“It knows how to not understand us!”

A study on what the concept robustness entails in design of conversational agents for preschool children

Julie Hagen Nilsen & Kristine Røyneland

Thesis submitted for the degree of
Master in Informatics: Design, use, interaction

120 credits

Department of informatics
Faculty of mathematics and natural sciences
UNIVERSITY OF OSLO

May 2019

“It knows how to not understand us!”

A study on what the concept robustness entails
in design of conversational agents for preschool children

Julie Hagen Nilsen & Kristine Røyneland

May, 2019

“It knows how to not understand us!”

http://www.duo.uio.no/

Printed: Reprosentralen, University of Oslo

Abstract

This thesis explores preschool children's interaction with a conversational agent (CA), and we position this work within the field of Human-Computer Interaction (HCI). Through a preliminary case study, we explored children as users of conversational technology and gained insights regarding how a CA could function as a pedagogic tool in a kindergarten setting. In a succeeding design process consisting of three distinct phases and four iterations, we developed and evaluated a prototype in order to gain a deeper understanding of what robustness entails in child-CA interaction. In a concluding field trial, we deployed the prototype in a kindergarten in order to study the interplay between all actants involved in use; the CA, the children, and the educators. We applied Steven Jackson's (2014) perspective of *broken world thinking* to guide our exploration throughout the design process. This had us focusing on how conversations are always breaking, as well as how the users conduct repair work when a communicative breakdown occurs. The findings indicate that various trouble sources complicate the communication between children and CAs. The trouble sources primarily derived from the subtle differences that make conversations between users and CAs different than ordinary conversations between people. This thesis shows how attempting to develop a prototype that addressed several of these trouble sources resulted in an increased understanding of how these should be mediated through design. At the same time, it illustrates how an exclusive focus on breakdown can devaluate inherent characteristics within the user group and thus become detrimental to the design result. We offer suggestions on how to support robust communication between children and CAs, and provide three overarching design recommendations that we believe practitioners within the field can benefit from: (1) support the user in understanding communicative constraints, (2) support the collaborative nature of conversation, and (3) understand what constitutes an effective conversation.

Acknowledgements

First, we would like to thank the children and educators participating in this study. Thank you for your time, patience and days of fun. Without you, this thesis would not exist. We would also like to give a huge thank to our supervisor, Hanne Cecilie Geirbo, for sharing your knowledge and expertise, guiding us through every step of this process. Your help has been extremely important in times of uncertainty, and your enthusiasm has motivated us all the way to the finish line. A massive thank you is also directed to our co-supervisor, Tone Bratteteig, who guided us through the design process, helping us to think in new ways when we found ourselves stuck. Your sharing of knowledge is very much appreciated. Thanks also to the Norwegian Institute for Air research for the collaboration which became the starting point of our research. Even though our study pivoted away from the subject of air itself, we would not have ended up where we did without you as a starting point. To our people on the 7th floor; thank you for being the best friends anyone could ask for. The fun times and lunch-quizzes will be forever remembered. Thank you for making our four years at UiO the best time ever. Last, but not least, thank you to our partners, families, and friends for their support and interest in our project. We love you.

Julie Hagen Nilsen & Kristine Røyneland
University of Oslo
May 2019

Contents

List of Figures

List of Tables

Prologue

> *The single biggest problem in communication is the illusion that it has taken place.*
>
> *- George Bernard Shaw*

"Hey, what's that on your head?" Annie, a three-year-old girl, sounds bemused as she directs her question to the embodied conversational agent sitting before her. She is referring to a big fluff of fabric that is supposed to depict a cloud resting atop a provisional robot we have made to represent the conversational agent we are currently developing. A silence ensues as the prototype attempts to figure out what she meant, and if it is indeed capable of answering the question. Annie looks to us and says, "he's not answering". "He's thinking", we explain, knowing that the agent is unable to answer this particular inquiry. Finally, the agent responds in its monotonous voice "Hi! I can give you missions, make stories or play a quiz? Do you want to receive a mission, make a story or play a quiz?" Nina, who has been watching the interaction from the sidelines looks to us and proclaims eagerly "it can talk!" whilst Annie hastily answers "mission!" Two more children decide to engage with the agent and playfully repeat the word "mission" several times over. Once again, the agent takes its time coming up with a response and Magnus groans, "he never understands!" His statement is confirmed when the agents eventually answers "sorry, I didn't catch that. Try saying it one more time"

Sustaining a mutual understanding of the unfolding action or the topical issue during interaction is a prerequisite for successful communication. If the participants fail to achieve this, the exchange will simply be an illusory manifestation of communication. Child-CA interaction is tricky, as neither party are proficient in the arts of conversing. We aim to contribute knowledge that can remedy this situation, so that children can be adequately supported during spoken interaction with machines in the future. After all, no one wants to talk to someone who never understands what they are saying.

1 Introduction

Have you ever had a frustrating experience with chatbots or digital assistants? If your answer is no, you should consider yourself lucky. The possibility of talking to machines holds a promise of effortless interaction because for most of us, engaging in ordinary conversations is as natural as breathing. Conversations are the way we share knowledge, how we get things done, and how we connect socially and emotionally with the people around us. However, making interaction with machines feel as easy and instinctive as conversations do has been a long and winding road and the current deficit in these agents' conversational abilities falls short of their implied humanness.

Conversational agents (CA) are becoming increasingly pervasive in our society today as their ability to understand natural language increases with access to progressively larger data sets. The possibilities the advances in conversational interfaces offer regarding human-computer interaction are exciting and novel, contributing to a steadily growing body of research within the domain of conversational user experience design. However, interactions between CAs and users frequently fail, attesting to the difficulties of designing for robust conversations.

In this thesis, we detail our approach, experiences and insights from conducting research on what characterizes spoken interaction between CAs and preschool children. Over the course of six months, we conducted an instrumental case study and a subsequent empirical study incorporating a user-centered design process within which we developed and evaluated a prototype along with children and educators from two kindergartens. Our experiences suggest that human-CA interaction is especially fragile when the user group comprises of young children, partly because they are in the midst of developing their conversational competencies, and partly due to their often playful, impulsive and associative behavior. We have found that this often triggers breakdowns in interaction, as the CA we observed in use were unable to handle these issues adequately.

As a consequence of the fragility in interaction between children and CAs, we applied *broken world thinking* (Jackson, 2014) as a perspective in the design process. This is a perspective within technology and media studies that foregrounds breakdown and disintegration as the starting point for technological development in the hopes of creating a turn towards a repair-centric view on progress and innovation. Breakdown and ensuing repair is a concept that is central in both human-computer interaction and in the

organization of ordinary conversation. Thus, we attempted to apply this perspective in a process in which we worked with dialogue as the primary design material. We asked what we could learn by making breakdown the starting point for understanding the challenges in interaction between children and CAs.

