"This is the Fluffy Robot that Only Speaks French": Language Use Between Preschoolers, their Families, and a Social Robot While Sharing Virtual Toys

by
Natalie Anne Freed

B.S. Computer Science
M.S. Computer Science with concentration
in Arts, Media, and Engineering
Arizona State University

Submitted to the Program in Media Arts and Sciences,
School of Architecture and Planning
in partial fulfillment of the requirements for the degree of
Master of Science in Media Arts and Sciences
at the
MASSACHUSETTS INSTITUTE OF TECHNOLOGY

September 2012

Author .... **Natalie Freed**
Program in Media Arts and Sciences,
School of Architecture and Planning
August 10, 2012

Certified by .... **Cynthia Breazeal**
Associate Professor of Media Arts and Sciences
Thesis Supervisor

Accepted by .... **Patricia Maes**
Professor of Media Arts and Sciences
Associate Academic Head, Program in Media Arts and Sciences

"This is the Fluffy Robot that Only Speaks French": Language Use Between Preschoolers, their Families, and a Social Robot While Sharing Virtual Toys

by

Natalie Anne Freed

Submitted to the Program in Media Arts and Sciences,
School of Architecture and Planning
on August 23, 2012 in partial fulfillment of the
requirements for the degree of Master of Science
in Media Arts and Sciences

Abstract

This thesis proposes an approach to language learning for preschool aged children using social robots as conversation partners within a shared play context for children and their families. It addresses an underserved age for language learning, where early learning can greatly impact later educational success, but that cannot benefit from text-based interventions.

With the goal of establishing a shared physical context between multiple participants without absorbing all of the children's focus onto digital content, a hybrid physical and digital interface was iteratively designed and play-tested. This interface took the form of a "café table" on which the child and robot could share food. A robot was programmed to introduce itself and name foods in French, eat some foods and express dislike towards others, respond with distress to a new object, show its focus of attention through gaze, and in one experimental condition, express feedback about its comprehension when spoken to in French or English.

The study found that some children as young as 3 years old would treat a social robot as an agent capable of understanding them and of perceiving a shared physical context, and would spontaneously modify their use of language and gesture in order to communicate with it – particularly when the robot communicated confusion. The study also found that parents tended to frame their scaffolding of the children's behavior with the robot in a social context, and without prompting aligned their guidance and reinforcement with language learning goals. After one exposure to the robot and new French vocabulary, children did not retain the robot's utterances, but engaged in communicative and social behaviors and language mimicry throughout the interaction. The system appeared to support multi-user social participation, including both caretakers and siblings of the participants.

Thesis Supervisor: Cynthia Breazeal

Title: Associate Professor of Media Arts and Sciences

"This is the Fluffy Robot that Only Speaks French": Language Use Between Preschoolers, their Families, and a Social Robot While Sharing Virtual Toys

by

Natalie Anne Freed

The following people served as readers for this thesis:

Thesis Reader..**Cynthia Breazeal**
Associate Professor of Media Arts and Sciences
MIT Media Arts and Sciences

Thesis Reader..**Rosalind Picard**
Professor of Media Arts and Sciences
MIT Media Arts and Sciences

Thesis Reader..**Nonie Lesaux**
Associate Professor in Human Development and Urban Education
Harvard Graduate School of Education

"You don't know what a joy it is to run in your own language until you've stumbled in someone else's."

- *Gregory David Roberts, author*

This thesis is dedicated to my French and American grandmothers,
Christiane Baudot and Marian Scudder,
who both in ways as different and wonderful as your styles of cooking
were masters in the Grandmother Method of encouraging,
and are still with me today.

Acknowledgements

To my readers: Many thanks and credit are due to my advisor Dr. Cynthia Breazeal both for her expertise on social robotics and for her vision and enthusiasm about the future of education. You pushed me to take on new challenges while providing the resources to make it possible. To Rosalind Picard, for sharing the possibilities of Affective Computing and for the thoughtful discussions you facilitated. To Nonie Lesaux at Harvard Graduate School of Education and your collaborators, for grounded and very motivating advice on language acquisition in the classroom.

To my group: Thank you Polly Guggenheim for your warmth, sense of humor, and for listening. Adam Setapen for everything you've built and taught me and for all the times you've made me fall off my chair laughing. Sooyeon Jeong for being a wonderful self-driven learner. Florian Pellet for your "grit," and arriving to help just at the right time. Angela Chang, for generously sharing your space and for your wise advice and example. I won't forget M.O.S.C.O.W! Fardad Faridi for creating extraordinary designs from last-minute requests and for always being willing to teach your craft. Jin Joo Lee for your accessible teaching style and your leadership. David Robert, for your empathetic insights and extreme creativity. Nick DePalma, for thoughtful answers and thought-provoking questions. Siggi, for pulling punches! Also for sharing much EE and R1D1 wisdom. And Kris, Nancy, Kenton, Jesse, Matt, Adam W., Phillip, Peter, Jason, David N. for your great work, encouragement, help, discussions, debugging, and general awesomeness.

To my friends: Jie Qi for being partner-in-mischief and shoulder to cry on, and for sharing your making and repair-skills. Ryan McDermott for always being there; whether to humor random requests to take photographs of fruit or write a choose-your-own-adventure website to help me face writing. Matt Miller for machine learning lessons and sober life advice. Forrest Green for keeping me going with puzzles, games, and being on my team. Crafternooners and movie buddies, for making spaces happy.

To mentors and teachers along the way: Dr. Edith Ackerman, for the trickster archetype and your poetic description of "song" of a language. Dr. Christopher Dede, for your support and feedback on motivation, engagement, and flow. Jay Silver and Eric Rosenbaum for radical learning, both upper and lower case.

And to my family: Emma, Inigo, Portia, Arthur, Colin, and my parents: thank you for cheering me on, for being patient, and for the motivation and inspiration. Love you!

Technical credit is due to:
Adam Setapen: created Dragonbot hardware and software platform.
David Robert: projected environment for Le Fonduephone.
Fardad Faridi: graphics for fruit, "French cafe" backdrop projections, Dragonbot animations and hands/feet/faceplate designs, ideas and feedback.
Sooyeon Jeong: projected table software, dragonbot software, teleoperation.
Florian Pellet: robot construction, logging software, teleoperation, voice of the robot.
Emma Freed: helped brainstorm fondue concept, edited conversational modules.
Forrest Green: log parsing scripts
Ryan McDermott: CGI programming help, fruit photography
Ehsan Hoque: Voice and facial expression analysis support.
Didier Roncin: voice of the robot.

Table of Contents

I. Introduction: Why Robot Learning Companions?

Preschool is an opportune age to introduce a new language, for newcomers to a country and beyond. Yet schools lack resources, passive media such as television cannot teach the structure of language, and technological interventions (eg. software games) that work for children who can read and write cannot be used. Rather, children of this age learn language through social interaction. Social robots combine the scalability of digital media with human-readable social cues and the creation of a social context for communication. The vision of this research is a robotic playmate that speaks a foreign language and motivates young children to practice speaking and understanding in order to communicate and play with an engaging character.

This thesis describes the design and evaluation of an immersive, socially situated language learning experience for preschoolers comprising:

- A French-speaking robotic playmate modeled as a younger peer.
- A physically and spatially shared context including changeable digital props.
- A multiuser activity for a robot, a child, and the child's family.

The values reflected in this work are the importance of entire-family participation in interaction with technology, of leaving room for imagination, and of instructing through emphasis on exploration rather than correctness.

Motivation: Preschool as Pivotal Moment

This research specifically targets preschoolers (3-5 years) as the age and grade level where adding foreign language acquisition support may have the greatest impact on later social and academic success.

Though the "critical period" hypothesis – that language is more easily acquired within a certain age range – still does not have a definitive answer in language acquisition research, certain advantages to learning at a young age are well-established. One is that accent and pronunciation - which can have a large impact on the social acceptance of a speaker as a competent communicator - are age-sensitive (Snow & Hoefnaget-Hohle, 1977). Another is that the sooner newcomers to a country can catch up on the language, the more successful they are likely to be socially and academically (Garcia, 2011). The academic achievement gap for children who are behind in English skills at Kindergarten age persists throughout later schooling, making a strong case for addressing language learning at the preschool level. Finally, while this factor may vary by personality, anxiety caused by "stumbling" in a new language may be more comfortable for children still establishing a notion of identity than for adults.

The Question: Can Technology Support the Social Nature of Language Learning?

There is evidence that initial language acquisition is "socially gated." Recent research by Patricia Kuhl shows that infants exposed to audio and video recordings of speech in English or Mandarin

Chinese did not learn to differentiate between phonemes in the new language, while infants who heard a person present in the room speaking did learn (Kuhl, 2007).

As children grow old enough to read and write and develop a sufficient grasp of their native language, they are able to use mechanisms other than social interaction to learn a new language. Textual representation breaks language into words and sentences that can be mapped to vocabulary and grammatical structures in their known language. Once children can read, drill and practice software based on text (as most are) can be effective technological interventions.

Before literacy, in the target age for this research, the only way for children to learn is through listening to and testing out the effect of speech. Moreover, their internal representation of language is still under development at this stage. Text-based software cannot help, and studies have shown that children learn only a limited subset of language from watching television. Viewing educational television shows can affect vocabulary, but grammar rules and complex sentence structures are not learned (Naigles & Mayeux, 2001). Young children do not initially learn language by connecting individual vocabulary words together, but (in both native and foreign languages) by applying "formulas" they have heard to new contexts and receiving feedback from those they are communicating with (Tannen). Interactivity is critical. In addition, without a shared context between speaker and listener or speech contingent on the child's focus of attention, it can be impossible to disambiguate meaning from a stream of unfamiliar speech. A human teacher can use shared attention, contingent actions in a shared environment, and prosodic variations in speech to help isolate what is being referred to and supply necessary feedback to the learner's adapting mental model.

So can technology still help support language learning at this age? This work proposes that components of socially supported language learning grounded in a shared world might be replicated with a robot as the social actor.

Social robots combine human social cues and sensory capabilities with the scalability and flexibility of technology. They leverage the way people already know how to communicate with each other to provide intuitive interfaces. When implemented effectively, cues such as gaze, expressions of affect, and contingent but independent-goal directed behaviors are interpreted naturally by humans as social cues expressing attention, desires, and intent. This has been shown in children as well as adults: as early as 18 months, children will follow a robot's gaze if they observe adults doing so (Meltzoff, Brooks, Shon, & Rao, 2010). These cues are critical for socially situated language learning.

Research Goals: Language Learning Companion

My research goals are to create and evaluate a novel second-language technology for preschool age children comprising the following components:

1. A social robot presented as a native speaker of another language with the goal of teaching skills related to how to communicate with a native speaker. Both children

learning a new language and children interacting with them can benefit from having a space to experiment with this type of interaction.

2. A robot modeled as a peer or slightly younger child. It is small, and non-threatening. It can be taught, as well as teach – allowing children agency, the possibility of directing the play, and the opportunity to practice both speaking and listening.
3. A robot that is a plush imaginary creature rather than a humanoid robot, with the goal of supporting pretend play and storytelling in the process of language use, and of supporting not only children who are excited about exploring mechanical toys but also those who gravitate towards social pretend play.
4. A shared activity that has changeable digital components but also has a cohesive physical presence and supports multiuser face-to-face interaction – with the goal of creating a shared physical context between humans and robot, as well as making possible entire-family participation with the technology.
5. Evaluation in a context that includes caretakers and children, and design to facilitate adult guidance and participation. The robot serves as a mediator around which to scaffold effective communicative and social behaviors.

Thesis Overview

II. Background and Approach: Related work and theory underlying the approach.

III. Explorations and Design Insights: Informal experiments that guided the design. Designing, play-testing, and building a shared activity for robot and child.

IV. Final System: Social Robot and Shared Context: Descriptions of components of the system including social robot and shared game designed for the final study.

V. Final Study: Randomized controlled study of children's language use during a shared activity.

VI. Conclusion: Contribution, lessons learned, and remaining challenges.

VII. Future work: Future directions for the work both in refining the evaluation and extending the system.

II. Background and Approach

Prior Work: Robot Language Instructors

In a survey of humanoid robots used as language teaching tools, Chang et al. list characteristics of robots that motivate their use in language learning: repeatability (necessary for practice), flexibility (curriculum developers, teachers, and students can customize), digitization (interaction data can be saved, manipulated, and learned from; all the resources of the web and the cloud are available), "humanoid appearance" (may appeal, engagement, and motivation), body movement (capability to express gestures to support understanding), interaction (they can participate in social situations), and anthropomorphism (the agent is functionally treated as a real speaker/listener while reducing speaking anxiety) (Chang, Lee, Chao, Wang, & Chen, 2010).

In a recent study of a language-teaching social robot named RUBI in an American preschool classroom, Movellan et al. (Movellan, Eckhardt, Virnes, & Rodriguez , 2009) found that children recalled words taught by the robot compared to control words that it did not teach. After 13 sessions, they also appeared to have adopted the robot as a social participant, showing social behaviors including pointing, hugging, imitation, social referencing, and gaze following. In order to teach new words, RUBI had a screen on its chest that children could use to play games, and would ask children to point to images of the new words on this screen, using physical gestures to direct attention to on-screen events. It would also pick up, name, and pass to the children physical objects.

RUBI spoke the children's language and introduced foreign nouns interspersed with English speech the children could understand. Children were able to learn the foreign vocabulary words framed by their own language, but the study did not explore whether they would be able to learn in a more immersive language environment. In other research, robots were modeled as foreigners not speaking the children's language – although the existing studies on this approach do not address children at pre-literacy and early language development stage. Kanda et al. conducted a two-week field trial of a humanoid robot, "Robovie," to teach English in a Japanese elementary school classroom (Kanda, Hirano, & Eaton, 2004). They found that after the second week, the children who formed a bond with the robot persisted in interacting and improved their vocabulary of conversational English phrases compared to peers who stopped interacting. In addition to the learning outcomes found, this may suggest that establishing a connection with the robot – affective and social factors - can affect motivation to learn the language.

Some robots for language teaching – in particular, several commercial robots currently on the market, as compared to research robots – come with a screen on the front to display images and other content relevant to the lesson. However, in these cases the interaction appears to be highly touchscreen-based; while the robot appears to be simply a more engaging form factor for a computer screen or tablet. These interactions may still lack the social aspects that are important for scaffolding language learning.

There exist inexpensive commercial robotic toys that teach reading or language skills (Ants, 2011). However, they lack the intelligence and expressiveness to be conversational partners, and they do not exhibit social cues or language-relevant contingent behavior. However, they do present a soft and appealing aesthetic that children are comfortable with, a goal with the character design of the robot for this thesis. Moreover, plush toys are associated with children's social pretend play, inventing voices for characters and creating narratives. At preschool age, "pretending" with a physical character helps children practice perspective taking, social and communicative skills, and anchors them in their narrative. I propose that more than only "believable" responses from a robot, but also some room for imaginative play with a character ("filling in the gaps"), will evoke rich language behaviors while - if well-designed - turn some technological limitations into assets.

Finally, though there does not appear to be current research on their use for teaching preschool aged children languages, telepresence robots are another proposed approach. In contrast to autonomous robots, in this approach the robot's form is a proxy in the world for a remote teacher.

Ultimately, a multiplicity of approaches can address a variety of language learning needs. As Chang et al. suggest (Chang, Lee, Chao, Wang, & Chen, 2010), robots can play different roles – from teacher's assistant to child's pet to supportive cheerleader – for different language learning skills and populations.

Within this space, this thesis seeks specifically to address preschool, pre-literacy aged children for the reasons discussed in the introduction: this is a high-impact age that has a unique need specifically for social language learning. This age is under-addressed in current robot language instruction research, and in language learning technologies in general.

