshop on Robot and Human Interactive Communication, RO-MAN*. IEEE, 591–594.

[3] Jenay M Beer, Arthur D Fisk, and Wendy A Rogers. 2014. Toward a framework for levels of robot autonomy in human-robot interaction. *Journal of Human-Robot Interaction* 3, 2 (2014), 74–99.

[4] Tony Belpaeme, James Kennedy, Aditi Ramachandran, Brian Scassellati, and Fumihide Tanaka. 2018. Social robots for education: A review. *Science Robotics* 3, 21 (2018).

[5] Allison Druin. 2002. The role of children in the design of new technology. *Behaviour and Information Technology* 21, 1 (2002), 1–25.

[6] Charles C Kemp, Aaron Edsinger, and Eduardo Torres-Jara. 2007. Challenges for robot manipulation in human environments [grand challenges of robotics]. *IEEE Robotics & Automation Magazine* 14, 1 (2007), 20–29.

[7] Jessica Lindblom and Rebecca Andreasson. 2016. Current challenges for UX evaluation of human-robot interaction. In *Advances in ergonomics of manufacturing: Managing the enterprise of the future*. Springer, 267–277.

[8] Michael McTear. 2008. Handling Miscommunication: Why Bother? In *Recent trends in discourse and dialogue*. Springer, 101–122.

[9] Thorsten Schodde, Laura Hoffmann, and Stefan Kopp. 2017. How to manage affective state in child-robot tutoring interactions?. In *Companion Technology (ICCT), 2017 International Conference on*. IEEE, 1–6.

[10] Paul Vogt, Rianne van den Berghe, Mirjam de Haas, Laura Hoffmann, Junko Kanero, Ezgi Mamus, Jean-Marc Montanier, Cansu Oranç, Ora Oudgenoeg-Paz, Daniel Hernández García, Fotios Papadopoulos, Thorsten Schodde, Josje Verhagen, Christopher D. Wallbridge, Bram Willemsen, Jan de Wit, Tony Belpaeme, Tilbe Göksun, Stefan Kopp, Emiel Krahmer, Aylin C. Küntay, Paul Leseman, and Amit K. Pandey. 2019. Second Language Tutoring using Social Robots: A Large-Scale Study. (2019). To appear in *the 2019 ACM/IEEE International Conference on Human-Robot Interaction*.

[11] Christopher D Wallbridge, Séverin Lemaignan, and Tony Belpaeme. 2017. Qualitative Review of Object Recognition Techniques for Tabletop Manipulation. In *Proceedings of the 5th International Conference on Human Agent Interaction*. ACM, 359–363.

[12] Bram Willemsen, Jan de Wit, Emiel Krahmer, Mirjam de Haas, and Paul Vogt. 2018. Context-sensitive Natural Language Generation for Robot-assisted Second Language Tutoring. In *Proceedings of the Workshop on NLG for Human–Robot Interaction*. 1–7.

[13] Guang-Zhong Yang, Jim Bellingham, Pierre E Dupont, Peer Fischer, Luciano Floridi, Robert Full, Neil Jacobstein, Vijay Kumar, Marcia McNutt, Robert Merrifield, et al. 2018. The grand challenges of Science Robotics. *Science Robotics* 3, 14 (2018), eaar7650.