During this narration, we tell the story of the challenges and realizations that accompanied our attempts to design a CA that could take young children's conversational competencies and inherent playfulness into consideration. There are numerous technological tools available to those who wish to develop CAs – ranging from software tools that require extensive programming skills to chatbot building platforms that allow you to make CAs without writing a single line of code. With these tools, the designers and developers can structure the dialogue according to the intended objective of use. However, the options available to us were limited as only one tool provided spoken input and output in Norwegian. Furthermore, this tool was limiting as it only allowed us to explore certain ways to organize a conversation. As the design process progressed, we witnessed continuous breakdowns in communication between the CA and the participants, and we found ourselves creating a strict and rigid dialogue design that left little room for exploration and play. Hence, we ultimately failed in our ambition to design a CA that was in harmony with the behaviors the children displayed.

This study point to various prevalent trouble sources in child-CA interaction and illustrate how these complicate cooperative meaning-making. By applying theoretical concepts such as breakdown, the cooperative principle of conversation (Grice, 1982), communicative repair (Sacks, Schegloff, & Jefferson, 1974), and graceful interaction (Hayes & Reddy, 1983), we suggest that designing CAs for this age group entails supporting the user in understanding communicative constraints, supporting the collaborative nature of conversations, and designing for an effective conversation. In sum, we argue that designers should identify and accentuate the subtle constraints that make spoken interaction between people and machines different from human-to-human communication, and thus minimize the breakdowns that ensue because the children incorrectly assume they can speak to the CA as if it were a person.

1.1 Motivation

Our decision to explore conversational technology originated from previous experiences with the design and use of intelligent assistants, as well as an aspiration to study a unique design material. As interaction designers, we wanted to learn more about what lies beneath a conversational interface and how interactions with such devices are structured and created. In short, what does the creation of conversations entail within the practice of interaction design?

Conversational technologies are in continuous development, and Forbes magazine have predicted chatbots to be one of the top digital transformation trends in 2019 (Newman, 2018). They believe that by the end of the year, up to 40 percent of large-scale businesses will have adopted Natural Language Processing as a tool to improve their services. Virtual or digital agents are making headway within service industries and have taken on a variety of roles such as customer service agents, loan processors, and museum guides, to name a few. Additionally, smart devices with embedded digital assistants are finding their way into people's homes. Consequently, children are becoming an emerging target group as these devices invade their private spheres. We are of the opinion that currently, the conversational technologies available to the mass market largely fail to account for children as a user group, although research shows that they frequently interact with such technology (Sciuto, Saini, Forlizzi, & Hong, 2018). The collection of funny YouTube clips of kids trying to make themselves understood by Alexa or Siri can attest to the neglect of younger children as potential users. The merging of this user group and this design material warrants a closer look in order to lay a firmer groundwork on which to base CA-design for children. We believe this creates a need to explore how children use and perceive such interfaces and additionally consider carefully how children's communicative abilities and behaviors should affect dialogue design. We found that we wanted to contribute to this endeavor, and thus do our part in ensuring great conversational user experiences for this group in the future.

1.2 Context

In the spring of 2018, we established a collaboration with the Norwegian Institute for Air Research (NILU). NILU works to increase peoples understanding of air quality, hazardous substances, the composition of the atmosphere, as well as the effects the

aforementioned have on climate change (NILU, 2018). They strive to make the public more aware of these effects and enable people to interpret the data they gather and distribute correctly. Thus, the main objective of NILU's work is inciting behavioral change that benefits the environment. Our collaboration revolved around how we can utilize technology to spark curiosity and facilitate learning and reflection by young children on a complex topic such as environmental sustainability. Thus, we positioned the CA within the domain of pedagogy and environmental education on air and air-quality.

The overarching aim for the CA as a pedagogic tool was to contribute to the children developing an *environmental identity*. Clayton and Opotow (2003, pp. 45–46) propose that an environmental identity is a way of understanding oneself in connection to the nonhuman natural environment and that this connection affects how we perceive and act towards the world. An environmental identity or lack thereof, is important because it guides a person's personal, social and political behavior. Consequently, inhabiting a strong environmental identity is a motivating factor for pro-environmental behavior. For the CA to fulfill this function, we needed to know more of how it could work as a pedagogic tool, and how children behave as users of conversational technology.

In early fall of 2018, we established collaborations with two kindergartens in the municipality of Oslo. Thus, we explain how we can situate the CA within established guidelines for education in kindergartens. In Norway, the Norwegian Directorate for Education and Training supervise education in kindergartens based on a National Curriculum (Norwegian Directorate for Education and Training, 2017) in which one of the core values is *sustainable development.* The curriculum states that kindergartens play an important role in promoting values, attitudes, and practices for creating more sustainable communities, thus kindergartens are obligated to help children understand how their actions have consequences for the future. Consequently, in relation to the stakeholders in this project, the objective was to contribute an understanding of how CAs could function as a digital tool and a resource for teaching children about important environmental issues, such as air pollution.

1.3 Target group

In our research, we positioned the children as *agentic* (Corsaro, 2000; Sorin, 2014); they are competent actors who actively participate in the world by simultaneously producing

their own children's culture as well as contributing to the production of adult society. The agentic child is an empowered and self-motivated learner and not a passive and innocent bystander in a stage of "becoming" an adult. They are in a state of "being", and can aid the construction of knowledge in collaboration with adults who in turn enable the children by sharing their resources with them (Sorin & Galloway, 2006). Consequently, we reiterate Sorin and Galloway's (2006) argument that we should conduct research with children rather than about them, and we strove to include them and support their participation throughout the project.

The target group for this project has been preschool children age 4-6. In Norway, children in this age group are in their last years of kindergarten. We decided to limit participation to the oldest children because we considered them capable of handling the complexity of the tasks we presented. However, two children age 3 participated in the project because we considered them capable of contributing valuable data due to personal aptitudes.

We recognize that this target group is heterogeneous. The participants in this project were in different developmental stages concerning communicative and cognitive abilities and had dissimilar backgrounds, interests, and opinions. Consequently, there were noteworthy variations in their abilities regarding social competencies and conversational skills. We elaborate on this further in Chapter 7.

We find it important to mention that we both have previous experience working with children. Julie has worked in a kindergarten for 2 years, and Kristine is educated within pedagogy and briefly worked as a teacher for children age 9-10. This familiarity prepared us for interacting with children of this age group, although the context of research comprised a different setting than the ones in which we have previous experience.

1.4 The process

Our research process has not been straight forward, and the concluding result was far from what we envisioned at the beginning of this venture. The model shown in Figure 1 is a visualization of the process and provides an overview of changes in both the research approaches and our focus of attention throughout the project.

The left side of the visualization illustrates what Sanders and Stappers (2008) call a "fuzzy front-end" as its ambiguity and chaotic nature often makes it unclear what is to be designed. This stage includes several activities that inform and inspire the exploration of open-ended questions, which is typical of the first phase in many design processes. We characterize our process as "fuzzy" prior to the case study. During this period, we worked towards narrowing our focus by exploring literature pertaining to different fields and made decisions that set the course for the subsequent process.