The studies on language learning from television show that preschool aged children can learn vocabulary words, but not grammar. Those require an interactive, social context with rich feedback. The RUBI study established that preschoolers can retain vocabulary words introduced by a robot and that preschoolers can adopt a robot as a social participant in a classroom, but did not yet specifically show the categories of language learning that cannot be learned from television, nor looked at supporting language use beyond object vocabulary learning.

The approach I take is to explore the dynamics of the conversation between children and a social robot. In contrast to on-screen games presented by the robot for children to play, I explore bringing the digital content into the physical space children and the robot share, with the goal of expanding the conversation beyond only the names of the objects. The robot is modeled as a peer and playmate, rather than a teacher, so that children have more of an opportunity to drive the interaction, The robot is contextualized as a pretend-play character with support from a shared activity and props, with the premise that the context of imaginative play will yield richer language use. Also in contrast to prior work, the robot is modeled as a native speaker of only the second language rather than combining vocabulary words with English, exploring an experience more analogous to language learning in a real–world context.

Finally, unlike prior work, this research includes children's caretakers and siblings as integral social participants in the technology, seeking to foster the interactions needed for language learning within children's everyday context.

Establishing a Shared Context by Blending Digital and Physical

Imagine if a robot-playmate that spoke another language could magic into being any object to play with. Children could practice speaking that language while having an ever-growing set of things to discuss and play with, adapting as they mastered the content. Because the objects would be shared and manipulable by both child and robot, the language could move beyond simple object recognition and be about a shared, grounded activity.

Digital content is flexible, adaptable, can be easily replicated, and is becoming increasingly ubiquitous in education. This is even more true now than when computers were new: today parents routinely hand their smartphones to their "digital native" young children and many schools are beyond infrastructure and maintenance problems and are experimenting with novel uses of mobile and desktop devices.

However, on the devices that act as portals to digital content - including mobile phones, tablets, and personal computers - all the action happens behind a screen on which the user's eyes are fixed. Contrast this to the gestures with props, movements in space, and orientation towards each other's expressions and actions of children playing pretend games. Immersion in a digital space brings benefits of its own, but for the educational goal of language instruction, I target the skills of interacting with other people in real time within a changing physical context full of objects of co-created significance. Shared meaning requires understanding of the other's context, which includes emotional and social states communicated through social cues as well as references to shared physical space and objects and the negotiation of shared attention. All of these things require looking up from the screen.

The company Toca Boca makes multi-user mobile apps designed for pretend play, such as a kitchen game and tea party game, which teachers have anecdotally documented are used by children in shared pretend play (McCrane). In this work, I extend the concept and transform the form factor of digital devices so that the affordances match the type of play I am encouraging and in particular, seek to help resolve the split in attention between digital content and playmates.

Socially Situated Digital Play Props

How do we leverage the flexibility of digital content while maintaining awareness of social interaction and physical context? The approach I took was to turn digital objects into play props. My goal was to create a shared context for the robot and one or more children playing with it that included digital elements and thus could be easily built on, but that focused attention naturally on social interactions with fellow playmates in a physical space. Further, creating a digital "shared context" for robot and child would reduce the technical challenges of the robot sensing the world. Physical toys and gestures in shared space require advanced sensing capabilities and new programming for new objects, but the computer knows the state of "virtual toys" by definition.

Research by Cati Vaucelle examined real video recording functionality built into video cameras for dolls, so that children could document their stories without interrupting their narrative (Vaucelle & Ishii, 2008). In previous work, I embedded remote communication capabilities into dollhouse objects (tiny phones and computers running video chat software) so that children could play at a distance without breaking out of the make-believe context (Freed, 2010). Both of these designs are examples of technological devices built seamlessly into play props so that the play with objects need not change in order to use the features of the technology.

"Sam" was an on-screen character (a "virtual peer") who told stories and shared physical toys with child playmates (Ryokai, Vaucelle, & Cassell, 2002). A child could tell Sam a story while manipulating a dollhouse character, and then give Sam a turn by moving the doll through a door into Sam's "virtual dollhouse" on the other side of the physical dollhouse. Studies with this system showed that children could successfully share a toy with a digital character, and anchor their verbal play in this shared context.

Building on these approaches to create a shared-context activity for robot and child, I will later describe several iterations of a "French café" game in which the digital content was the food the child and robot ordered and ate together. The form factor evolved from a play table with real plates and projected food to a robot-scale table with an embedded touchscreen displaying digital plates and food. In both cases I approached the problem by placing the digital content on a surface visually and spatially congruent with a play-context-appropriate physical prop.

Choice of Language

I chose to test with the French language for the following reasons: it is a relatively unlikely language for children to have already begun learning at preschool age in the Boston area, it contains enough cognates to study their impact, it is well supported by existing speech recognition and synthesis systems, and it is a language I am familiar with.

Family Participation

Throughout the design process, pilot studies, and the final evaluation of the technology, children in the target age range were invited to participate with their parents and siblings, and I analyzed the behavior of all participants. I argued earlier that a "human teacher," as opposed to non-interactive, non-social media, can supply the necessary context, modeling, and feedback to a child acquiring language. Specifically, young children's primary language teachers are their caretakers; parent behaviors are critical to the process this technology aims to grow on. Moreover, designing for whole-family participation takes the use of social technology further: not only could it make possible scalable learning that is not possible within a standard technological context, but could potentially do so whilst fostering positive social interactions between children and their families.

III. Explorations and Design Insights

The following experiments and explorations were conducted alongside the more direct development of the study specifically about social robots and preschoolers, and influenced my approach and ideas.

First, I developed and compared a personalized "learning experience" designed to create language immersion, and a more traditional language practice game.

I also developed a project examining performative challenges of communication. It featured a physical character that helped the user modulate their voice while reading a story.

I discuss explorations with simple speech recognition and multimedia content that led to the development of the final shared activity for the robot and child.

Finally, I describe iteratively designing and evaluating the shared activity between robot, child, and other social participants. These include hybrid digital/physical props created to help establish a flexible shared context.

The robot platform used in the final study was developed by another student in parallel with these experiments. In the spirit of "prototyping early and often" when refining the interaction, I tested with as much of the robot as was complete at each point in time. First, a plush toy with an embedded speaker stood in for the robot; next, I used a robot with only facial expressivity. Finally, I was able to test with the complete platform and programmed in the full range of social behaviors appropriate to the shared activity for the final study.

Personalized Language Immersion Experience

Half of my family is from France, and my relatives do not speak English. Growing up, I spent most summers visiting them. I also spent my 8th grade year in an all-French school when my mother moved back to her home country. I am fascinated by the feeling of immersive language experiences. It can be quite embarrassing to be less comfortable in a language than the people you are surrounded by, but there is also something about being so deeply focused on communicating that is fascinating and wonderful. I ultimately want to create technology that helps bring these experiences to a wider audience, but to do this I needed to understand: what are the components of an immersive language experience?

Explorations

As part of a course on alternative educational models, I conducted an informal study with adults in order to gather more personal experience about the process of new language instruction. While quite informal and in many ways not directly applicable to language acquisition for preschoolers, what I learned surprised me and influenced both this work and thoughts about future work.

In summary, I created an immersion-learning brunch for a non-French speaking friend. I served authentic French crepes, fruit, and chocolate. In addition to these objects on the table, I put a map and a calendar on the wall, with the plan that we could discuss an upcoming trip we were taking together. During the experience, I spoke only French while my friend was allowed to speak her choice of English or French. My learner initially said she would be embarrassed to practice speaking a language she didn't know. I was also nervous about leading the experience.

The experience lasted two hours and ended with both of us mentally exhausted, exhilarated, and surprised. We unexpectedly turned an initially awkward and socially challenging situation into a high-flow state for both of us.

I followed this experience with a more traditional drill-and-practice game I designed, simplifying the words and amount of vocabulary based on what I had observed was difficult in the first experience and building this into a matching game with tiles for two French learners.

Observations

While the vocabulary was indeed visibly easier to learn in the drill-and-practice game, and the game was engaging, something important was lost in the process. In a follow-up interview, my friend described it this way:

> *The experience at your house more felt real, it felt like learning a language. It was immersive, it felt like being in a play. It was more authentic. The game was more classroom-y. It was really fun and I probably learned more words from it. But in the long term, I think I might learn more from the immersive one.*

This experiment revived my own memories of the social nature of language learning. It reinforced the idea that physical, manipulable props can help establish shared meaning and teach more language than the names of the objects themselves. It guided me to focus on imparting the structure and rhythm of sentences in another language and treat vocabulary as a secondary component. This meant that the rhythm, pace, and prosody of the interaction would be critical areas for development. I also observed that *learning how to ask for clarification* in the new language was critical; as the teacher, I needed feedback about what the learner did and did not understand that I would not have had received had she not been able to speak English during the interaction.

Finally, I learned in a concrete way that making a learning experience "easier" can in some cases result in a loss of engagement, motivation, and flow – at least in adults. These observations reinforced my interest in modeling a robot as a native speaker of the foreign language that does not understand the child's language, rather than having it introduce foreign vocabulary words through the child's language.

Key Insights

1. In the long term, authenticity and immersion may be more important than immediate vocabulary learning.
2. Physical props can be seeds for conversation, beyond the names of the objects themselves.

3. The learner needs to be able to give feedback to the teacher when they do not understand.
4. The learner does not need to understand all the words immediately to make meaning.

Performative Speech and Sock Puppets

This project was motivated by my interest in helping people become more comfortable taking performative social risks, including practicing a new language but also public speaking, improvisation, and children's imaginative play. The goal was to create a tool to help people be playful while practicing modulating their voice.

The original inspiration was a young girl who asked me to "do the character voices" for the dollhouse game we were playing, because she "didn't know how" – and the realization that I couldn't do them either. A moment of success was when I found myself reading a book in a silly high-pitched voice while animating a sock puppet of a frog in a labcoat during a presentation, something I previously would not have thought possible given my own performance anxiety.

Explorations

I designed handpuppets augmented with hardware and software so that changes in the mouth movements could control playback of a voice recording from the user. There were two puppets (a frog with a high pitched voice and a whale with a low voice) and depending on which puppet was chosen, the pitch of the voice was different. The amplitude of the mouth movements controlled the volume and the speed controlled the speed at which the voice was played back.

Figure 1: Voice modulation play/practice tool with puppets

This design, shown in Figure 1, allowed real time exploration of what a change in the user's voice would sound like, but did not bridge to helping practice modulate his or her own voice. So I designed another interaction. I placed a speaker inside the frog puppet and connected it to an Android tablet. I then augmented the children's story "Five Little Monkeys," a counting-down story with multiple characters, with pages where the frog puppet (in the role and costume of the "doctor" in the story, as shown in Figure 2) would read the appropriate lines.

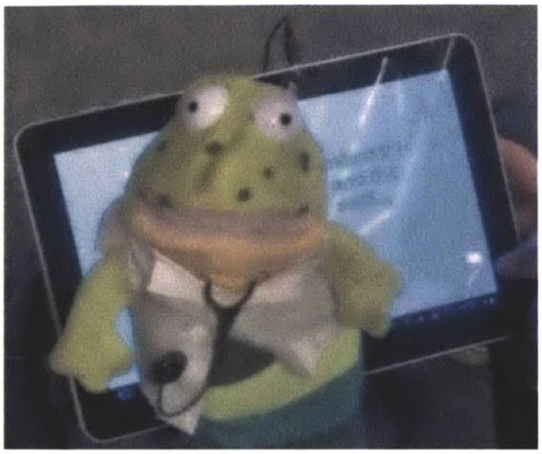


Figure 2: "Doctor Frog" hand puppet and interactive storybook on Android tablet

The storybook recorded the reader's voice and searched for an audio recording closely matching prosodic properties of the user's voice such as pitch and speech rate. If the user read "Mama called the doctor and the doctor said..." in a slow, low-pitched voice, the doctor puppet would say its next line: "No more monkeys jumping on the bed!" in a slow, low-pitched voice. Normal reading triggered a more ordinary voice, with the goal of motivating vocal exploration.

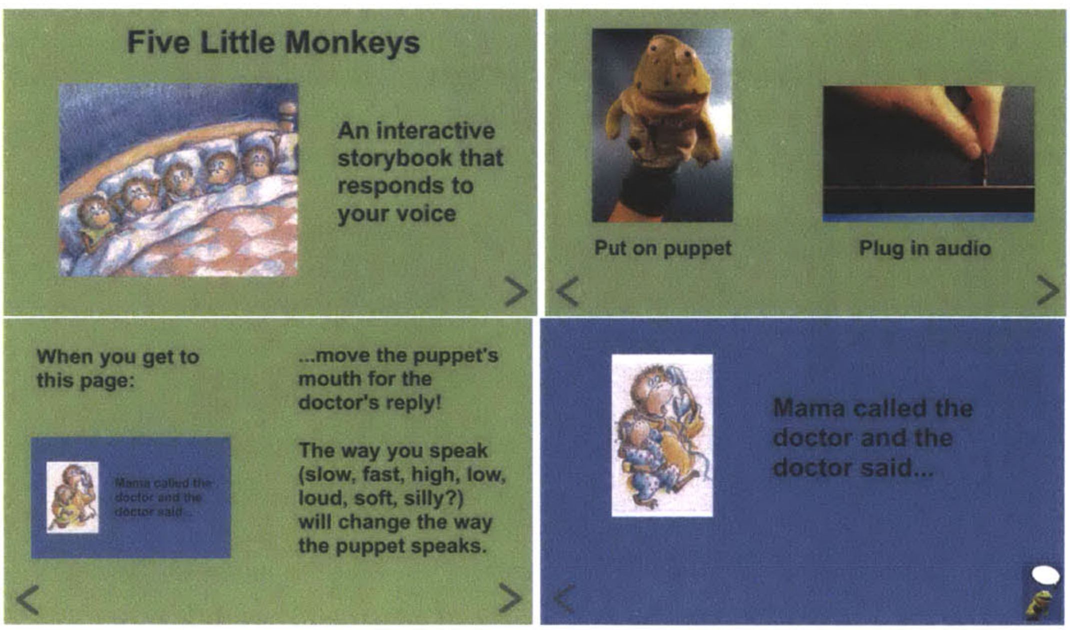


Figure 3: Five Little Monkeys interactive storybook instruction pages

Observations

I brought the "Five Little Monkeys" book and its character to a preschool classroom and found that the children were amused by the story and playing with the puppet, but that the most significant benefit from my perspective was the increased ease of speaking in a silly voice when I had a prop to essentially give me an excuse to perform.

This project suggests that there is value simply in making it possible for parents to comfortably help their kids learn. Both parents who don't speak English and second-generation parents who don't speak their other language with each other in the home but want to pass it on to their kids might be able to use an interactive character as a mediator, in ways that don't feel comfortable or natural to do alone.

A design goal of this project was that of helping people be silly and playful when practicing. Performance and speech is quite an anxiety-provoking experience for many people. And modulating pitch, timing, loudness, etc. are things that there is no "right" way to do. Rather than having someone practice voice modulation by trying to immediately perfect it, you want them to begin by playing around with their voice and its limits. They should try shouting and whispering. They should try extremely silly high-pitched and low-pitched voices. They should try exaggeratedly slow and impossibly fast timing, and they should experience what it feels like to create extreme binary variations in range versus speaking in as many different ways as possible. I believe this type of exploration would also be beneficial for practicing pronunciation and accent in another language – which again, might be best learned through freedom to vocally explore.

This project highlighted the need to design socially supportive language practice tools not only for children directly, but to help their parents and teachers engage in social behaviors that will in turn help the children learn.

Key Insights

1. Interfaces can be mediators for communicating, bypassing shyness and performance embarrassment.
2. There is value in empowering children's caretakers to engage in better teaching behaviors, in addition to impacting the children's learning behaviors directly.
3. Silliness can both defuse anxiety and encourage full-range exploration.

Le Fonduephone

This exploration was about the following ideas:

1. How might one establish a shared context for play that is digital, and therefore can be adapted easily, but also maintains focus on social cues and not only a screen?
2. What is a play pattern that supports conversation anchored in a specific set of objects and actions but is still open-ended, and how might the technology's form prompt this pattern?

Explorations

Playing "tea party" or eating food together is a pretend play game that has a more defined set of actions than playing with dolls, yet still allows for open-ended play and conversation.

I designed a dining experience in which virtual food was projected onto physical plates, preserving some of the benefits of tangible physical objects and some of the flexibility and ease of sensing by the robot of the digital content. This experience will be referred to as *Le Fonduephone* – "fondue" for the chocolate fondue virtually served, and "phone," the Greek root for speech or voice.

In *Le Fonduephone*, the player would sit at a child-sized café table with the robot or plush character. Around them would be projected an immersive virtual environment resembling an outdoor French café, as shown in Figure 5. On the table, a projector would display images of fruit onto physical plates. The child could use a special tracked fork to pick up and take a bite of the virtual fruit, or eat it completely by taking multiple bites. They could also dip the fruit in chocolate by plunging the fork into a bowl of dark, white, or milk chocolate and pulling out a fruit covered in dripping chocolate of the appropriate color.

Figure 4: 3-year old playing in *Le Fonduephone* projected environment

A few brief notes on implementation follow. Both the teacher panel and the projected fruit application were implemented in Java and Processing. The trackable fork (Figure 7) was made of lasercut plywood, conductive paint in engraved channels for the circuitry, an infrared LED, and a coin cell battery. The bright point created by the LED was tracked by a webcam with a visible light-blocking filter constructed from the inside of a floppy disk. The immersive environment was created by 7 stitched projectors on the floor and wall of a scaffolded structure (this part of the system was already in existence at the time of this project and was built by collaborators). The fruit were projected by a projector mounted on scaffolding above the café table and aligned with the camera image so that they could be accurately moved by the fork. Figure 5 depicts the layout of the projected environment. Figure 6 shows the play table with projected food and physical props, and Figure 7 shows the fork prop with embedded tracking circuitry.

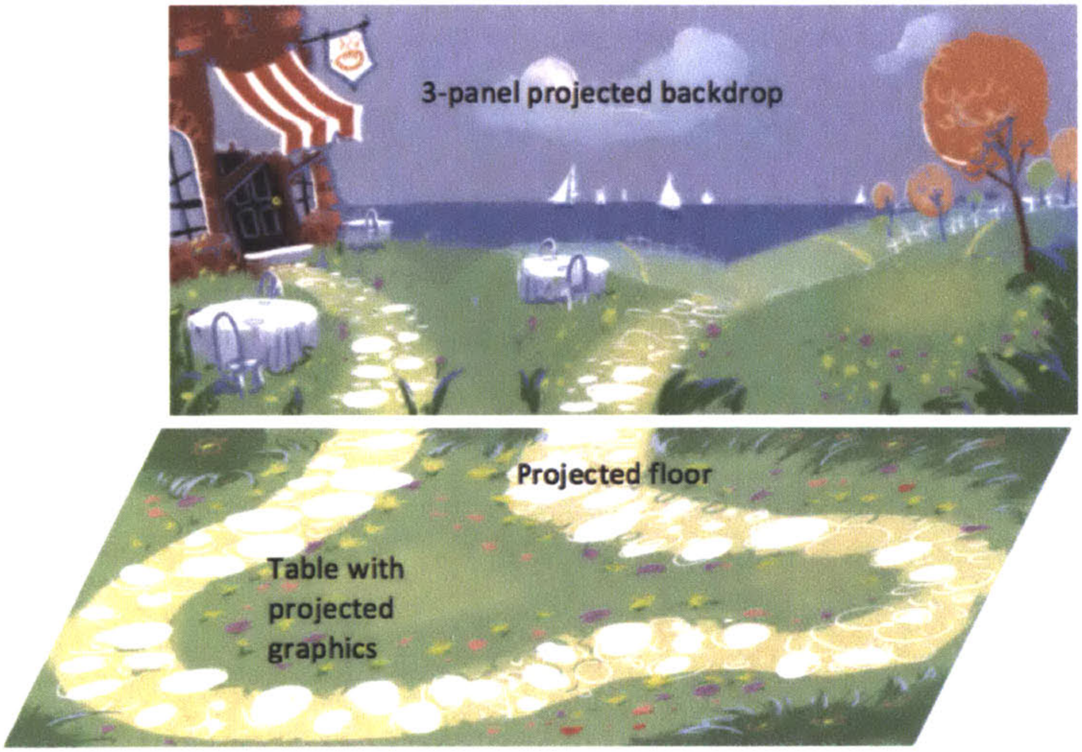


Figure 5: Diagram of virtual environment with projected graphics

Figure 6: Play table with physical plates, trackable fork, and projected digital fruit and chocolate dip.

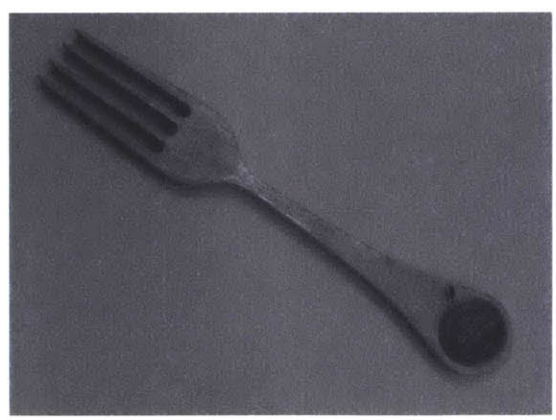


Figure 7: Trackable fork that children could use to move and eat projected fruit

In this activity, the child could order a fruit by asking the robot how to say the name in French, then repeating it as though calling out an order. When a piece of fruit was successfully ordered, a projected image of the fruit would move from the serving plate to the player's plate. The player could then use the fork to "pick up" the virtual fruit and dip it into the chocolate or eat it, triggering sound effects and changes in the projected fruit. Figure 8 depicts one such interaction.

a. Initial table state

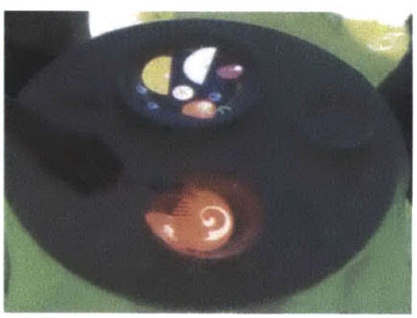
b. Robot:
"Que veux-tu manger?"
(*What do you want to eat?*)

c. Reply:
"Une fraise, s'il te plait."
(*a strawberry, please*)

d. Strawberry appears on plate

e: Taking a bite out of the strawberry by removing it from table (crunch sound)

f. Strawberry now has a bite missing



g. Dipping strawberry in chocolate

h. Eating chocolate covered strawberry (chewing sound)

i. Strawberry has now been "eaten"

Figure 8: Ordering a strawberry. Conversational module 5 (see Appendix 1: *Conversational Modules*) and interaction with virtual fruit through trackable fork.

The fruit available were the following:

1. Une pomme (an apple)
2. Une banane (a banana)
3. Une fraise (a strawberry)
4. Une orange (an orange)
5. Un raisin (a grape)
6. Une myrtille (a blueberry)

I chose these to include two cognates ("banane" and "orange"), one single-syllable and easier to pronounce ("pomme"), and three more difficult to pronounce.

The interaction was controlled behind the scenes by a "teleoperator interface" that allowed the operator to choose what the robot would say, confirm what they thought the child had said, view a record of the conversation, and reward the child's utterances with fruit appearances.

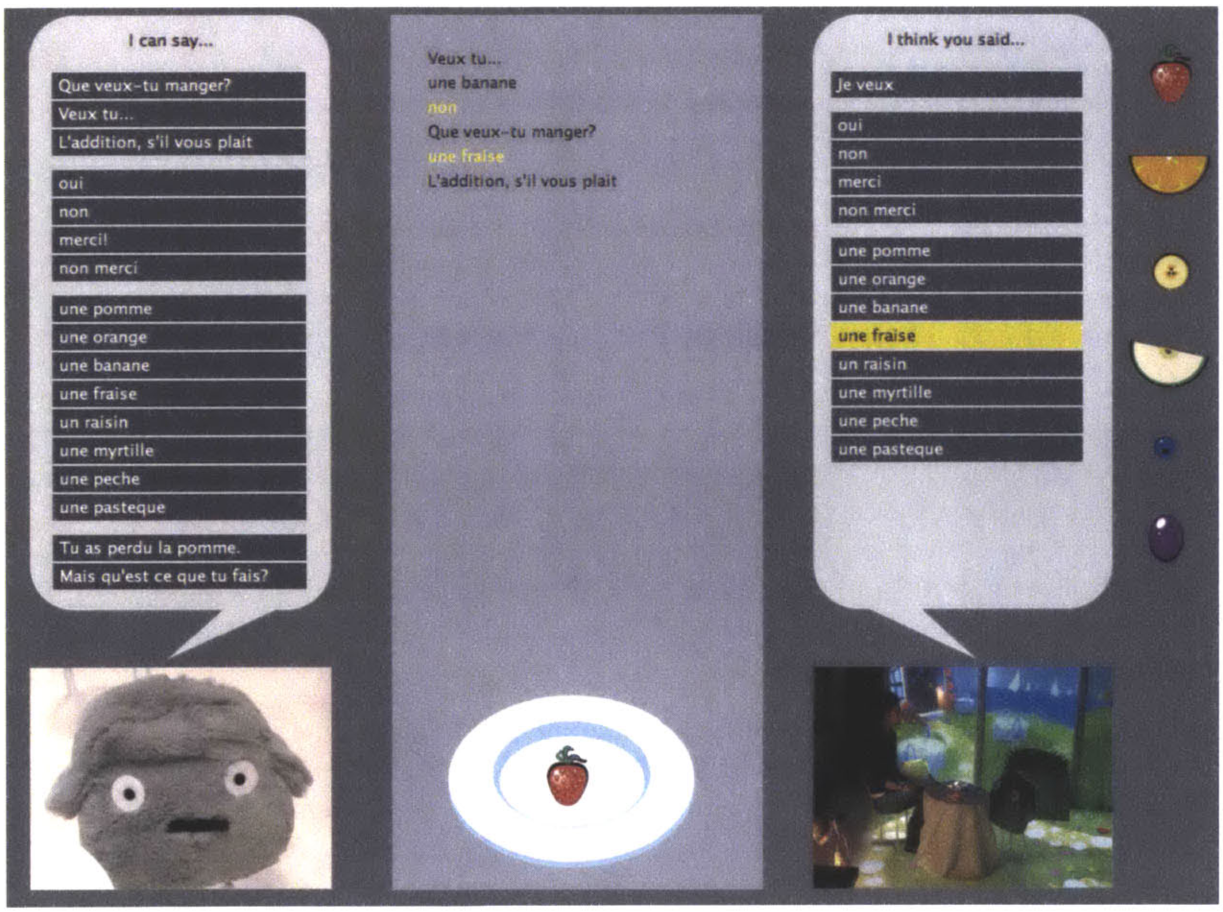


Figure 9: Teleoperator Interface for Le Fonduephone

Observations

Several groups of preschool-aged children (3-5 years old) visited and played with the system with their families.

Figure 10: Preschoolers and older siblings playing with Le Fonduephone and different iterations of the French-speaking robot.

An initial observation was that it required adult scaffolding to teach them to say the word to the robot. The speaker sound coming out of the robot may not have been clear enough for them to understand, and the system does not explain itself. Challenges for future robot design will include getting their attention, clear feedback that it is listening and what they said was heard (even if not understood), and modeling the pronunciation clearly enough to repeat. One parent suggested recording each utterance multiple times, because hearing the same phrase repeated differently would help the listener filter for the sound of the word.

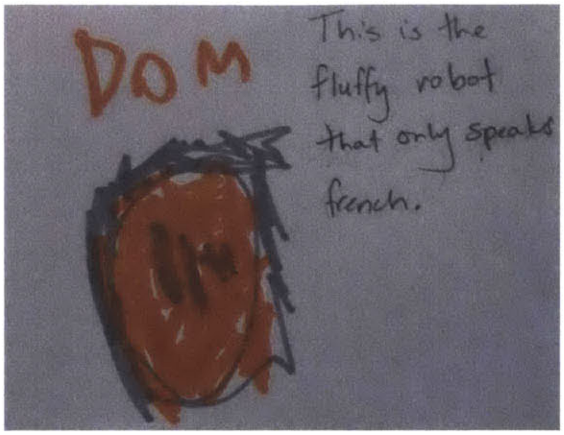
Many children who visited immediately sat down at the small table and started pretending to eat the virtual food, including the fruit they could not move with the trackable fork. Children who were not holding the fork also mimed eating the food even though it did not disappear as they did. The particular physical props of this system appeared to successfully support a pretend play setting without any adult prompting. However, they argued over turns to use the fork. I had not anticipated having the system work for multiple children at once, but it was clearly engaging for the children to be able to interact with each other and frustrating to wait for control of the shared objects.

Several children independently discovered that not only could they eat the digital food but could move the fork towards the robot to feed it, a part of the tea party play pattern that I had not identified ahead of time. This game actually appeared to interest them even more than eating the digital fruit themselves.

Another user interface detail is that the fork was used in an unexpected way. Several, including the 3-year old boy observed individually, wanted to "stab" with the fork rather than holding it flat (which is what the system could recognize). A future requirement for the system is to be able to recognize the fork in any orientation.

I also observed that the children constantly rearranged the plates on the table, exploring everything with their hands. When they move the plates, the fruit currently stays where it was projected, not moving with the plate. In future, an improvement would be to track the plates and make the fruit follow.

Several of the preschool children sent drawings back, showing that they remembered the experience. These are shown in Figure 11.



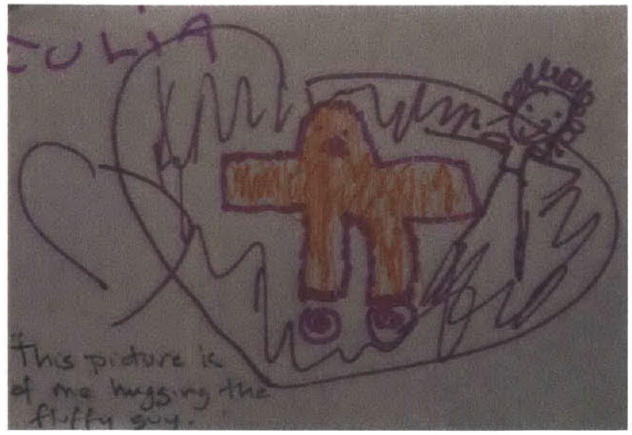


Figure 11: Drawings from preschooler visit.

Left: "This is the fluffy robot that only speaks French."

Right: "This picture is of me hugging the fluffy guy."

Key Insights

1. Familiar play props and setting ("tea party") helps children identify a play pattern and make sense of the integrated digital objects.
2. The technology needs to support multiple children playing with the digital objects.
3. Adult scaffolding is critical, and needs to either be naturally incorporated into the system, or the system needs to do its own scaffolding.
4. Children love to "feed" the robot!

Creating a Portable "Shared Play Context"

With its integration into a children's play set, "Le Fonduephone" appeared to support similar play behaviors as a non-digital game and to allow multiple children to share an activity with the robot and each other.

Could this be achieved with commercially available hardware in a low cost and portable form factor? I explored and tried a few ideas before choosing one for the final robot study.

Projector Boat

The first designs were of a small table with built in seats for two children and the robot, into which the projector and all the sensors for data collection would be built in. To support more flexible play than simply eating at a table, the design was envisioned as a child-sized play boat.

Figure 12: Early prototypes for a "portable play table" with embedded sensors

These fantasy tables and boat ideas did not achieve the desired level of portability. They could be moved from the lab in a car, but not tucked into a bag with the robot and taken to preschools and homes. The final interface, described in the following chapter, was designed for greater portability and lower cost.