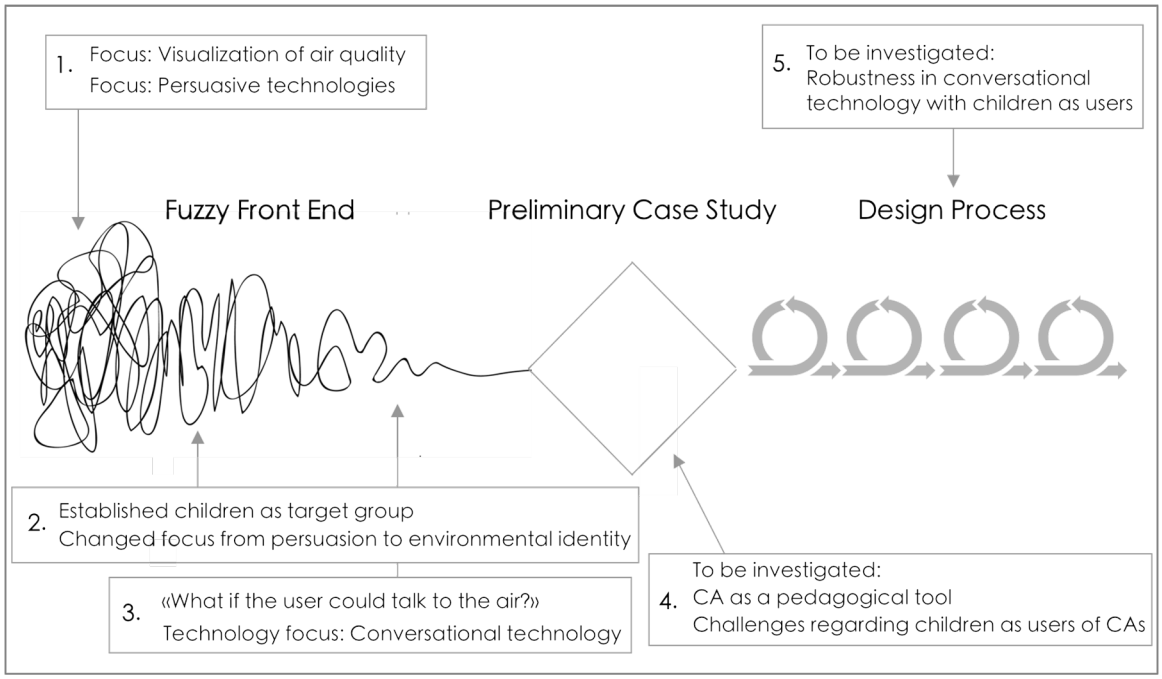


Figure 1 Visualization of our process

The part in the model representing the preliminary case study is inspired by "the double diamond" developed by the British Design Council ("The Design Process," 2015), which illustrates commonalities to a vast range of design approaches (see Figure 2). The first diamond in their model has two stages; *discovery* is the phase in which one gain insight into the problem, and *definition* is the phase in which one establishes the focus area. We conducted the case study to explore a context of use with the aim of specifying our area of research and thus find it suitable to compare it to the left-hand side of the double diamond. We diverged within an area of interest and subsequently defined a focus area for further exploration.

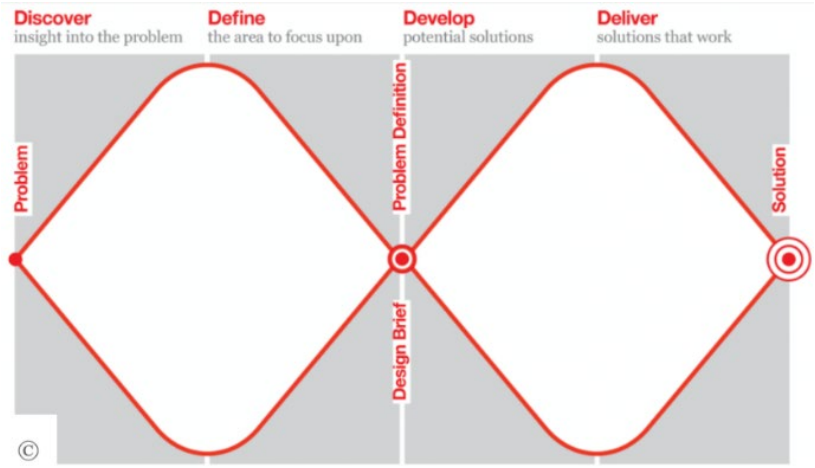


Figure 2 The Double Diamond developed by the British Design Council

In the final part of our model, we illustrate a design process by using a figure[1] of four iterations. During this time, we developed and evaluated a prototype. The culmination of the entire process was a field trial during within which educators and children tried to use the prototype.

As the model illustrates, we have made a continuous effort to specify our context of research. The starting point was exploring how we, through technology design, could contribute to communicate information about air quality and persuade behavioral change in a target group. However, we quickly diverged from this approach when NILU suggested that we design for a kindergarten context as it raised concern regarding small children's influence over decisions that affect air quality. Our argument was that children are not responsible for the degradation of air quality. It is futile to design with an objective to change the nonexistent. Regarding air pollution, children seldom perform harmful acts or possess bad habits as they neither drive nor decide to heat their home by wood burning. The adults make these decisions. In sum, we were concerned with placing too heavy a burden on the children by making the improvement of their parent's environmental behavior their responsibility, and we decided to abandon persuasive strategies as the focus area in the design process.

As an alternative, we found that supporting a growing environmental identity was a more suitable approach and decided to focus on exploring how design could support learning and reflection about the existence of air as a substance, how we are dependent upon it,

[1] Element representing one iteration is collected from https://www.hallnet.co.uk

and how human behavior affects it. Then we asked, “what if children could talk to the air?” This question led us down the path of exploring conversational technologies with young children as the target users.

The process of chiseling out an interesting area of research culminated in an attempt to combine three rather complex and distinct themes in order to explore the learning potential in the intersection between them; (1) designing with and for young children, (2) design of conversational agents, and (3) *broken world thinking* (Jackson, 2014) as a perspective in technological developments. These three areas are challenging in their own right because they comprise of knowledge gained from a wide range of research domains. Their interdisciplinary nature demanded us to familiarize ourselves with everything from techniques and methodologies related to including young children as users and informants in design processes, *broken world thinking* (Jackson, 2014) and different conceptualizations of breakdowns in design, to dialogue design and *conversation analysis* as a way of understanding the organization of ordinary conversation. Consequently, we have gone through much trial and error in order to learn how we should design and evaluate a prototype in a way that proved valuable to our primary objectives.

1.5 Research question and objective

We argue that an indispensable prerequisite for ensuring that digital tools in educational contexts fulfill their purpose is simply that they must work without causing too much trouble. In the case study, we identified several trouble sources in child-CA communication, which led to breakdowns in interaction. This had us reflecting upon what characterizes a robust CA, and how the concept robustness might comprise different things depending on the users and the technology. Consequently, this thesis attempts to contribute insights relevant to research and design of conversational technologies by examining the following research question:

> *What does the concept robustness entail in the design of conversational agents for preschool children?*

Our research project encompasses two distinct stages: a preliminary case study and a design process. During the case study, we used qualitative methods of inquiry such as interviews, participant observation and focus groups as a way to construct an

understanding of the context of use, as well as begin to study children as users of CAs. Through an empirical study involving a user-centered design (UCD) process, we built and evaluated a prototype that we ultimately deployed in a field trial with children and educators as participants.

We applied *broken world thinking* (Jackson, 2014) as an investigative lens in the design process and we offer a debate on its merits and faults in the final discussion. Additionally, we account for and use concepts from conversation analysis pertaining to the orderly organization of natural conversation as tools to understand the structuring of dialogue between users as well as the complexities of recovering from communicative breakdowns. We also compare and contrast our notion of robustness to traditional definitions within interaction design practices in order to construct our own understanding of this concept in relation to designing CAs for children.

1.6 Thesis structure

We have structured the thesis in the following way:

Chapter 2 – Background presents conversational agents as a technology and continue to explain concepts relevant to understanding dialogue as a design material. We describe the concept graceful interaction (Hayes & Reddy, 1983) as well as Dix et al. (2004) definition of robustness in interaction design. The concept of breakdown within the field of technology design is presented, with particular emphasis on *broken word thinking* as a perspective on technology development. We summarize the chapter by repeating key theoretical perspectives and concepts applied throughout the thesis.

Chapter 3 – Research methodology presents our philosophical perspective, case study as a research methodology and user-centered design as a design methodology. We also describe what roles children can have in a design process and how we consider the children to have participated in this process. We continue by describing the methods of inquiry we applied, the prototyping tool, as well as methods for data analysis. We end the chapter by discussing methodological challenges and ethical considerations relevant to this project.

Chapter 4 – Preliminary case study presents the activities done and insights gained through the preliminary case study, as well as how these insights informed the research question for this thesis.

Chapter 5 – The design process presents the design process as a whole, which consisted of three phases. Phase 1 was an introductory stage in which engagement of the participants was the focus. Phase 2 comprised of three iterations of prototyping, evaluations, and analysis. Phase 3 encompasses one iteration entailing prototyping, a field trial, and a succeeding analysis of the interplay between the three actants in use: the CA, the children, and the educators.

Chapter 6 – Discussion presents our experience with using broken world thinking as a perspective in a design process. We continue by presenting and discussing our insights related to the research question regarding what the concept of robustness entails in child-CA interaction. We end our discussion with three design recommendations, which we believe can contribute to more robust child-CA interaction in the future.

Chapter 7 – Critical reflections presents an evaluation of methods and tools used in this project, a reflection upon the children's role throughout the research process and thoughts regarding the introduction of conversational technology in a kindergarten setting.

Chapter 8 – Conclusion presents our contributions and possibilities for future work.

2 Background and theory

In this chapter, we present four main themes: conversational agents, dialogue as a design material, graceful interaction in dialogue design, different definitions of robustness within various technological domains, and perspectives on breakdowns within design and technology studies. In our final remarks, we summarize the content of the chapter by reiterating the key theoretical concepts for this thesis.

2.1 What are conversational agents?

Conversational agents are software programs that interpret and respond to human speech in real-time based on Natural Language Processing (NLP), text to speech conversion, and artificial intelligence (Hall, 2018; Luger & Sellen, 2016). The aforementioned software comprises the core of a CA as it enables it to communicate with its users. However, as with humans, other characteristics affect a CAs communicative ability such as its tone of voice, prosody, choice of words, phrasings, how it structures a conversation and the personality it presents. These are key elements significant to the field of interaction- and user-experience design.

In the following section, we provide a brief history of conversational technology and offer some examples of how this technology is currently applied. We attempt to make clear the connection between CAs and the field of pedagogy by explaining the use of conversational pedagogic agents (CPA) within education and continue by relating some studies that have examined children as users of CAs. Lastly, we account for a few important distinctions and features within the realm of conversational technology.

2.1.1 A brief history of CAs and its application areas

Although recent advances have launched an influx of conversational technologies, CAs have been around for some time. Best known is perhaps "ELIZA" – a software program launched in 1966, which behaved as a Rogerian psychotherapist[2] (Weizenbaum, 1966). ELIZA could engage users in natural written conversations by using key-word spotting and pattern matching to construct a response in the form of a prompt for further

[2] A therapist who assumes the role of compassionate facilitator and lets the client take the lead

elaboration. Since the days of ELIZA, conversational technology has slowly evolved and been applied within a range of domains and contexts. Today, most of us are familiar with CAs through the fields of e-commerce and customer service. Most of these agents are task-oriented or provide the user with support in navigation and information retrieval on complex websites. Digital assistants (e.g. Apples Siri, Google Assistant, Amazon's Alexa and Microsoft Cortana) embedded in personal devices are also well known and becoming more prevalent as smart speakers connect us to the services and tools in our homes. Another timely example is "Kommune-Kari"[3], a task-oriented assistant that several municipalities in Norway have deployed in order to lighten their workload by filtering question posed by their citizens.

Besides serving a practical purpose, CAs can provide entertainment and fulfill different social needs in its user's lives. Presently, a project collaborating with the website Ung.no are developing a CA aimed at supporting young peoples need for information and guidance. The children interviewed comment that they might feel more comfortable talking to a CA rather than a human about certain sensitive topics (Ditlefsen, 2018). Furthermore, emotional agents are currently on the rise, and some chatbots have been developed with the sole purpose of providing the user with emotional support (Stoltenberg, 2018). An example of this is Replika[4] for which the tagline is "if you're feeling down, or anxious, or just need someone to talk to, your Replika is here for you 24/7". These examples underline the fact that people employ CAs within a wide range of domains, attesting to the variations in both sought and possible functionality. Relevant to this project is research relating to the use of CAs in educational contexts. The following section provides a short overview of this field of study.

2.1.2 Conversational pedagogic agents in contexts of learning

In this project, we intended the CA to be an interactional tool embedded in a process of knowledge construction. Consequently, we briefly comment on the history of pedagogical agents (PA) and describe a small collection of relevant studies in order to make clear the interdisciplinary connection between the field of education and conversational technology.

[3] https://www.prokom.no/kari/
[4] https://replika.ai/

PAs are computer-generated characters employed in educational settings that serve various pedagogical goals (Gulz, Haake, Silvervarg, Sjödén, & Veletsianos, 2011). Conversational pedagogic agents (CPAs) belong to a subgroup that distinguishes itself by its ability to engage in conversations with a learner (Veletsianos & Russell, 2014). The studies related to PAs mainly focus on how the agent can contribute to the objective of learning, or how different characteristics related to appearance or personality influence the learning outcomes. Mostly, PAs are embedded in virtual environments. However, some studies have explored PAs in the form of physical robots installed in classroom settings (Kanda, Hirano, Eaton, & Ishiguro, 2004), which more closely resembles how we employed the CA in this project.