The Matching Game

Turning to a tablet device as a more portable shared activity than the projected display, I explored simply building a game on a tablet for the child and robot to play together.

a) "Where is the apple?" Robot requests help. **b) "The apple is there" Child assists robot.** **c) "An apple! I love apples!" Robot reinforces.**

Figure 13: Collaborative Memory/Matching Game design

Play testing a “matching” game on a tablet computer with a 3 year old girl revealed the possibility that this game could be played in solitaire. The young girl had to be reminded frequently to give her father a chance to take a turn. This was partially a result of the nature of the game chosen, which was not an open-ended imaginary play game as *Le Fonduephone* was. However, the form factor tablet itself did not naturally encourage sharing the interface in the way that I had observed with *Le Fonduephone*, which already matched a type of play that requires social interaction. It reminded me that the physical shared context was important, and the game (or shared activity) chosen would influence the conversation and social dynamic. I decided to go back to the “French café” game, but to think about using a tablet device in a new way. The final system combined a tablet with the activity developed in *Le Fonduephone*.

IV. Final System: Social Robot and Shared Context

Having explored my research questions broadly, I narrowed down a final system to evaluate formally. The components of this system and the design decisions are described in this section.

The Robot

The social robot used for the final study was the DragonBot platform designed by Adam Setapen and collaborators (Setapen, 2012). This platform consists of a squash and stretch robot controlled by software running on an Android cell phone, the screen also serving as the face.

Squash and stretch robots can express a range of emotions without the need for many actuators, and reflect a basic principle of animation (Thomas & Johnson, 1981) (Wistort, 2010). In addition, the design creates softer, more organic motions than the visible angle changes in individual joints of a humanoid robot's arm or neck: the entire body changes shape as though it is a soft and flexible mass rather than a mechanical contraption.

Figure 14: Squash and stretch expressive possibilities, mapping posture to valence in this illustration

For the study, I created a custom skin and other aesthetic elements for the Dragonbot "Sophie Neon" shown in Figure 15 below. The name "Sophie" was chosen because it is pronounced very similarly in French and in English, and thus more likely to be parsed as a name in a stream of unfamiliar French.

Figure 15: Sophie the Dragonbot with her table

In the aesthetic and character design, I focused on using materials that engage the senses: soft fur and fabric of different textures, reflective and layered surfaces that highlighted secondary motion. The goal was to create an appealing and memorable embodied character emphasizing the physical world. Figure 16 shows snapshots of the design process, with explorations of tactile and visual texture.

Figure 16: Dragonbot aesthetic hands-on design process (some images are of Sophie, some from other characters in the Dragonbot family)

Software

The software running the system relied on the codebase R1D1, developed by the Personal Robots Group as a general-purpose robotics and cognitive architecture platform (Gray, Hoffman, Adalgeirsson, Berlin, & Breazeal, 2010). On top of this ran a software framework developed specifically to control the Dragonbot (Setapen, 2012). Finally, I developed software for the specific behaviors of the robot in the study and collaboratively designed custom facial expressions and body animations to express "listening to French" and "listening to English" states.

As described above, each Dragonbot contains an embedded Android phone that both runs control software and displays a digital face including eyes and a mouth. This face displays the robot's emotions and the direction of its gaze. For the language learning application, the robot was also given the capability to display visemes, the mouth shapes corresponding to phonemes in speech. The 9 basic phonemes implemented on the simple dragon mouth are shown in **Figure 17** below.

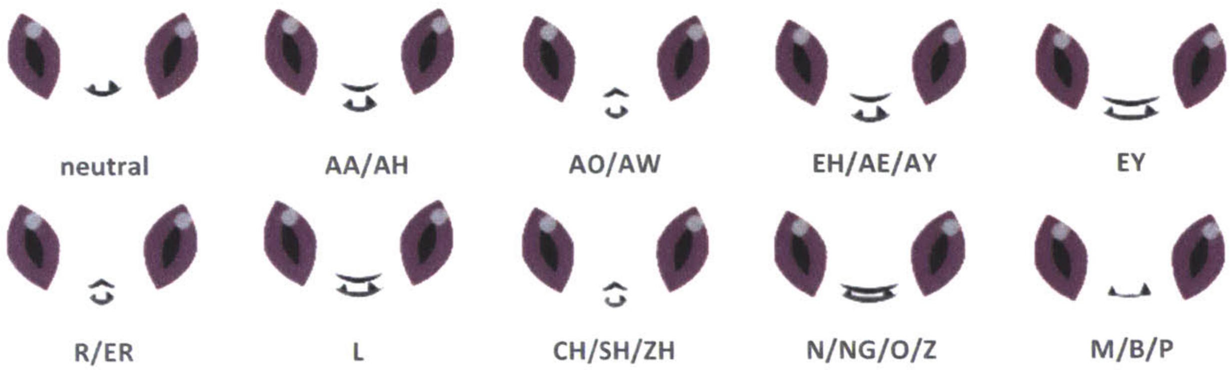


Figure 17: Dragonbot visemes implemented by Fardad Faridi

Play Prop: The Magic Table

In *Le Fonduephone*, I used projections rather than screens because of the difficulty of seamlessly blending two-dimensional digital content behind glass with a physical world. However, screen-based devices have the advantage of being low-cost, scalable, and portable compared to a custom projector setup. Because I wanted to study a technological intervention that would eventually be accessible to educators and families, I used a tablet device design for the final study. I designed a small table for the robot (at the robot's scale, similar to a prop for a stuffed animal toy or large doll) and build the screen of an Android tablet into its surface.

Figure 18: "Magic Table" : Small toy table with embedded Android tablet

Earlier play testing with preschoolers showed that many are comfortable with tablet devices and familiar with both multi-touch interaction and the concept of a menu system. The 3 year old who tried the memory game was very proficient at using a tablet device and accessing the menu to change games – although she occasionally asked her father for help with the fine motor control of picking a specific menu item she wanted. When I tested the sock puppet game, several children also decided they wanted to try another game afterwards and visited the menu to choose one. In order to encourage them to treat the surface and objects as props for play within the context of a social interaction, rather than a multipurpose screen-based device, I designed the table to cover the menu bar and blend as seamlessly as possible into the screen. The background of the display matched the wood grain of the table, and the tablet's "frame" gradually curved in to meet the glass surface of the tablet.

Design of the Game

In designing the game, I sought to preserve what had worked well in Le Fonduephone: the "feeding the robot" and sharing food play pattern and the table and plates prop (in a more portable form factor). The shared activity used the small "magic table" containing an Android tablet, as pictured in Figure 19. The tablet displayed two plates: one for the robot, and one for the child.

Figure 19: Tablet screen in table displaying plates and a fruit; Touching the plate to "order" a fruit.

When the child touched the table and held their finger down for a few seconds, a "magic wand" sound (recording of chimes) would play, and a fruit would appear (one of the six fruit in *Le Fonduephone*).

The child could then eat the fruit by moving it off the plate towards them. The first time, a bite would disappear, and the second time it would disappear completely. Or, the child could offer the fruit to the robot by pushing it to the robot's plate at the other end of the tablet. The robot was programmed to "like" some fruit and not others, and would either express dislike by refusing to eat the fruit, leaning away, and displaying a "disgusted" facial expression, or would "eat" the fruit by leaning in, biting, and chewing it synchronized with the disappearance of the fruit on the plate.

Figure 20: Sophie likes and eats the fruit she is offered.

Figure 21: Sophie does not want the fruit she is offered, leans away as it is pushed towards her.

Because it would take two bites to eat each fruit, the child could also choose to share with the robot; either taking a bite first and then feeding the robot, or finishing a fruit the robot had taken a bite of. Each time a fruit would appear, the robot would exclaim the name of the fruit in French as though excited to see it appear.

The robot had a simple model for "saliency" of objects determining what it should look at. If the child was touching the table to order or feed it a fruit, the robot would be most likely to follow the progress of the fruit with its eyes. Otherwise, the robot would be most likely to look ahead at the child, with the assumption that the child was sitting directly in front of the robot. There were a few mechanical problems with the robot during the sessions, so in some cases the robot's head would lean to the side and its gaze at the child would appear off-center.

Randomly, and not more often than every 10 fruit, a shoe would appear on the table when the child put their finger down to order a fruit. The robot would respond to this event with a downcast and afraid expression. The goal of this event was to learn what children would do when they were the ones "in the know" and the robot needed an explanation and reassurance.

Figure 22: The shoe appears. Sophie reacts with sad/afraid expression and leans away from table.

V. The Final Study: Sophie and the Magic Table

Following iterations on the interaction and study design, I ran a controlled study with 16 participants between the ages of 3 and 6, their parents, and in some cases their older or younger siblings.

Study Description

I designed a study to learn if feedback from the robot could influence the child's use of language towards it. In the experimental condition, it was presented as a completely French-speaking robot that non-verbally expressed interest whenever the child spoke to it – confused interest if the child spoke English, as though it was listening hard but did not understand, and pleased interest if the child spoke French to it. In the control condition, it engaged in the same shared activity but did not give the child feedback when it was spoken to.

The primary research question was about children's use of language during the interaction with the robot in response to its behaviors. Would they treat the robot as a conversational partner? Would they try to speak to it in French, or use other strategies to communicate? Would they be more likely to speak to it in the "feedback" condition? A secondary question was whether children would learn language from the robot by listening to and/or using its utterances.

I was also interested in observing parents' participation throughout the interaction. Would they speak to the robot directly, modeling treating it as a social agent? Would they "play pretend" as though they were playing with dolls with their child? In what ways would they scaffold the child's social behaviors and speech? Would the "play pattern" of eating together support parent's participation?

The hypotheses were as follows:

1. Children will speak more to the robot in the "feedback" condition.
2. Children will learn more of the fruit names in the "feedback" condition.
3. Children will mimic the robot's utterances.
4. Parents will treat the robot as a social agent and engage both in modeling (teaching by example) and direct scaffolding of their children's communication with the robot.

More open-ended questions included:

1. Will children treat the digital objects as part of physical shared space between them and the robot?
2. What will children do and say when the robot is confused and needs guidance (when the shoe appears?) Will they explain to it what is happening?
3. What actions of the robot's will trigger children speaking to it?
4. How much will the robot be treated as an autonomous agent, and how much will children "fill in" its behaviors and speech as though playing with a stuffed animal? Do different actions and non-actions of the robot's trigger different levels of each?
5. Will children understand the explanation that the robot speaks only French, and will they try to speak to the robot in French?

6. What kind of modeling, scaffolding, and/or reinforcement will parents engage in?

Overall, this study was designed to learn how preschoolers would communicate with a robot that spoke only a different language, a question that has not been addressed by prior research and, if successful, would signify language learning in a more natural context. It was also designed to investigate whether specific types of communicative behaviors – expressing non-understanding, expressing interest and positive feedback, and expressing surprise and fear at an unknown object – would prompt particular communicative behaviors in the children. Finally, it sought to observe how parents and other family members would participate in the interaction, in particular with behaviors that could support the children's learning.

Conditions

Two conditions were randomly assigned to each participant: the "feedback" condition in which the robot expressed interest whenever spoken to (differently for French and English), and the "no-feedback" condition in which the robot did not react when spoken to.

Figure 23: Sophie reacts to hearing French spoken to her by learning in, saying "hmm!," and smiling.

Figure 24: Sophie reacts to hearing English spoken to her by learning towards the child, making "hmm?" sounds, and expressing that she does not understand.

The final software was evaluated with 10 children in the feedback condition and 6 in the no-feedback condition. These were assigned randomly by the teleoperator at the start of each session via a coin flip; with the small total sample size, it is assumed to be coincidence that a larger number of children were assigned the feedback condition.

Introduction and Pre/Post Testing

As the facilitator, I administered a pre test measuring expressive and receptive knowledge of vocabulary and asking the child about his or her favorite fruit. I was blind insofar as was possible (since occasionally I was called in to answer a question) to both the experimental condition and the set of fruit shown to each child during the session.

The pre and post test (Appendix 2: *Pre-Test of Receptive and Expressive Vocabulary* and Appendix 3: *Post-Test of Receptive and Expressive Vocabulary*) included six common fruit as in *Le Fonduephone*: apple, orange, banana, strawberry, blueberry, and grape. The participant was tested on receptive and expressive knowledge of all six fruit names in French.

Some fruit images may have been ambiguous (the grapes were described by some children as cherries), but this would not affect measurement of vocabulary learning in the context of the study since the same images were used in both in the tests as in the interaction itself. The fruit names that are cognates (banana and orange) were not ambiguous, based on the children's answers when I asked them to name the fruit in English.

In pilot tests as I was finalizing the study design, I found that it was difficult to get the children to focus on answering questions, particularly after playing with the robot. I made some rubber stamps for them to stamp directly onto the answer sheets to make the testing more participatory.

Figure 25: Custom fruit stamps for pre and post tests. Left photograph shows the images the children saw, right photograph shows the bottom of the stamps. Images on top of the stamps matched the images of fruit that appeared on the "magic table". The apple stamp was changed to a slice rather than a whole apple for the study (not shown here), and an additional "I don't know" stamp with a blank top was added.

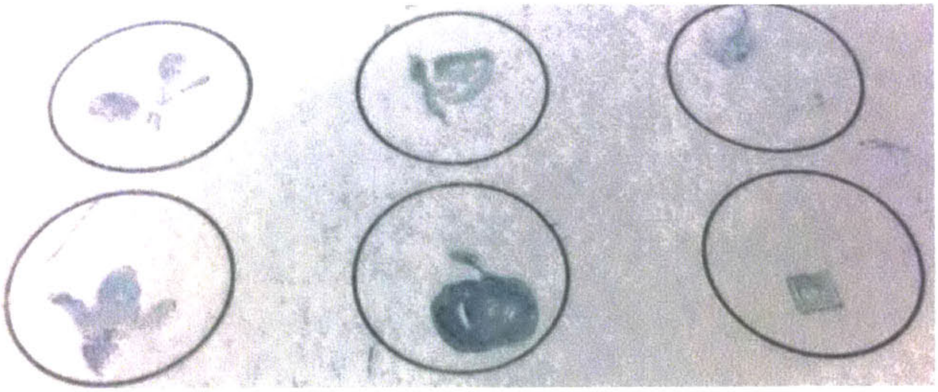


Figure 26: Answer sheet with stamped responses, including "I don't know" (square with a question mark)

After the pre-test, I gave the parent a one-page description to read (

Appendix 5: Instruction Sheet for Parents) introducing the robot as Sophie, a shy robot who only speaks French but loves to be talked to, and describing how to interact with the table. I explicitly described the robot as a social agent in this narrative ("Sophie likes some fruit…"). I gave the parents 10 minutes alone with their child and this written description to read while I readied the robot. I then showed the child and parent to the area where the robot resided, then left the room.

Some of the participants came with older or younger siblings or a second parent. In these cases, I told the families it was fine if everyone participated, although I did not pre and post test children not in the target age range. Where relevant, data from these sessions was excluded from the analysis, as I will discuss in more detail in the evaluation section.

The participant was exposed to a randomly selected set of five fruit that appeared and were named by the robot during the session. Of these five, the robot "liked" – or was willing to eat – two fruit, and "disliked" three, also randomly selected.

Each play session lasted approximately 10-15 minutes, after which I returned and asked the child to say goodbye to Sophie. Finally, I administered a post-test identical to the pre-test other than the replacement of the question "What is your favorite fruit?" with "What is Sophie's favorite fruit?" (Appendix 3: *Post-Test of Receptive and Expressive Vocabulary*)

While the children were answering the post test questions, I gave the parents a survey about their children's play interests, prior knowledge of French, French in their household, and any other feedback. This survey is included in Appendix 4: *Parent Survey*. In addition, I asked the parents to provide a manipulation check for the feedback/no feedback conditions using the following questions, on a 7-point Likert scale:

1. I felt that the robot was trying to understand what we were saying.
2. It was difficult to speak to the robot.

Participants and their siblings were thanked with an envelope of Dragonbot stickers.