Gulz et al. (2011) explain that the origins of PAs were called Intelligent Tutoring Systems (ITS), which first appeared in the 1970s, "the *classic* ITS was an impersonal, non-social and abstract pedagogic agent whose sole purpose was to tutor" (Gulz et al., 2011, p. 130). These systems could assess learning progress, answer or ask topical questions and provide feedback. Since then, advances have transpired relating to aspects such as visualization, embodiment, and interactional capabilities. However, the most noticeable shift concerns the different social roles assigned to the agents. The domain of PAs is multidisciplinary, and a prominent perspective is the Computers as Social Actors paradigm, asserting that people respond and treat machines in a similar manner as they would humans (Veletsianos & Russell, 2014, p. 760). Multiple studies explore the various social roles taken on by a PA, such as tutor, motivator, companion, mentor, etc. (Gulz et al., 2011; Kanda et al., 2004; Ryokai, Vaucelle, & Cassell, 2003; Sun, Leite, Lehman, & Li, 2017).

We proceed by providing some examples of studies relating to the employment of PAs. Kanda, Hirano, Eaton, and Ishiguro (2017) examined interactive robots as social partners and peer-tutors for children. Their research objective was to explore how robots can form relationships with students in order to fulfill their role as tutors. Their main findings stress that for human-robot relationships to emerge, the robot must inhabit basic social competencies and possess the ability to learn so that the relationship has a natural progression. Kim, Baylor, and Shen (2007) studied PAs as learning companions in a virtual environment and more specifically how emotional interactions influence the agents' instructional impact. The study affirmed the utility of emotion and empathy as resources in educational contexts.

Finally, we bring up Gulz et al. (2011) who argue for a design-based approach to PA-research. They comment on the challenges they faced when designing agent-based educational software by providing examples from a project in which a conversational PA taught mathematics to children via a virtual game. Their concluding remarks contain a guiding framework for the design of PAs and a discussion of challenges relating to PA-design. The four challenges they identified were all connected to their goal of developing a "virtual human": (1) handling student expectations of the agents knowledge, (2) handling student expectation of the agents social skills, (3) handling student over-engagement in off-task conversations, and (4) handling abusive comments from the student. They argue that these are challenges others will likely face when designing conversational PAs.

2.1.3 Children as users of conversational technology

As we have previously stated, our target group comprises of children ages 4-6. Consequently, we have sought out studies related to children's interactions with different conversational technologies. We found that a commonality in the studies were findings pertaining to the strategies for communicative repair applied by the children in situations of miscommunication. We give a more detailed account of the concept of *communicative repair* in the chapter on dialogue as a design material. The following section describes the main findings from the studies on child-agent interaction.

Cheng, Yen, Chen and Hiniker (2018) examined how children repair conversations in interaction with a voice interface, as well as the parent's supporting role during the interactions. Druga, Williams, Breazeal, and Resnick (2017) conducted an initial exploration of child-agent interaction focusing on how the children perceived the technology. Both the former and latter studies found that the children applied different strategies to make themselves understood by the agents. Both studies point to rewording and varying loudness of voice as important repair strategies. However, Cheng et al. (2018) found repetition to be the most common repair strategy, while Druga et al. observed the children punctuating words within sentences to make the message clearer.

In addition to findings related to repair, Druga et al. (2017) present key findings connected to perceived intelligence, identity attribution, playfulness, and understanding. The children related the agent's intelligence to what it seemingly "knew" about their own

topics of interest. Additionally, the researchers observed probing into the identity of the agents, as well as the children playfully testing the limits of the agent's capabilities. The findings from the study suggest that voice and prosody affected the perceived friendliness, and that "mobile and responsive agents appeal to children" (Druga et al., 2017, p. 599). In their concluding remarks, the authors suggest that there is a need to explore how we can transfer the facilitation provided by adults during the interactions into the design of the agent.

Cheng et al. (2018) found that the children were persistent in attempting to interact with the prototype and that they rarely asked for help or showed frustration when it failed to work. The adults supported interaction in a variety of ways. They started by suggesting ways to remedy the interaction. Next, they would commonly intervene and attempt to interact with the technology themselves. Ultimately, they resigned and discontinued interaction.

2.1.4 Distinctions within conversational technology

We conclude this section about CAs by clarifying some important distinctions within the realm of conversational technology. Firstly, Jurafsky and Martin (2018) separate CAs into two general classes, the first being *task-oriented dialogue agents*, which provide interactions in the form of short verbal exchanges in which they collect information from the user in order to complete a given task. Conversely, *chatbots* are CAs that engage the users in lengthy conversations, imitating the natural feel of dialogue between humans. The mode of communication through which the CA interacts with its users can be either text-based, verbal, or both and it can be multi-modal or not. Secondly, dialogue design in CAs can be *open*, providing the user with control of the conversational activities, actions, and topics. Conversely, there are *scripted* CAs, which directs the conversation by providing the user with a limited set of options and successively guides them through a specific dialogue flow. Thirdly, some CAs are *embodied*, they have some sort of physical form representative of the agent either in the virtual (e.g. avatars) or physical world (e.g. robots), and others are simply an *abstract* voice interface (e.g. Apples Siri). Lastly, agents can inhabit different roles in the interactions in which they take part and therefore display different personalities. Hence, designing a suitable and consistent persona is important. Designers need to focus on behavior, creating a compelling tone-of-voice, sensibly selecting words, and deciding what role the CA should play in the user's lives. The

continuing focus in this thesis is organization and interaction through dialog. Thus, the succeeding section describes how we might understand dialogue as a design material.

2.2 Dialogue as a design material

The organization of conversation is the foundation upon which interaction with CAs are built. Conversations are transitory cooperative language transactions governed by intricate rules and agreements that unfolds when the participants have the same perceptual or mental focus (Ninio & Snow, 1996; Sacks et al., 1974). Beyond the words that are uttered, social cues, non-verbal indicators and context, dictates the construction of meaning between multiple actors. There are several disciplines devoted to the study of the organization and structure of language and conversations (linguistics, semiotics, pragmatics, conversation analysis, cybernetics, etc.). Although they all bare relevance to our goal of designing for a dialog-based system, describing them all in meaningful detail is too big an undertaking. However, in relation to our thesis, it is pertinent to explain our thoughts on what designing dialogue entails. What are the implications for the design endeavor and what do we need to know about conversations to be able to handle dialogue as a design material properly?

We continue by describing some important issues concerning the design of conversations. We comment on the different functions of conversations in social life and on how conducting a conversation is a cooperative endeavor. We proceed by accounting for important principles and building blocks in achieving an effective organization of communication before we continue by describing the organization and the importance of communicative repair. We end this section by pointing to two significant challenges pertaining to designing dialogue for CAs, namely paralinguistic as well as enabling correct assessment of intelligence and capabilities.