Teleoperator Tasks During Session

The condition (feedback or no feedback) was chosen randomly and known to a teleoperator, but not to the facilitator (myself). The teleoperator monitored the interaction remotely via a video and audio connection. When the parent and child had sat in front of the robot, the teleoperator triggered the "intro" script, in which the robot introduces itself:

> Moi, c'est So-phie. Et toi?

This was chosen to be as simple as possible and to contain the robot's name, which was deliberately selected as a name that is pronounced very similarly in French as in English. The goal was to give children a bridge to recognizing part of the robot's utterance. The context was also intended to help with understanding the introduction: the narrative given to the parents to read prompted the family to introduce themselves, and gave the children the robot's name before they heard the introduction from the robot directly.

The teleoperator's instructions were then to press different buttons on a tablet device whenever the child appeared to be addressing the robot directly in either English or French. In the feedback condition, these buttons would trigger the corresponding "listening" animations on the robot.

Figure 27: Teleoperator's tasks during each session

Environment

The experiment itself took place in a curtained space decorated to be child-friendly. The robot was placed on the floor where it would be at eye level for children and where the table's surface would be easily visible. Small cameras were placed unobtrusively in the environment.

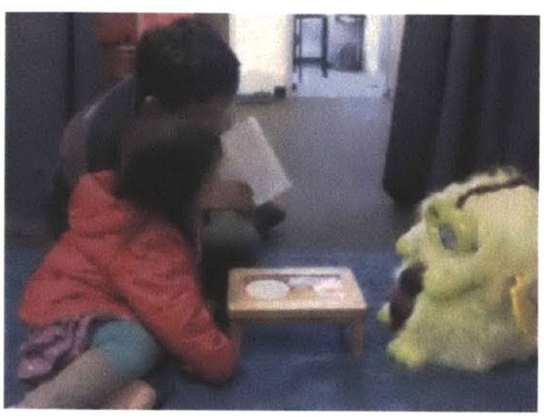


Figure 28: Physical layout of Sophie, table, and participants as viewed from first camera. Second camera was placed behind the robot's head in order to see children's gaze, facial expression, and determine whom they were addressing when speaking.

Recruitment and Participant Demographics

I sent out the following call for participants to my graduate program's mailing list, to a local volunteer list, to the parent mailing list for my university's preschool program, and to acquaintances in the Boston area.

> Call for preschool play testers!
>
> The Personal Robots Group at the MIT Media Lab is designing fun robots to help teach young children foreign languages.
> We're looking for play testers between 4 and 5 years old (both girls and boys) for a study running April 8th - 17th.
>
> If your child is interested in playing with a fuzzy, French-speaking robot dragon and giving us some feedback, please visit the link below to sign up!

The sign-up form asked parents to share why they were interested in the study. As sign-ups came in, several parents gave answers similar to:

> **Generally interested in robotics personally, and I have been getting my kids to be engaged in technology.**

> **My son loves technology and learning about things and how they work.**

and simply:

> **Robots!**

Social robots are even more unfamiliar to the general public than robotics in general, and there is currently a strong association between robots and taking-things-apart play. In order to help clarify this expectation, I added the following line and sent out a second call:

> The study is particularly well-suited for kids who are inclined towards social play (storytelling, imagination games, etc.), as these are social as opposed to construction-play robots.

Although I did not specifically ask parents which message they received, as was the goal the final group of participants reported a wide range of reasons for signing up. These included specific motivations for language learning as well as interest in both the technology itself and its application:

I'm interested because my husband is French and I speak French, but we have had trouble exposing our daughter to enough French so that she is fluent. And I'm generally interested in the topic of language acquisition.

I am a high school language teacher and am interested in learning about other ways to have kids learn languages.

The study sounds like a fun opportunity for my very social-play-oriented daughter to have a chance to play with a neat piece of novel technology.

Confirming that the final group of participants was social pretend play oriented overall, a majority (9/16 participants) reported in the parent survey (Appendix 4: *Parent Survey*) that pretend play was the activity their child spent the most or second most time doing, out of a range of other common play activities.

Perhaps impacted by specifically mentioning French in the announcement, half (8/16) participants reported that the highest French level spoken in their household was 3.5 or greater out of 7, which meant that some of the parents and children knew some French already. In addition to these, at least two participants did not speak English as their native language. In future studies, it might help establish a similar baseline to recruit children who do not speak the target language, and to decide whether to include bilingual children or not. In this study, the variety of participants brought up interesting aspects to investigate further in future studies.

38% of children in the final study (6/16) were male, and 62% (10/16) female. Of the primary participants (not including their siblings who came along), there were 25% 5-year olds (4/16), 6% 6-year olds (1/16), 13% 3-year olds (2/16), and 56% 4-year olds (9/16).

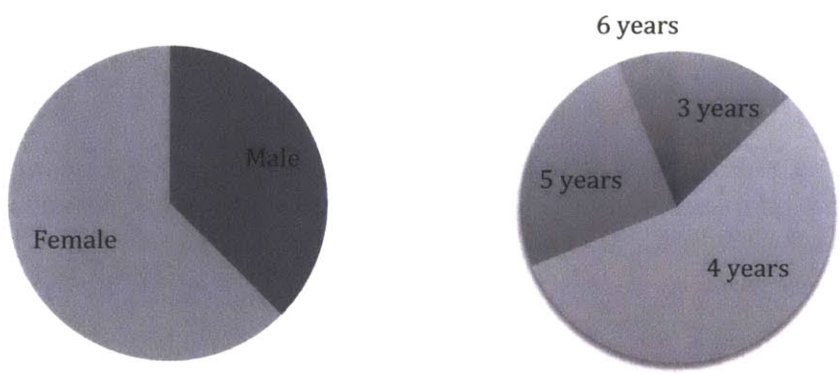


Figure 29: Gender and age distribution of participants

Data on the racial background of participants was not collected.

Results

Video recordings of each session were transcribed to text and annotated by two video coders blind to the purpose and conditions of the experiment. The two coders each analyzed half of the videos, randomly assigned. Inter-coder reliability was not computed, as the coders did not analyze the same video.

Utterance annotations included the following:

1. Who the speech was directed toward (for example, "to Robot").
2. Mention of robot's emotional or social state (eg. "she likes this")
3. Language spoken (English, French, or other)
4. Guiding behavior (for example, parent suggesting to child that she speak to the robot)
5. Reinforcing behavior (parent's praise or enthusiasm)

I first describe findings based on my own observation of the video and transcriptions, then a statistical analysis of the independent-coder-tagged utterances and participant survey and testing data.

Communicating with the Robot When She Doesn't Respond

I observed that there were moments when the children spoke to the robot and expected a response as evidenced by repetition or use of other strategies to evoke a reaction, particularly a verbal one. I propose the following grouping of their strategies:

1. **Repetition**: repeating utterance.

For example, participant 7 repeated a greeting multiple times when she did not receive a response.

> P: Bonjour.
> C1: Bonjour.
> C1: Bonjour, Sophie.
> C1: Bonjour, Sophie!
> C1: Bonjour, Sophie!
> C1: Bonjour, Sophie!

In the dialogue above and those that follow, the parent is denoted by P, the child by C, and the robot by R.

2. **Simplification**: repeating a simplified form of the utterance.

For example, participant 13 repeated one or two words of a longer sentence several times during the interaction in response to the "hearing English" state:

> C: Are you really shy?
> *R: Hmm?*
> C: Shy?

3. **Persuasion**: verbally trying to convince the robot to respond. This strategy is the only one that assumes the robot understood but is voluntarily not replying.

For example, participant 4 continued to convince Sophie to respond to an instruction.

> C1: Eat it.
> C2: Eat it, Sophie.
> ...
> C2: Eat the pear, Sophie.
> C2: Eat that pear. It is what you eat.
> ...
> C2: Eat the pear Sophie please.
> C2: You don't want it?

There are elements of both persuasion and elaboration in the above dialogue.

4. **Elaboration**: rephrasing the content of the utterance in different words.

In an example of parent-prompted elaboration, participant 7 echoed her parent's rephrase of a question when the robot displayed the confused expression and "hmm?" sound of the hearing English state.

> C: Is that good Sophie?
> R: hmm?
> C: Is that good Sophie?
> P: Do you like the orange?
> R: hmm?
> C: Do you like the orange?

5. **Emphasis**: repeating utterance with changes in loudness and/or emphasizing syllables.

For example, one of the two children in session 16 - French speaking twins – spoke louder, slowed his speech, and emphasized the syllables as he repeated what his sister said to the robot.

> C1: Bonjour, Sophie.
> C1: Bonjour, Sophie!
> C2: Bon-jour, So-phie-- Sophie.

The instances I observed did not specifically appear to be the type of prosodic variations used to emphasize the important parts of a sentence as in when adults scaffold children's learning of language, because the entire sentence was emphasized.

6. **Speaking in French**: repeating the content of the utterance in French.

When the child was prompted by the parent to do so, in some cases this was an explicit suggestion to switch to French, in others the parent modeled switching to French and the child repeated it.

In two cases (sessions 8 and 16) there was also an instance of the child trying to speak to the robot in French ("Bonjour") followed by the same greeting in English "(Hello") when the robot did not respond.

This strategy was not possible for all participants because only about half of the parents spoke conversational French.

7. **Mime**: using gesture to communicate the meaning of the original utterance.

In the single instance of this strategy that I observed, participant 13 attempted to ask the robot if she liked a fruit. After using several other verbal communication strategies and receiving the "listening to English" confused response, he pointed at his plate while looking at the robot and said "look here." Then he mimed biting something and made a "chomp" sound.

8. **Touch**: interest in touching the robot directly following an unsuccessful communication.

In three cases, children became interested in touching the robot following attempts to make the robot respond verbally, needing to be pulled back by their parent (who had been asked to wait for the facilitator before touching the robot). It is possible that when they couldn't get through to it through language, they were drawn to touch it to communicate. There could be a number of other reasons for this: perhaps children became less interested in the interaction after receiving no response and sought out more interaction possibilities. There is not enough information in this study to determine this, but I include this category in the analysis for future examination.

I additionally propose the following grouping for these strategies:

1. **Repetition** and **emphasis** are strategies that do not *necessarily* attribute the ability to understand meaning.
2. **Persuasion** assumes the robot understood, but is voluntarily not responding.
3. **Simplification**, **elaboration**, **switching to French**, and **miming** imply the speaker perceives the robot as an agent capable of understanding the *meaning* of language.
4. **Touch** might indicate another means of communication, but there is not enough data to tell what the children's intent was.

In other cases, also excluded from this analysis, children spoke to the robot but seemed satisfied with the response. For instance, participant 1 spoke many times to the robot throughout the interaction about likes and dislikes while offering fruit ("You don't like to eat this, right?"), but appeared to be satisfied with the robot's response confirming this (by eating or not eating the fruit). Here I specifically examine communication strategies the children tried when they were not satisfied with the robot's acknowledgment of their attempted communication as evidenced by continued communication of the same content.

Figure 30 shows the frequencies of these communication strategies by condition, counted if used by the child at least once during the session.

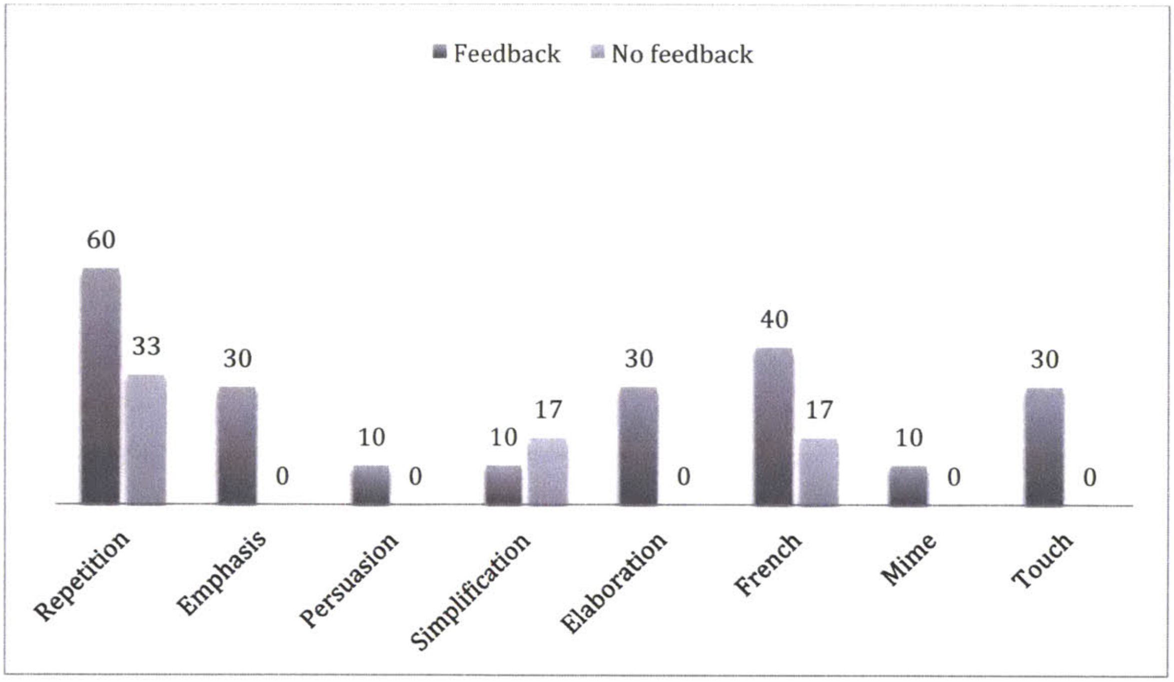


Figure 30: Percentage of children in each condition using re-communication strategies

Table 1 shows p values calculated for each communication strategy and meaningful groupings of the strategies using Fisher's exact test. Statistically significant results using alpha=0.05 are bolded.

	repetition	emphasis	persuasion	simplification	elaboration	mime	French	touch
F	6 / 10	3 / 10	1 / 10	1 / 10	3 / 10	1 / 10	4 / 10	3 / 10
NF	2 / 6	0 / 6	0 / 6	1 / 6	0 / 6	0 / 6	1 / 6	0 / 6
	p = 0.61	p= 0.25	p= 1	p= 1	p= 0.25	p= 1	p= 0.59	p= 0.25
	Attribute ability to recognize		Attribute ability to understand					Other means
	p= 0.14		p= 0.12					p= 0.25
	All communication strategies							
	p= 0.01							

Table 1: Frequency distribution of children's use of communication strategies analyzed using Fisher's exact test. F represents "feedback" condition, NF represents "no feedback" condition

Table 1 shows that participants in the "feedback" condition were more likely to use re-communication strategies than participants in the "no-feedback" condition.

While no statistically significant differences were found for *specific* strategies, In all cases except "simplification," a higher percentage of the children in the "feedback" condition used these re-communication strategies than in the "no-feedback" condition.

As an illustration of an especially communication-rich session, following is a longer excerpt from Participant 13's session, demonstrating simplification, elaboration, repetition, and miming.

P: Do you think she likes it?
C: Do you like it?
R: hmm?
C: Like it?
R: hmm?
C: Look here. <Child points to plate while saying this. After robot looks down, child mimes biting something and makes a "chomp" sound>
C: Get it!
R: hmm?
C: Get it.
R: hmm?
C: Get it.
R: hmm?
C: Geeeet it!
R: hmm?
C: Why's she not...? Saying huh?"
P: Well, she speaks French. So she may not understand what you're asking. Now let's see if she wants to eat the strawberry. Look. Do you think she l-- she doesn't like strawberries, huh. Do you want to eat it?
C: Yeah.
P: Yeah, me too.
C: I ate it! <directed at parent>
R: hmm? <(may have been accidentally triggered by teleoperator)>
C: I ate it! <looks at robot before speaking, appears to direct speech at robot>

Figure 31: Participant 13 gesturing at plate and saying "look here" to robot.

A coincidental but seemingly contingent social cue from the robot occurred during this interaction. Because the robot's head position was raised to look at the child in the "hearing English" pose and went back to its neutral pose after the child stopped speaking to point at the plate, the robot appeared to have responded to the child's prompt to "look here."

It is ambiguous whether "get it" was intended to convey "go get the fruit," a continuation of the message the child was trying to communicate, or a frustrated command to "understand it!" In either case, the overall interaction appears to have been one of trying to convey some meaning to the robot. First, he tried simplifying the utterance which would address the possibility that the language was too complex. He did not repeat this strategy when he received another confused response, but switched to another. These actions imply that he thought of the robot as an agent capable of understanding *meaning* (as opposed to recognizing speech) and modified his communication accordingly. They also indicate that he interpreted the robot's "confused" expression as intended, which was to convey a lack of understanding, and maintained motivation to communicate despite this. Finally, the miming gestures may show that he believed the robot could perceive him and their shared spatial context, encompassing the table and the objects on the screen.

Further on in the conversation, the child said "I ate it!" to his parent. Perhaps due to teleoperator error, the robot then expressed the "speaking English" state and said "hmm?" The child looked up at the robot and repeated the utterance while continuing to look at the robot. Because he was not looking at the robot, it appears that the sound alone was enough in this one case to communicate non-understanding.

Overall, these observations suggest that the robot's feedback of "I don't understand you" prompted re-communication by the children, revealing a variety of strategies for communicating with the robot; including some that attributed comprehension of meaning to the robot.

Manipulation Check

Hearing French and Hearing English Reactions

The robot's reaction to hearing French was designed to express that it understood what the child had said, in contrast to the "hearing English" reaction, which was meant to indicate that it did not. As shown in Figures 21 and 22, the reaction to French was an "ooh" mouth expression transitioning to a smile, combined with a lean in as though listening and a "hmm!" sound. The reaction to English also included a lean towards the child, but the facial expression was a "confused" one and the sound was "hmm?"

Based on many instances of children re-explaining (through repetition and other strategies in the *Communicating with the Robot When She Doesn't Understand* section) in both the French and English cases, it appears that these two expressions were both interpreted as the robot not understanding. The actual intention of the "hearing French" expression seemed to be less readable: the "hmm!" sound accompanying it was too close to the "hmm?" sound of the other expression, and perhaps most importantly was not a satisfactory response to a direct question asked in French, supposedly the robot's language, as in Session 16 :

C1: I said French to her for good morning and she was like 'huh'.
F: What do you think she meant by that?
C1: She didn't-- I, I said it louder, she was like 'huh' when I said it louder.
P: That would mean she heard you.
F: What do you think she meant?
C1: What is it.

Parent Survey Questions about Feedback and No Feedback Conditions

Table 2 and Table 3 below show parent responses to the manipulation check questions relating to the two conditions of the experiment, asked following the interaction with the robot. These questions were designed to measure if there was a perceived difference between the conditions.

1. I felt that the robot was trying to understand what we were saying.
2. It was difficult to speak to the robot.

Condition (n)	***Mean***	***Standard Deviation***	t (*a<0.05*)
Feedback (10)	3.40	1.71	-1.94
No feedback (6)	1.83	1.47	

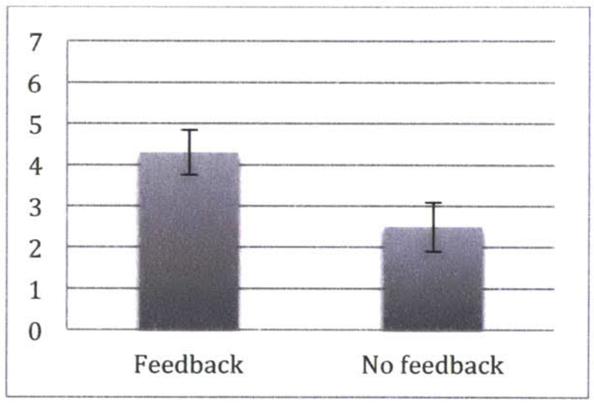


Table 2: Parent Survey responses: "I felt that the robot was trying to understand what we were saying." on 7-point Likert Scale

Condition (n)	***Mean***	***Standard Deviation***	t (*a<0.05*)
Feedback (10)	3.90	1.29	-0.51
No feedback (6)	3.50	1.64	

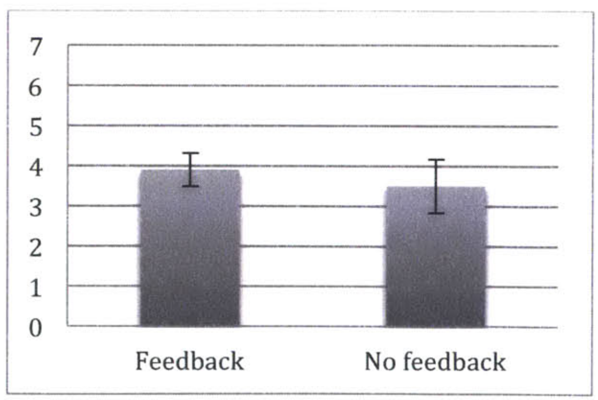


Table 3: Parent Survey responses: "It was difficult to speak to the robot." on 7-point Likert Scale

While not statistically significant, answers to "I felt that the robot was trying to understand what we were saying" trended higher in the feedback condition. Given the small sample size, which made it difficult to achieve statistical significance, it would have been helpful to gain some qualitative information by asking parents why they answered the way they did.

There was little difference between the answers to "it was difficult to speak to the robot." This question was included as a negative phrasing of the previous question to make sure there was not a bias towards a positive answer, a common problem with Likert-type questions. I believe the rephrasing did not address the difference between conditions, but rather a separate question. The answers revealed that on average, participants found it between easy and difficult to speak to the robot. This could have been because they did not understand the question as applied to a robot character, because they found that pretending to speak to a robot character was of medium difficulty (eg. as compared to speaking to a stuffed toy), because the robot gave a medium amount of verbal feedback, or for many other reasons. Here again it would have been helpful to ask parents to explain their answers. Alternately, this question could have been broken down into more specific questions addressing how comfortable the parents were with speaking to the robot, for example "Was it easier to speak to the robot than to one of your child's toys?"

I included both these questions in the survey as an indirect manipulation check that did not present too great a burden on the participants, in particular intending not to overwhelm the children with questions. In future studies, I would prompt parents to ask their children the question and discuss it with me out loud in an unstructured interview format.

Mimicking the Robot

Although this did not appear to influence recall of either receptive or expressive language as measured in the post test, 6/16 participants spontaneously mimicked Sophie's naming of the fruit, repeating the word in French after her (sessions 1, 2, 3, 4, 8, and 11).

Making Sense of the Robot's Behavior

When children and their parents discussed making sense of the robot's behavior, in many cases they gave socially-framed explanations. Some reflected those in the written explanation I gave the parents ("Sophie is a shy little robot..." full text in

Appendix 5: Instruction Sheet for Parents) and some were new descriptions.

Some of the parent's and children's explanations revolved around her movements, in some case paired with an interpretation of the social action or communication, as in this example of a parent's explanation from session 5:

> P: Huh, he's backing off. He doesn't want it.

And a parent in session 3:

> P2: Oh! Does that mean-- when her eye gets big like that, do you think she likes it?

A child offered an explanation in session 4:

C2: If she goes down, that means I think she doesn't like it. <interpreted as describing the robot's body moving away>

Several parents pointed out the robot's gaze, including in session 15:

P: Look, she's way over here looking at you. <Mother sitting to the right of child mimics robot's coincidental lean towards the left, the side that child is on>

...
P: Does she see you?
...
P: Look, she looks at it, see? Watch her, she's looking back and forth. <as child moves object on plate>

Participant 7 built on the robot's expression and added a more detailed narrative about the robot's state:

R: Une pomme!
C: She's chewing.
C: She ate.
C: Now she looks happy and full.

The robot crashed towards the end of a few of the sessions, which rendered it immobile. One child said:

C: Sophie! Sophie, wake up!

Other explanations for the robot's behaviors included "she is just learning to speak French" and "doesn't have too many words," "maybe he's not hungry," "she is asleep."

In session 7, the child was uncomfortable and afraid of robots, and her mother explained that it did not have agency:

P: You see? You say, "hello Sophie". She doesn't do anything, she's a fuzzy toy. That's all she will do.
C: Why is she moving back?
P: You know why she's moving?
C: Why?
P: They are moving her, they're moving her. There's a battery in there somewhere. See? Look at that. Look at that back there, see? They are moving her.

Later, this parent went on to describe Sophie's likes and dislikes:

P: She likes it too! She gonna eat more? Oh!
P: Ooh, she's smiling! You wanna try another one?

There appeared to be a conflict here between wanting to reassure by speaking about the robot as a machine, but also naturally guiding the narrative and play through social explanations.

Parent Behaviors

Parent behaviors consisted primarily of scaffolding and reinforcement: guiding the child in the interaction and giving positive reinforcement, both in a variety of different ways.

Appendix 5: Instruction Sheet for Parents) describes the narrative that was given to each parent to read. Other than this document, no prompt to scaffold any particular behaviors in the child was given by the facilitator.

Following are the behaviors by parents observed during the sessions:

1. modeling speaking to robot
2. modeling speaking to robot in French
3. prompting child to speak to the robot
4. prompting child to speak to robot in French. For instance:

   P: Ok, can you say goodbye?
   C1: Goodbye
   C3: Goodbye
   P: Say- how do you say "bye" in French?
   C2: "Au revoir"
   P: And how do you say "see you later – soon"
   C2: Uh
   P: "A bien"
   P: "A bien"
   C2: "-tot"
   P: "A bientot"

5. modeling speaking about the robot as a social agent
6. technical help about how to use the table
7. guiding social interaction, turn-taking, and sharing ("now feed her another fruit," "alright, now why don't you give yourself something?")
8. reinforcing through praise and/or enthusiasm
9. pointing out the robot's social cues ("she is looking at you")
10. "dialogic" guiding, for eg. asking child to make connections (Session 14: *"why didn't you think she ate it? What do you think?"*)
11. Repeating Sophie's utterances to the child
12. Prompting child to learn the names of the fruit
13. Modeling touching the robot, when they were allowed at the end of each session

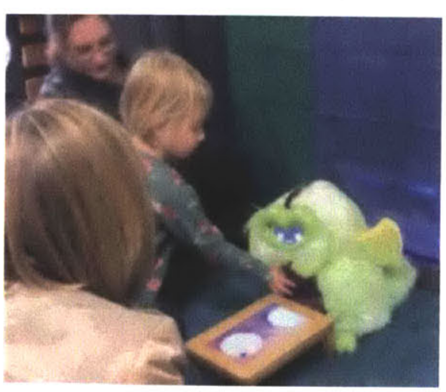


Figure 32: Tactile exploration

I analyzed specific parent behaviors by condition, choosing the following most frequent categories of behavior relevant to language learning:

1. Modeling behaviors
   a. speaking to the robot
   b. speaking French to the robot
   c. treating the robot as a social agent during the shared activity
2. Prompting/scaffolding behaviors
   a. Prompting child to speak to the robot
   b. Prompting child to speak to the robot in French
   c. Prompting social behaviors, eg. turn taking, sharing, responding to robot's communicative actions.
3. Dialogic guiding: asking child to predict, explain, or describe the robot's behaviors.
4. Acting as mediator between child and robot; specifically, repeating and reinforcing robot's actions or utterances to child.

A similar analysis of these behaviors as that used on the children's re-communicative actions during the sessions as described in *Communicating with the Robot When She Doesn't Respond,* as shown in Table 4 below, revealed no statistically significant differences between the conditions. This shows that the robot's "feedback" behavior (communicating that it did not understand) did not specifically affect parent's behaviors.

	Speak to robot	Speak French to robot	Treat robot as social agent	Prompt child to speak to robot	Prompt child to speak French to robot	Social interaction prompts	Dialogic questions	Reinforcing robot's actions
F	6 / 10	1 / 10	10 / 10	5 / 10	2 / 10	8 / 10	7 / 10	7 / 10
NF	2 / 6	1 / 6	5 / 6	2 / 6	1 / 6	5 / 6	1 / 6	5 / 6
	p= 0.61	p = 1	p= 0.38	p= 0.63	p= 1	p= 1	p= 0.12	p= 1
	All supportive behaviors							
	p= 0.21							

Table 4: Frequency distribution of parent supportive behaviors analyzed using Fisher's exact test. F represents "feedback" condition, NF represents "no feedback" condition

The activity itself, beyond the two conditions, was designed to support and leave room for these behaviors. It appears that it successfully did so, as the great majority of parents did engage in some or all of them. Learning if this specific activity elicited these behaviors more than conventional language learning tools, or more than a different activity, would require another study.

In the *Statistical Analysis of Utterances* section, I will describe the finding that children's communicative behaviors were positively correlated with parents' modeling and prompting of those behaviors during the sessions.

The Shoe

The shoe mainly evoked surprise and amusement from children and their parents. Counter to my hypothesis, in no cases did they explain the shoe to the robot. However, it appeared to be a compelling event, provided an opportunity for parent-child discussion, and was revealing of children's perceptions of the table interface.

The parent from session 15 pointed out the robot's expression to the child:

> P: Aw no! The shoe! Look how sad she is.

After her parent helped remove the shoe from the table, one of the children in session 4 said:

> C2: What. Mama you just ate a shoe right now.

This may show that the metaphor of the table and "eating" by pulling objects off the table was accepted, at least within the context of the game framed by the table mechanics. In session 16, one child expressed distress at the idea of eating the shoe, and did not want to move it from the table until her parent helped because this would represent "eating" it.

> C1: I'm not going to eat this shoe.
> F: What does she think of the shoe? She likes the shoe?
> C1: She doesn't eat it, and I don't.
> C1: She's not eating it.
> F: She's not eating it? What are you going to do with it then?
> C1: She shook her head no.
> F: Ah, ok.
> C1: I don't want-- I don't want to eat this shoe either

The shoe did not fit the make-believe context created by the table and the families' discussion, and silliness that did not break the context ("you just ate a shoe right now!") was needed to resolve it. Again, this implied that the make-believe context, in which moving an object off the table represented eating it, was accepted.

Shared Social Context

Instances of sharing and of group participation in the interaction are described below.

Most children fed the robot the fruit it liked, and ate the ones it did not. However, two participants shared bites with the robot, as in session 1:

> C: I eat first, and then you eat first

One child, from session 16, wanted to leave a small bead as a gift for Sophie:

> C2: Ah, can I give this to Sophie?
> F: You wanna give that to her? Uh, sure! Ok.
> P: What is that? You can try. You can see if she wants it.

F: You could maybe put it on her plate.
P: Where'd you get that?
C2: School.
C1: She won't eat that.
C2: Sophie, you can have this. I don't need it, ok?

In three instances where there was an older and younger sibling, and in one case where the siblings were twins, I observed one sibling (older when different ages) helping the other to feed the robot, either by helping move the fruit on the table or by verbal scaffolding. Session 5 shows one such interaction:

C2: This is a little sibling. Yeah. <directed at robot while pointing to younger sister.>
...
C2: <younger sister's name>, wanna have a-- wanna try?

And in session 16, a girl guided her twin brother as he pushed a fruit towards the robot:

C1: No she doesn't like that!
C1: If it goes in there, she's not going to eat it, so you have to put it here, so you can eat it.
C1: If she does, you don't eat it.

The photograph below shows a mother, an older and younger brother's interaction with the robot. First, the older brother helped the younger move the fruit towards the robot, while explaining out loud. Next, the mother moved it forwards, and finally the younger child moved it the rest of the way. When the robot ate the fruit, the child turned towards his mother and exclaimed and gestured in celebration.

Figure 33: Family collaboration to feed robot, followed by celebration.