2.2.1 The different functions of conversations

We find that understanding the conversation and its functions in social life are central to the design of CAs because we need to understand what motivates people when they engage in conversations with machines. So, why do we talk? Conversations connect people. They are a focal point of human communication and thus often referred to as the most intuitive interface there is. Conversations are also a resource for action in that

people fulfill their objectives through the act of talking. They are useful because they provide us with the ability to learn something new, enable us to coordinate our actions in a mutually beneficial manner and support collaboration through facilitating discourse and the convergence of objectives (Dubberly & Pangaro, 2009). However, conversations are not merely vehicles of information propagation or a means through which people accomplishes practical work. There are many genres of conversations, such as stories, instructions, lectures, explanations, jokes, and so on. What types of conversations we apply to a situation depends on what we want to accomplish.

There are many requirements that must be fulfill in order for communication to take place. Roman Jakobson (1960, p. 353) was a prominent linguist who described the process of effective verbal communication as an *addresser* sending a *message* to an *addressee* and argued that for the message to be operable it required a *context* that must be sizable; the receiver must be able to perceive it. Moreover, the message contains a *code*, which must be at least partially common to the participants (encoder and decoder). Finally, what he calls a *contact*, a physical channel and psychological connection, must be maintained in order for the addresser and the addressee to uphold communication. Jakobson is best known for utilizing this model of communication to establishing six functions of language. They are as follows:

1. *The referential function* concerns descriptive statements, which correspond to the context and refer to phenomena in the real or imagined world.
2. *The expressive function* reveals information about the addressers internal state
3. *The conative function* engages the addressee, e.g. gets their attention.
4. *The poetic function* describes language for languages sake, e.g. language in poetry.
5. *The phatic function* refers to the language utilized as a means to obtain social connections with others.
6. *The metalingual function* refers to when language is used to discuss language itself.

According to Jakobson (1960), effectiveness in conversation relates to how well the participants communicate in order to reach their objectives. However, the objective may differ depending on the underlying motivation for initiating contact with other actants. Baron (2015) accounts for essential motivations behind most conversations by mentioning three key rationales: *pragmatics*, *social companionship,* and *control*. Firstly,

the pragmatic function of a conversation manifests when participants communicate information as a means to get something done. This is similar to Jakobson's *referential function*. Secondly, Baron proceeds by stating that humans are social beings and that we use talk because we seek social interaction, similar to *the phatic function* of language. Hence, not all communication vehicles information, but is rather designed to keep the channels of communication open. Lastly, people use language to assert control. She defines "control" as directing the course of the conversation, and emphasize that this can manifest in numerous ways (Baron, 2015, p. 3). Baron goes on by discussing how these motivations affect dialogue design and requirements concerning conversational proficiency in robots. Digital voice assistants are primarily meant to fulfill a pragmatic purpose through conversation and are required to be excellent conversationalists in order to be effective. However, does this also apply to robots that fulfill various social needs? Baron imagine that robots we interact with based on a social motivation can inhabit less idealized conversational skills because displaying ordinary and imperfect language would not prevent it from providing companionship. Regarding control, however, people would probably accept neither diminished autonomy in the private sphere nor a completely autonomous CA that controlled the conversation by disregarding the user's intents in utilizing the machine in the first place.

Dubberly and Pangaro (2009) also comment on the need to design for *effective* conversations. They define dialogue as a "progression of exchanges among participants" (Dubberly & Pangaro, 2009, p. 1) and describe an *effective conversation* as an interaction in which the participants are changed in a way that has lasting value to them. This might entail getting to an agreement, reflecting upon a topic, or simply developing social bonds and having fun. The participants might also agree to perform an act beyond the conversation that has taken place (Dubberly & Pangaro, 2009). In the design of CAs, this entails creating a conversation that is of value to the user, e.g. them gaining new knowledge, extending their perspective, or having an enjoyable experience.

2.2.2 The conversation as a cooperative endeavour

In the book "Studies in the Way of Words", philosopher of language Paul Grice (1982) articulated a set of maxims that are now recognized as guiding principles in effective interpersonal conversations. He suggested that the foundation of regular conversations is a general *cooperative principle*, which entails that every participant must contribute for it

to be successful. Regarding the development and design of CAs, this highlights the cooperative nature of the practice in which it is to take part. The four principles Grice proposes as guidelines for becoming a competent social communicator is as follows:

- *Quantity* is contributing statements that are as informative as the situation requires. Providing too much or too little information is unhelpful. The actors have to assess and possess the amount of information their conversational partner will find adequate.
- *Quality* pertains to truthfulness and involves transparency concerning one's motives and objectives, as well as not portraying something as true if one is, in fact, uncertain about its legitimacy.
- *Relation* concerns being relevant to the purpose of the conversation. Hence, one should be appropriate and correctly assess when to provide more information or when it is suitable to let others speak. In short, be context-aware.
- *Manner* connects to being specific, brief and unequivocal in one's statements.

If participants do not comply with the abovementioned guidelines, the communication might collapse, or the exchange may be unsuccessful. Designing dialogue that is inconsistent with Grice's maxims would entail withholding information, not telling the truth, straying from the topic of the conversation, conveying information inefficiently, or being ambiguous. In real life conversations, people frequently fail to abide by these principles, thus, conversations between people are often inefficient. Hence, being a perfect conversationalist is challenging because situated language is improvised and far from idealized. Regarding CAs however, we as designers have the opportunity to consider the implication of these maxims closely concerning the replies and prompts posed to the users.

Because of the less than idealized way people speak in ordinary situations, Erika Hall (2018) argues that literal interpretation and application of natural human conversations do not necessarily lead to well-designed CAs. She uses Grice's conversational maxims as groundwork for establishing several design principles for human-like conversational interfaces and proposes that in order to create valuable exchanges with users, a CA should be cooperative, goal-oriented, context-aware, quick and clear, turn-based, truthful, polite and error-tolerant. Furthermore, Hall describes four key moments that are essential when developing a user's relationship with a system: *introduction*, *orientation*, *action,* and

guidance. During the *introduction*, the identity and potential value of the CA should present itself. This is the moment where the CA should communicate its usefulness and attempt to establish trust, interest and an emotional connection with the prospective user. During *orientation*, the CA should provide the user with appropriate contextual information and offer the right choices at the right time. When in the *action*-phase, the CA must be able to interpret the user's intent and provide feedback, so the users know if the action moved them closer to their objective. The sequence of interactions should be logical, and the users should perceive the exchange as cooperative, clear, and non-ambiguous. Lastly, opportunities for *guidance* might present itself in many ways, the most relevant one being contingency messages where the CA should help the users get back on track if an error has occurred.