Observations and Feedback for Further Investigation

Requesting French from parent

In several cases, children asked their parents how to say something specific in French so that they could communicate with the robot. Not retention of vocabulary, but simply motivating children to ask this question might be the hardest part for a bilingual parent wanting to teach their children their language, because the children can already communicate in the language they are more comfortable in. This speaks to language motivation, and to the technology acting as a mediator to make parents' scaffolding easier.

Children with a native language other than English

Two different participants spoke a language with their parents other than English – Hebrew and Urdu. In both cases, the children spoke this language and some English with their parents, and spoke English to the robot. This might have come from the parent modeling the behavior or from the child's guess that a stranger in the United States would speak English. But in these cases they seem to have understood that the robot spoke a *different* language. This may be worth investigating further.

Learning Sophie's likes and dislikes

Most children stopped feeding the robot fruit she had indicated she disliked after a few trials, but a few children continued to offer her every fruit. It is unclear whether this was because they did not learn what she liked, or whether part of their exploration was trying every object.

Audio quality

Many parents commented that the audio was "soft" and they thought their child had trouble hearing it. One parent said that although she spoke French herself, she had difficulty understanding the robot's speech. From the teleoperator station, the robot's speech was easier to hear than the parent and child's speech, so enunciation and expressiveness from the robot to announce it was going to speak may have been the causes of this difficulty understanding more than actual volume. One parent commented that a more attention-getting tone in the robot's voice would be more likely to alert the child that there was something to be listened to.

Learning the digital activity

While most figured it out quickly, several families took a few minutes to understand how to use the table even with the instructions, which took up a large fraction of the total 10 minutes. The primary confusion was how long to touch the plate before fruit appeared. One family initially thought that to "order" meant they needed to say the name of the fruit out loud rather than do it through touching the table.

Parents' feedback about this included:

> **Sophie greeted us first which was promising but because we didn't immediately understand "how to play" (i.e. how to touch the table and that we were playing restaurant, etc.) my child's eagerness to interact with Sophie quickly faded. Hence she became "afraid" because the interaction had failed.**
>
> and
>
> **It was a bit confusing knowing when/how to start the interaction, so we probably used up some of our interaction time. <Child's name> was clearly engaged with Sophie and I think he would have figured out the language bit with a slightly longer interaction.**

While this did not affect many of the participants, In a future study I would let kids and their parents play with the table before they played with the robot, because I was not trying to measure how easy it was to learn to use the table, but rather how engaging the interaction was once they had understood the basic functionality.

Learning

In the pre and post tests, I measured children's expressive and receptive knowledge of the six fruit names. While I was not expecting children to retain the fruit names from this brief, single-exposure interaction, I hypothesized that they would remember some of the names the robot spoke if I tested immediately after the interaction.

In each session, Sophie "liked" two fruit. In the post-test, I asked the question: "what was Sophie's favorite fruit?" which may have led children to answer only even if they would have remembered two. In the data analysis, I used "number correct" as the independent variable. If they gave one answer and it corresponded to one of the two fruit Sophie liked, the number correct was 1. If they gave two answers and they were both correct, the number correct was 2.

There was not a statistically significant difference in the learned fruit between the "feedback" and "no-feedback" conditions, by age, gender, preference for pretend play, comfort speaking to others (as measured in the parent survey), nor was there one overall between the pre and post tests.

Statistical Analysis of Utterances

For analysis of total utterances in each session normalized over time, only sessions in which there were a single parent and a single child were included, because they could not be directly compared to the number of utterances in sessions in which there were a larger number of participants. Simply normalizing over number of participants would not be valid, because a linear relationship between number of participants and number of utterances could not be assumed. This left a total sample size of 9 subjects for this part of the analysis.

It turned out to be difficult to tag parental "reinforcement" consistently because some of it seemed to be expressed implicitly, such as enthusiasm from parent expressed through voice pitch and laughter. A clearer definition and additional video coder training would have been required for accuracy. For this reason, the reinforcement tags were excluded from this analysis.

Correlations between survey responses and parent verbal behaviors, represented as Pearson's r-values, are shown in Table 5 below. For a sample size of 9, an r-value above .666 reveals a statistically significant correlation, here bolded.

	Child utterances /min	**Child utterances to robot / min**	**Child speaks of robot as social agent / min**	**Child utterances in French / min**
Child's comfort speaking	-0.33	0.13	-0.03	-0.17
Child's pretend play	0.21	-0.28	-0.16	-0.37

rank				
Child's age	-0.52	**-0.72**	-0.05	-0.03
Household French level	0.20	0.47	-0.07	-0.11
Child's French level	0.22	**0.75**	0.21	0.03
Parent utterances/min	0.61	0.42	0.58	0.42
Parent utterances to robot / min	0.26	0.26	**0.67**	0.52
Parent speaks of robot as social agent / min	0.32	0.34	0.45	0.38
Parent utterances in French / min	0.42	0.55	**0.87**	**0.90**
Parent guiding U/min	0.64	0.49	0.34	0.29

Table 5: Correlations between survey responses and parent verbal behavior

a) **Children's verbal behavior and survey responses**
b) **Significance of scaffolding: Child and parent behaviors**

From this analysis, it appears that children who spoke more French at home spoke more to the robot overall. One possible explanation for this is that these children had more vocabulary to try that they believed the robot would understand. This could suggest that empowering children with more phrases to try would give them more motivation to try speaking to the robot. Another possible explanation is that French-speaking children had an easier time understanding the idea that the robot spoke a different language. Or, it might simply be that French-speaking children felt more comfortable with a French-speaking robot than children who did not speak its language. If parent reports of their children's overall comfort speaking were accurate, neither household French level (r=0.07) or children's French level (r=0.29) was correlated with their comfort speaking, eliminating the possible explanation that the French-speaking children were more comfortable speaking in general.

It appears that younger children spoke more French to the robot. Given the small range of ages and very small sample size, I believe this would require more investigation. It's possible that the 3-4 year old age range may have been more comfortable with role playing with a character than the 5-6 age range, or less susceptible to shyness in a new social situation.

The above results show that the amount the parent spoke in French was correlated with the amount the child spoke in French. It also appeared that children whose parent spoke more to the robot and children whose parent spoke more in French to the robot treated the robot as a social agent the most. These findings suggest that the parent's role in the interaction, specifically in modeling speaking behaviors, was important. Relating to Meltzoff's findings that infants will treat a robot as a social agent if they observe an adult doing so (Meltzoff, Brooks, Shon, & Rao, 2010), it may be that at this older age scaffolding from an adult helps reinforce and guide specific behaviors towards a social agent. It was significant, as discussed in the Parent Behaviors section, that parents spontaneously engaged in this modeling behavior with no more than a written narrative about the robot as a prompt. If the robot can prompt parents to scaffold their children's communication, and this communication does in fact help children practice and learn, part of the robot's role could be as a mediator for learning interactions.

VI. Conclusion

This thesis sought to design, implement, and evaluate a novel system for language learning using a French-speaking social robot and a shared activity to draw out language practice from preschoolers and encourage learning-supportive behaviors from their families. First, the space of language immersion and socially situated language learning contexts was broadly explored. A play pattern was narrowed down and iterations of technology props were built and play tested, with the goal of developing a shared activity for robot, child, and family. The technology was adapted to a commercially available tablet device to eventually be accessible outside a research lab. A final system was created to evaluate children's communication with the robot, vocabulary retention, and family participation within a "shared meal" play context.

In summary, results from the evaluation showed promising language behaviors from many children, including mimicry of the robot and communication strategies indicating they thought of it as an agent capable of understanding by multiple means. Children's actions showed that they accepted the hybrid make-believe context and saw it as part of a shared physical space. Parents and siblings spontaneously scaffolded and modeled communication with the robot, suggesting that the technology might also be able to support children's learning indirectly through their family's participation.

Work remains to be done to validate these results: perhaps in part due to the small sample size, differences in communicative behaviors between conditions were not statistically significant. Retention of vocabulary also was not seen after the single exposure to the system.

Contributions

Technical/Design

The technical and design contributions of this thesis included the Iterative design and implementation of a hybrid game bringing digital objects into a spatially situated context, with contingent actions and speech by a social robot. This robot integrated expressive gestures, contingent gaze, speech and association mouth movements (visemes) into a French-speaking character. Finally, the aesthetics of a novel robotic character designed to look like a stuffed animal and support associated role-playing play patterns were developed.

Study

The unique contributions of this study in the domain of robots for language instruction was the evaluation of the following combination of approaches: first, a robot modeled as a native speaker of a new language that does not understand the child's language, is a learner as well as a teacher, and gives feedback when it does not understand the child's language. Second, a novel shared activity combining flexible digital content congruently with a physical "play prop" anchoring a particular social play interaction. Finally, evaluation within a social context including caretakers and siblings as participants.

Hypothesis Results Summary

Specific results to the research hypotheses and questions posed in the final study are summarized in Table 6.

Children will speak more to the robot in the "feedback" condition.	No. However, children used a larger number and wider range of communication strategies to make themselves understood by the robot in the feedback condition.
Children will learn more of the fruit names in the "feedback" condition.	No. The duration and number of exposures to the new vocabulary in this study was not sufficient for most of the children to learn the fruit names.
Children will mimic the robot's utterances.	38% of the children spontaneously mimicked the robot's utterances.
Parents will treat the robot as a social agent and engage both in modeling (teaching by example) and direct scaffolding of their children's communication with the robot.	Yes. All parents engaged in some form of supportive communication. These are detailed in the "What kind of modeling, ..." question below.

Will children treat the digital objects as part of physical shared space between them and the robot?	Yes. One child in particular pointed to the fruit on the digital plate while explaining something to the robot through gesture, showing that they perceived both digital and physical elements as being in a shared visual context.
What will children do and say when the robot is confused and needs guidance (when the shoe appears?) Will they explain to it what is happening?	No, children did not explain to the robot what was happening – rather, the shoe prompted discussion between parent and child about how to treat an object that did not fit the play context (table surface) without breaking the metaphor, showing evidence that this digital/physical context was accepted. It also caused great hilarity.
What actions of the robot's will trigger children speaking to it?	The robot's "confused" reaction (in the feedback condition) to being spoken to in English prompted a wide variety of communication strategies in order to re-explain their meaning.

How much will the robot be treated as an autonomous agent, and how much will children "fill in" its behaviors and speech as though playing with a stuffed animal? Do different actions and non-actions of the robot's trigger different levels of each?	The robot was primarily treated as an autonomous communicative agent with regards to its verbal behaviors – children did not "speak for" the robot. With regards to its actions in general, there were a few instances of imaginative elaboration and explanations both at a higher level (explaining its intentions) and relating to specific gestures and expressions ("she is happy and full").
Will children understand the explanation that the robot speaks only French, and will they try to speak to the robot in French?	Inconclusive. 31% of the children addressed the robot in French. Effects of children's and parents' prior knowledge of the language was not controlled for, and their comprehension of the idea that the robot spoke only French was not explicitly measured.
What kind of modeling, scaffolding, and/or reinforcement will parents engage in?	Parents engaged in modeling behaviors, including speaking to the robot, speaking to the robot in French, treating the robot as a social and conversational agent during the shared activity, and touching the robot. They also guided the children's behaviors by prompting them to speak to the robot in French and in English, managing turn-taking, sharing, and responding to the robot's social and communicative actions. They engaged in dialogic guiding, asking children to think about, predict, and discuss the robot's behaviors. Finally, they acted as mediators between the robot and their children, repeating, pointing out, and reinforcing the robot's actions and guiding children to an appropriate response.

Table 6: Summary of research hypothesis and major questions results

Study Successes and Research Goals Achieved

Family Participation

Both parents and siblings participated in the social interaction and spontaneously reinforced behaviors naturally in line with the language learning subject matter: social behavior, practicing speaking to and understanding the robot, interpreting the robot's gestures and expressions in a social context.

Parents modeled language behaviors by speaking to the robot in English and, when they knew the language, in French. They also scaffolded children's communication with the robot, pointing out social cues such as facial expressions and gaze, explaining its actions in terms of social agency, prompting children to speak and to share, and reinforcing children's behaviors through praise and enthusiasm. Modeling behaviors were correlated with children's increased speech directed towards the robot and with treating it as a social agent.

Siblings helped each other interact with the robot by prompting actions, discussing the robot's behaviors, and helping use the tablet interface. The tablet and robot system supported shared use.

The significance of these findings is that the design supported family participation, successfully including important parent and sibling behaviors that language acquisition research shows help foster language learning. This aspect has not traditionally been explored in other work with second language learning technologies.

Children's Language Use

In the final study, all but one child spoke to the robot at least once. Over a third of children mimicked the robot's French utterances, a behavior supportive of language learning.

In several cases where the parent spoke French, children asked their parents how to say phrases in French to the robot. In this interaction, children were motivated enough to communicate with the robot that they engaged in "language pull."

Use of strategies to re-communicate with the robot trended higher (though not significantly so) when the robot gave feedback that it did not understand. In addition to prompting more speech from the children, this robot behavior also turned out to be a method of revealing the children's mental model of the robot's speech and understanding abilities. Over a third of the children used communicative strategies that indicated they thought of the robot as a social agent capable of understanding the meaning of speech rather than as a software program that would not have the ability to understand something explained differently (rather than repeated or said more loudly).

Shared Context

Children's actions showed that they accepted the hybrid make-believe context and saw it as part of a shared physical space. One child gestured to the robot to explain something, pointing at a digital object within the physical space between robot and child – indicating a belief that the robot could perceive this shared physical space and that the digital objects were part of it. Children's actions when they encountered a novel object that did not fit the make-believe context (a shoe, when the only make-believe action possible was eating) implied that they accepted the blended physical/digital metaphor and shaped their play around it.

Remaining Challenges

Based on the pre and post testing, children did not retain the vocabulary presented during the interaction. This was not surprising given that it was only a single exposure, but further evaluation would be needed to learn if the system supports vocabulary retention. Prior work with RUBI established that children will retain vocabulary presented by a robot; full evaluation of the system in this thesis would establish that this was possible in a fully language immersive (ie. in which the robot does not speak any English) context.

Differences between the study conditions in terms of use of communicative behaviors by the children were trending, but not significant, perhaps due to the small sample size.

Some of the robot's expressions were not readable by the children and in some cases by their parents. The robot's "interested, acknowledging" reaction to hearing French spoken was not sufficient to confirm to the children that they had been understood, and they continued to seek a response. The robot's reaction to the shoe was overshadowed by the child's own reaction to the novel object.

VII. Future Work

Evaluation

Validating Comparison to Existing Technologies

Value of physical/digital prop: Compare table to tablet

Results from the study showed that at least some children believed the digital objects could be perceived by both themselves and the robot. For nearly all of the children and their families, the physical prop appeared to have facilitated shared play rather than single-user screen-based interactions. However, in this study these new interfaces were not compared to existing, commercially available technologies. A future study with a sufficient sample size comparing children's behaviors with the "spatially congruent digital props" to the same activity with a standard tablet computer would help validate the observations.

Value of contingent behaviors: Compare Social Robot to Interactive Toy

In the final study, a robot with socially interactive behaviors (gaze reflecting attention to child or food, introducing itself at the appropriate moment, and responding to being spoken to) appears to have been treated as a conversational partner with the ability to comprehend meaning and speech. Again, to validate this result, a study would need to be conducted comparing this socially interactive robot to a moving robotic toy without such behaviors. This would address the question of whether commercially available robotic toys without complex sensing capabilities for speech, awareness of their environment, nor contingent social behaviors would achieve the goal, or whether this responsiveness is indeed critical.

Compare Social Robot as Language teacher to existing interactive games

While television does not teach language beyond vocabulary, there exist an increasing number of interactive games for children today – some of which may be accessible to preschool age children. The premise of this thesis is that socially situated language learning as made possible by a physical social agent would support aspects of language learning that are not possible outside of a social, communicative context. This thesis showed that at least some children treat a social robot as a social communicative partner with the ability to understand the meaning of language. Future studies will attempt to refine behaviors of the robot such that most children are guided to treat it this way, and such that they will also try to understand its language. Then, a multiple-exposure trial will be needed to measure learning. Eventually, it will be important to run a study comparing the language teaching robot to existing technologies.

Conversational Analysis

The coding scheme designed for the final study, as described in the Study Setup section, revealed few significant correlations between parent and child utterance data, nor differences between study conditions. In part this was due to the small sample size, but the coding scheme may also need to be improved to capture the subtleties of conversation.

In future work, video coders trained in the CHAT transcription format might be employed so that the conversations between children, parents, and the robot could be compared against existing databases and standard conversational analysis might be used (MacWhinney, 2000). Of particular relevance is the CHIP framework (Sokolov & MacWhinney, 1990), designed to measure verbal imitation and expansion. Questions to ask include what kinds of utterances modeled by parents are children repeating? By the robot?

In addition, reliability was not established in the language coding for this study. Objective coders analyzed randomly assigned video, but they each coded separate video. In future work, all video should be analyzed by at least two independent coders and inter-coder reliability computed.

Audience

As touched on in the participant demographics section, within the same age group language learners have situation-specific needs. Within the United States, the situation for a native English speaker learning a second language for educational purposes is quite different from that of children recently moved to a country and learning the language they are newly immersed in. In the former case, immersion and motivation to use the language are key needs. In the latter case, a less-overwhelming space to practice in and positive experiences with the new language may be the most important contributions. This preliminary study did not distinguish between these populations, but there is a great opportunity to narrow in on the specific requirements of language minority children.

Improving and Extending the System

The Robot's Interactive Behaviors

The robot programmed for the final study possessed a few simple contingent behaviors:

- Introducing itself at the appropriate moment
- gaze up at child's eye level or following the fruit on the table depending on whether the child was manipulating the fruit
- response to the language spoken to it (as detected by a teleoperator)

Children did show that they believed in the robot's ability to understand them – even though they often did not succeed in making themselves understood given its limited interaction capabilities. Moving forwards, I would add to its capabilities:

- more accurate "look at" behaviors and the ability to detect and follow where the child is gesturing.
- a wider range of possible verbal responses to the child's speech.

The "hmm?" sound of the "hearing English" state appeared to be particularly effective in communicating a request for clarification. I would also experiment with other non-language sounds expressing the robot's level of comprehension, encouragement, and interest in the child's utterances. Simple conversational feedback cues such as these are valuable to identify because they can extend a conversation without highly complex natural language processing capabilities nor too strict a script, giving children a more open-ended, pretend play space in which to practice while also giving them encouragement to continue. It is of course also important to give verbal feedback and new vocabulary, but elements meant to maintain the rhythm and flow of a conversation are a critical part and under-addressed aspect as well.

Prior to fully implementing these capabilities I would run a study with a teleoperator controlling all the robot's behaviors and sensing, and learn which of the above were critical and at what level of precision.

Teleoperator Interface

A control panel on a tablet device for teachers and parents to drive the robot remotely was explored in this thesis but not fully developed. The concept was to create an intelligently adaptive interface for remote control of high-level aspects of the interaction, while automating lower-level robot behaviors to reduce the operator's cognitive load.

The planned interface would adapt intelligently to the system's current state so that it would present only the most likely actions for the operator to choose from given the shared activity context, including such actions as speech, gaze, expressions of emotion, and game moves. The system would automate expressive behaviors of the robot as much as possible using sensor and activity state data to allow an untrained operator to puppeteer at a higher level.

This interface could potentially enable even parents and teachers who do not speak the language the child is learning to help a child learn, as it could be presented to the parent in their own language, and then translated to and from the robot.

A design goal of the teleoperator interface would be to be fun for the adults too – perhaps even enabling remote parents and grandparents to play at a distance.

Language Capabilities

The study with Sophie the Dragonbot explored children's language use with a robot that did not speak very much. As a learning tool, this was appropriate for an initial introduction to the robot and the language. However, to effectively teach, Sophie needs to be able to introduce more language and respond appropriately to the child's utterances. The shared activity has the benefit of grounding the conversation on a specific topic, limiting the language knowledge needed to something that could be implemented with current technology.

There is an opportunity to do an in-depth analysis with a sufficient sample size of the categories of questions children want a response to in a given shared activity such as the French cafe, and to use these to build a model of language capabilities needed for this or another specific shared activity.

Expanding the Digital Play Context

The hybrid digital/physical game designed in this thesis established that children could play with digital objects while situating them in shared physical space with a robot. The opportunity provided by this type of system is in the combination of the grounded nature of physically shared objects with the adaptability and responsiveness of digital objects.

Recent longitudinal studies of English language learners show that these children acquire everyday vocabulary successfully, but struggle with the abstract vocabulary required for academic reading comprehension (Mancilla-Martinez & Lesaux, 2011) (Jean & Geva, 2009). Angela Chang's "TinkrBooks" (Chang, Breazeal, Faridi, & Roberts, 2012) address this challenge of literacy learning for the general population. This research showed that digital picture books allowing children to "tinker" with story elements helped support literacy learning behaviors, and learning of higher-level written language concepts. Inspired by TinkrBooks' approach to helping children explore the meanings of abstract words and motivated by the need in the English language learner population, scenarios might be designed that employ the digital elements of the robot language learning system in more nuanced ways than for simple object word learning in order to impact abstract vocabulary.

References

Baumgartner, E. (2003). Design-Based Research: An Emerging Paradigm for Educational Inquiry. *Educational Researcher*, 5-8.

Bers, M. U. (2007). *Blocks to Robots: Learning with Technology in the Early Child- hood Classroom.* . Teachers College Press.

Chang, A., Breazeal, C., Faridi, F., & Roberts, T. D. (2012). Textual tinkerability: encouraging storytelling behaviors to foster emergent literacy. *Proceedings of the 2012 CHI Extended Abstracts* (pp. 505-520). New York, NY: ACM.

Chang, C.-W., Lee, J.-H., Chao, P.-Y., Wang, C.-Y., & Chen, G.-D. (2010). Exploring the Possibility of Using Humanoid Robots as Instructional Tools for Teaching a Second Language in Primary School. *Educational Technology & Society*, 13-24.

Freed, N. (2010). *Toys keeping in touch : perspective taking and make-believe-centered design for children's remote communication.* Arizona State University, M.S. Thesis Computer Science.

Garcia, O. (2011). *Bilingual Education in the 21st Century. A Global Perspective.* Wiley-Blackwell.

Gardner, E. (2011). *About Where are Your Keys.* Retrieved February 1, 2012, from Where Are Your Keys: http://www.whereareyourkeys.org/

Gray, J., Hoffman, G., Adalgeirsson, S. O., Berlin, M., & Breazeal, C. (2010). J. Expressive, Interactive Robots: Tools, Techniques, and Insights based on Collaborations. Presented at workshop "What do collaborations with the arts have to say about HRI" at HRI 2010.

Jean, M., & Geva, E. (. (2009). The development of vocabulary in English as a second language children and its role in predicting word recognition ability. *Applied Psycholinguistics* (30(1)), 153-185.

Kuhl, P. K. (2007). Is speech learning 'gated' by the social brain? *Developmental Science*, *10*, 110-120.

Kanda, T., Hirano, T., & Eaton, D. (2004). Interactive Robots as Social Partners and Peer Tutors for Children: A Field Trial . *Human-Computer Interaction*, *19*, 61–84.

Kidd, C. D. (2008). *Designing for Long-Term Human-Robot Interaction and Application to Weight Loss.* PhD Dissertation: Massachusetts Institute of Technology.

Lee, S., Noh, H., Lee, J., Lee, K., & Lee, G. G. (2010). Cognitive Effects of Robot-Assisted Language Learning on Oral Skills. *Proceedings of Interspeech Second Language Studies Workshop.* Tokyo.

Lepper, M., & Henderlong, J. (2000). Turning "play" into "work" and "work" into "play": 25 years of research on intrinsic versus intrinsic motivation. In Sansone, & J. Harackiewicz, *Intrinsic and extrinsic motivation: The search for optimal motivation and performance* (pp. 257-307).

Lesaux, N., Koda, K., Siegel, L., & Shanahan, T. (2006). Development of literacy of language minority learners. (). In D. August, & T. Shanahan, *Developing literacy in a second language: Report of the National Literacy Panel.* (pp. 75-122). Mahwah, NJ: Lawrence Erlbaum Associates.

Naigles, L. R., & Mayeux, L. (2001). Television as Incidental Language Teacher. In D. G. Singer, *Handbook of Children and the Media.*

MacWhinney, B. (2000). The CHILDES Project: Tools for Analyzing Talk. *3rd.* Mahwah, NJ: Lawrence Erlbaum Associates.

Mancilla-Martinez, J., & Lesaux, N. (2011). The gap between Spanish speakers' word reading and word knowledge: A longitudinal study. *Child Development* (82(5)), 1544–1560.

McCrane, M. (n.d.). *Using Toca Boca iPad Apps with Pre-K.* Retrieved March 15, 2012, from http://www.maggiehosmcgrane.com/2012/03/using-toca-boca-ipad-apps-with-pre-k.html

Meltzoff, A. N., Brooks, R., Shon, A. P., & Rao, R. P. (2010). "Social" Robots are psychological agents for infants: A test of gaze following. *Neural Networks* , 966-972.

Movellan, J., Eckhardt, M., Virnes, M., & Rodriguez , A. (2009). Sociable robot improves toddler vocabulary skills. *HRI '09 Proceedings of the 4th ACM/IEEE international conference on Human robot interaction* (pp. 307-308). New York, NY, USA: ACM.

Picard, R. W., Papert, S., Bender, W., Blumberg, B., Breazeal, C., Cavallo, D., et al. (2004). Affective Learning — A Manifesto. *BT Technology Journal , 22* (4), 253-269.

Setapen, A. (2012). Creating Robotic Characters for Long-term Interaction. Massachusetts Institute of Technology, S.M. Thesis Media Arts and Sciences.

Turkle, S., Breazeal, C., Daste, O., Scassellati, B.. (2004) Encounters with Kismet and Cog: Children Respond to Relational Artifacts.

Smarty Ants Preschool Kindergarten & Grade 1 Reading Games. (2011). Retrieved October 27, 2011, from http://www.smartyants.com/

Ryokai, K., Vaucelle, C., & Cassell, J. (2002). Literacy learning by story-telling with a virtual peer. *Proceedings of the Conference on Computer Support for Collaborative Learning: Foundations for a CSCL Community.* International Society of the Learning Sciences.

Robert, D. Y. (2011). *Imaginative play with blended reality characters.* Massachusetts Institute of Technology, S.M. Thesis Media Arts and Sciences.

Turkle, S., Breazeal, C., Dasté, O., & Scassellati, B. (2004). Encounters with Kismet and Cog: Children Respond to Relational Artifacts.

Tabors, P. O. (2008). *One Child, Two Languages: A Guide for Early Childhood Educators of Children Learning English as a Second Language.* Paul H Brookes Pub Co; 2nd edition.

Tannen, D. Implications of the Oral/Literate Continuum for Cross-Cultural Communication. In J. E. Alatis, & J. J. Staczek, *Perspectives on Bilingualism and Bilingual Education.*

Vaucelle, C., & Ishii, H. (2008). Picture this!: film assembly using toy gestures. *Proceedings of the 10th international Conference on Ubiquitous Computing* (pp. 350-359). New York, NY: ACM.

Wistort, R. (2010). *TofuDraw: Choreographing Robot Behaviors through Digital Painting.* Massachusetts Institute of Technology, S.M. Thesis Media Arts and Sciences.

Appendices

Appendix 1: *Conversational Modules*

The following five dialogues of increasing difficulty were designed to teach and test new vocabulary in the context of a "sharing food at a café" game. The first dialogue teaches yes and no and introduces the ordering process. The next is a practice for yes and no, and introduces a choice between different types of fruit. The next teaches colors and is a game about dipping the correct piece of fruit into the chocolate. The next introduces numbers, and the final one is practice combining numbers, colors, and fruit names.

1. veux-tu une pomme?
prompt: oui (nod) ou non (shake head)
words to recognize: oui, non
(looks towards plate/waiter) une pomme, s'il vous plait
apple move to plate animation
voila!
OR
go back to the beginning with a new fruit

2. veux-tu une pomme ou une orange?
(fruit wiggles as each word is said)
words to recognize: [fruit list] (if one not in request list is requested, switch plate!)
(looks towards plate/waiter) une pomme, s'il te plait
apple move to plate animation
voila!

3. donne moi une fraise avec du chocolat
donne moi un raisin sans chocolat
(animations for dipping each fruit into chocolate)
(if it's wrong) non merci (goes back to plate)
merci!

4. combien de fruits veux-tu?
prompt: un, deux, trois, ...
(number of fruit wiggle as each word is said)
words to recognize: [numbers list]
(looks towards plate/waiter) trois pommes, s'il vous plait
apple move to plate animation
voila!

5. que veux-tu manger?
prompt: pour moi, trois framboises, s'il vous plait
prompt: une pomme, une orange, une fraise, un abricot
(fruit wiggles as each word is said)
prompt: une pomme, deux pommes, trois pommes...
(number of fruit wiggle as each word is said)
words to recognize: [fruit list, numbers list]

Appendix 2: *Pre-Test of Receptive and Expressive Vocabulary*

What is this fruit in English?

1.
2.
3.
4.
5.
6.

What is this fruit in French?

1.
2.
3.
4.
5.
6.

What fruit is this? (play audio)

What is your favorite of these fruit?

Appendix 3: *Post-Test of Receptive and Expressive Vocabulary*

What is this fruit in French?

1.
2.
3.
4.
5.
6.

What fruit is this? (play audio)

What is Sophie's favorite fruit?

Appendix 4: *Parent Survey*

1. How old is your child? _________

2. What is your child's gender? ________

3. What is your child's experience level with the French language?

1	2	3	4	5	6	7
Has never heard the language		Knows a few words				Fluent speaker

4. What is the highest experience level with the French language in your immediate household?

(For example, you or a spouse speak the language)

1	2	3	4	5	6	7
Has never heard the language		Knows a few words				Fluent speaker

5. ***Rank*** how much time your child spends on each of the following activities (5 = most time spent, 1 = least time spent)

___ pretend play (storytelling/role-playing with dolls, action figures, stuffed animals)
___ construction play (building with blocks, legos, sand,...)
___ craft and art activities (modeling clay, drawing, decorating, cooking)
___ play with computers, touchscreen devices, handheld gaming toys, or video games
___ outdoor games or organized sports (eg. tag, soccer, playground play)

6. How comfortable is your child speaking to (communicating using language) with new peers or adults?

1	2	3	4	5	6	7
Completely comfortable						Finds this very difficult

How much do you agree or disagree with the following statements?

7. I felt that the robot was trying to understand what we were saying.

1	2	3	4	5	6	7
Completely disagree						Completely agree

8. It was difficult to speak to the robot.

1	2	3	4	5	6	7
Completely disagree						Completely agree

Appendix 5: Instruction Sheet for Parents

Hello there! Thanks for coming to meet one of our brand-new robots. Please read carefully...

Sophie is a shy little robot who only speaks French. She's very excited to meet you! You'll have 10 minutes to visit with her and then her caretakers will come back, but it's okay if you would like to stop sooner.

Sophie loves to be talked to. She does not understand English yet and sometimes feels a little lost, but she wants to learn and to teach you her language.

Sophie gets a little worried when things are not where she expects them, so please try not to move the table and other objects too much. She also only likes to be petted when one of her caretakers is with her.

Step 1: Greet the robot and introduce yourselves!

Step 2: Touch the plate and *hold for a few seconds* to order a fruit for you or Sophie to eat. She might tell you the name of the fruit in her language.

Step 3: Move the fruit towards Sophie to feed her...
or towards your end of the table to take a bite!

You can order as many fruit as you like, one at a time. If you want to order a new one, eat the one that's on the table!

Note that there are some fruit Sophie likes better than others, and that magic tables sometimes hold surprises.