2.2.3 Children in conversations – communicative competencies

Liv Gjems (2009, p. 46) presents sociocultural perspectives on language and learning, and states that language and conversation are essential tools for children to learn about both concrete and abstract phenomenon and events. Conversations are crucial for children to understand the complex situations in which they find themselves every day and is immensely influential on their understanding of the social and cultural world. Children begin to appropriate conversational competencies early on. However, they do not necessarily adhere to conversational conventions the way adults do. They might therefore have trouble finding words and phrases, as well as maintain conversational topics or comprehend their conversational partners understandings and experiences (Gjems, 2009, p. 64). Ninio and Snow (1996) comment on several skills that children have to acquire in order for them to function as conversational partners. They must learn that in conversations people expect you to answer quickly, and that frequent interruptions and long pauses disrupts the conversational flow. They must understand the important workings of responsiveness in conversation; when someone speaks to you, they usually expect a reply. Moreover, as a listener you are obliged to remain attentive and express either comprehension or incomprehension to what others say. Furthermore, people expect you to keep your response relevant to both the theme of the conversation and to the last speaker's utterance, something children often struggle to master. Lastly, children have to understand and apply strategies for communicative repair if needed. Because it takes time to become a competent conversationalist, children often require cooperative adults to help

them manage and compensate for the problems they experience during spoken interaction. Children learn and develop these competencies in conversations with adults who support and guide them.

2.2.4 Central concepts from conversation analysis

Conversation analysis is an approach to the study of social interaction and talk-in-interaction that have contributed several descriptive concepts that are useful for analyzing interactional structures (Richards & Schmidt, 1983). Designing dialogue requires us to obtain a basic understanding of the organization of ordinary conversation. Moore et al. (2017) argues that current conversational agents such as Siri and Cortana often produce disappointing user experience because user experience designers lack a technical understanding of the structures of natural conversation:

> *"Although it is easy to get a system to produce words, none of the current agents or bots display general conversational competence" (Moore et al., 2017, p. 492)*

He contends that understanding how conversations are organized will aid designers in the practice of creating a dialogue that feels natural to the users. Thus, we find that designing dialogue requires a basic understanding of the structure and logic of conversations. Next, we describe the organization of *turn-taking* in conversations, the *sequence* as an interactional component in conversations, and lastly the most common form of conversational patterns, *adjacency pairs*.

The established code of conduct regarding turn-taking in conversations is presently an axiom in exchanges involving a CA. In the practice of turn-taking, the participants speak one at a time in alternating turns and produce talk that is inspectable and inspected by co-participants (Sacks et al., 1974). The building blocks that make up a turn are called *turn-constructional units* (TCUs) and can be sentences, clauses, phrases or lexical items (Schegloff, 2007). Sentences are complete and descriptive ("I am making dinner for you"), while clauses are less descriptive but can stand alone ("I am making dinner"). Phrases are incomplete sentences, which may work as a response to an inquiry ("in the house", "a rainy day") and lexical items are one-word statements ("yes", "here"). Central to the organization of turn-taking is how turns are distributed amongst participants and

how transitions between speakers occur. When a TCU draws to a close, a transition-relevant point occurs, by which time the next speaker might begin to construct a TCU in response. There are three *allocation mechanisms* that govern the transition between speakers: (1) the producer of the current TCU selects the next speaker ("what do you think, Eric?"), (2) the next speaker appoints him or herself ("I think …"), or (3) no one self-selects and the current speaker continues talking until the next transition-relevance place occurs (Schegloff, 2007, pp. 3–4).

The smallest interactional component of human-to-human conversation is *the sequence* – a universal pattern we can apply to a variety of intents, settings, and situations (Schegloff, 2007). Schegloff describes it as "the vehicle for getting some activity accomplished […] some course of action gets initiated, worked through, and brought to closure" (2007, p. 20). Thus, a sequence is a grouping of turns relating to the same course of action. In this context, an action refers to "what that bit of talk appears designed to do" (Schegloff, 2007, p. 8) such as agreeing, answering, disagreeing, teasing, confirming, requesting, noticing, etc. One TCU can perform more than one action, or it can be the vehicle for the following action. Thus, sequence organization is important for understanding how we arrange actions in relation to each other through turns at talk.

Finally, a*djacency pairs* are a crucial pattern around which numerous sequences are organized (Schegloff, 2007). In its most basic form, an adjacency pair consists of two turns composed by different speakers, one after the other, differentiated into *first pair part* (initiation of some exchange) and *second pair part* (responsive to the first part), and lastly categorized into *pair types* such as greeting-greeting, question-answer, offer-accept/decline (Schegloff, 2007, p. 13). Adjacency pairs are flexible patterns, and participants can elaborate upon the base sequence in a multitude of ways. Conversation analysts categorized the expansions in *pre*-, *insert*-, or *post* expansions depending on when they occur during the sequence (Schegloff, 2007). Expansions serve different purposes but mainly support the participants in coordinating and balancing irregularities in the conversation. Thus, *screening* serves as a preliminary inquiry ("do you know any restaurants?"), *repeating* aids mutual understanding ("a few *what*?"), *closing* ends a particular sequence ("ok, thanks"), *eliciting* brings forth additional and required information ("what kind of food do you want?"), and finally, *paraphrasing* clarifies the

meaning behind an utterance that was deemed inadequate in building shared understanding ("what do you mean?") (Moore, 2018, pp. 184–185).

2.2.5 When things go wrong – communicative repair

Knowing how to design for different kinds of communicative repair is pivotal when developing well-functioning CAs because mutual understanding is a perpetual prerequisite for effective communication. Constructing meaning is a collaborative activity in which the participants monitor and interpret the statements uttered and subsequently act out an appropriate response (Schegloff, 2007). The deceivingly effortless ability of jointly building understanding through spoken interaction is hard to emulate and is one of the main challenges of conversational design. It requires the participants to inhabit basic competencies related to language, cognition, social rules of conduct, and contextual awareness. Discrepancies in the participants' understanding of the subject matter or intents lead to misunderstandings and might ultimately cause a communicative breakdown.

We proceed by relating some important concepts in communicative repair-processes. Conversational or communicative repair is loosely defined as the work done by speaker, hearer, or both, to correct an unsuccessful part of the communication (Richards & Schmidt, 1983, p. 147). It is a generic term that includes the correction of a variety of phenomenon ranging from overlapping talk and other errors related to the turn-taking organization of conversation, to ambiguity or corrections of factual errors in an utterance (Hutchby & Wooffitt, 1998). Repair procedures comprise of three components: the trouble source, repair initiation, and repair solution (Hoey & Kendrick, 2017). The *trouble source* or *repairable,* are terms referring to the point in an exchange that instigates a need for repair (Schegloff, Jefferson, & Sacks, 1977). It is not a given that the outcome of all repair sequences is mutual understanding and in order to design robust dialogue, it is important to recognize what might instigate a communicative breakdown. In natural conversations, there is a range of trouble sources necessitating repair, such as choosing the wrong words, interruptions, mispronunciation, mishearing, slips-of-tongue, misunderstanding, and so on.

Conversation analysts categorize repair-types based on whom initiates the repair sequence as well as who carries out the repair. Repair initiated by the speaker of the repairable is

called *self-repair,* whilst *other-repair* is elicited by any party other than the speaker (Schegloff et al., 1977, p. 363). Thus, there are four repair types. The speaker refers to who produces the repairable and the recipient is the listener(s):

- *Self-initiated self-repair*: the speaker both initiates and carries out the repair.
- *Other-initiated self-repair*: The speaker carries out repair initiated by the recipient.
- *Self-initiated other-repair*: the speaker attempts to get the recipient to carry out repair, for instance, if one cannot remember something.
- *Other-initiated other-repair*: This is closest to what we know as correcting others, as the recipient both initiates and carries out the repair.

We can apply different strategies to indicate the need for reparative efforts as well as to conduct the actual repair. Incomprehension is communicated in several ways; we can echo words or phrases to signal incomprehension ("milk?"), certain TCUs act as markers for lack of understanding (*huh*, *what*, *sorry*, etc.), and paralinguistic cues such as facial expressions, prosody and intonation, and gestures indicate confusion. After repair initiation, one of the participants must conduct the actual repair. Research have uncovered three overarching categories of communicative repair: (1) *repetition*, saying the same thing over again (2) *augmentation* such as speaking clearer or louder to get the message across, and (3) *substitution* which entails finding alternative words or courses of action to express the same sentiment (Golinkoff, 1986). Participants in a conversation mostly carry out repairs in relatively close proximity to the trouble source, as it would otherwise involve a cognitive and organizational challenge regarding the alignment of topical and referential coordination. It would create trouble concerning the shared understanding of the content of the conversation if someone was to initiate or conduct repair long after the repairable was produced (Hutchby & Wooffitt, 1998, pp. 64–66). Integrating adequate competencies concerning communicative repair is a central challenge in CA-design as well as a focus area in this thesis. We conclude this section of the chapter by accounting for two additional challenges in CA-design, paralinguistic communication, and perceptibility of the agent's capabilities and intelligence.

2.2.6 Additional challenges in the design of CAs

We will briefly comment on the aspect of paralinguistic communication as well as challenges related to conveying the agent's level of intelligence and capabilities correctly.

Paralinguistic communication

Significant to designing CAs is addressing what we cannot convey solely via spoken or written language. Therein lies the paralinguistic aspect of communication (Baron, 2015). We refer to two paralinguistic categories: the way we say something, e.g. prosody, tempo as well as tone and pitch of voice, and aspects relating to communication with the body, such as posture, gestures, and facial expressions. These are all communicative aids that contribute to the process of meaning-making. When users engage in conversations with CAs, important information is potentially lost as the agents is often unable to make use of paralinguistic cues to construct meaning; it is unable to adapt its prosody and body language in a manner consistent with the emotional tone of the conversation because it is not able to perceive nor send such communicative signals. Paralinguistic communication is a challenge concerning CAs that rely exclusively on verbal communication. Thus, this fact assigns even greater importance to the clarity, efficiency, and effectiveness of the dialogue. Alternatively, the CAs hardware or physical form must contribute to its communicative efforts.

Assessment of intelligence and capabilities

Many studies show that the need for communicative repair in human-CA interaction largely derive from the difficulties people experience when assessing capabilities and level of intelligence. Watt (1968) introduced the concept of "habitability", which describe peoples inclination towards presuming that computers have complex conversational competencies after exhibiting elementary skills. Luger and Sellen (2016) conducted a study examining the interactional factors that affect the everyday use of CAs. They discuss important areas of consideration when designing CAs. How to reveal system intelligence is one of these areas. They found that people who were technically skilled were usually able to adjust their mental model of the system in accordance with its actual capabilities. Those with less technical knowledge, however, did not alter their expectations as to what the system was able to do, which led to frustration when these expectations were not met. The authors state that this indicates the need to scaffold the user's expectations through the gradual revelation of intelligence. Another area of

consideration is how to indicate capability through interaction. Often, the CAs limited capabilities become visible when failures occur. Thus, the users learn what the CA is unable to do. Luger and Sellen argue that designers should consider how to convey system limitations and capabilities in moments not related to the failure of a task.

Moreover, the CAs physical or virtual form can affect and contribute to the users understanding of the CA's intelligence and capabilities because physical attractiveness and aesthetics can have a significant impact on social influence (Fogg, 2002, p. 93). "The halo effect" refers to how people assign admirable qualities to someone based on their appearance. Fogg states that technological artifacts can benefit from the halo effect in that users perceive them as intelligent, capable and reliable. However, it is crucial that the user's perception of the CAs level of intelligence does not exceed its actual capabilities. Those involved in the practice of CA and dialogue design comment on the importance of handling user expectations regarding the CAs capabilities and intelligence. We revert to Luger and Sellen (2016) who found that when the CA used humor, the users' expectations about the system capability increased because it contributed to framing the system as anthropomorphic. Luger and Sellen refer to Shedroff and Noessel (2012, p. 193), who suggest that framing a system in an anthropomorphic way will raise the user expectations about the systems capabilities, intelligence, language, judgment, autonomy, and compliance with social norms.

2.3 Graceful interaction and robustness in CA design

The preceding sections continue to provide an outline of the interdisciplinary and complex endeavor of designing CAs, with a focus on how to create robust dialogue without compromising the natural and human-like feel of the interaction. Human-to-human communication is simultaneously fragile and robust; misunderstandings and difficulties frequently arise, however, the participant's ability to make use of the communicative resources available to them as well as inhabiting adequate conversational competencies ensure that troubles are corrected. The question is whether it is possible to emulate the inherent robustness in human conversations when designing human-CA interactions. We continue by describing Hayes and Reddy's (1983) principle of "graceful interaction" before we provide some definitions of robustness in order to lay the groundwork upon which we base the concluding discussion of this thesis.

2.3.1 Graceful interaction – an idealistic concept?

Hayes and Reddy (1983) establish the concept of *graceful interaction* in conversational technology. Graceful interaction involves dealing appropriately with any input from the user by conducting the dialogue as human-like as possible. The user should be able to talk to the system as they would to another human, and the system should emulate this behavior in return. Hence, the strategies the system applies to clarify incomprehension or resolve ambiguity in the user's utterance must be the same as the strategies people normally use. However, this is challenging because the technology's ability to detect causes of incomprehension and appropriate strategies for repair is limited. Hayes and Reddy (1983) describe a set of abilities and behaviors they find to be essential for graceful interaction to occur between human and computers. We continue by presenting these, as they suggest how to design for constructive repair implementation and thus avoid strenuous repetitive sequences.

Three obstacles may arise in human-to-human interaction when attempting communication: the listener may not receive the message, the listener may be unable to interpret parts or the entire message, or the listener may misinterpret the message. Thus, the term *robust communication* is defined as "the set of strategies needed to ensure that a listener receives a speaker's utterance, and interprets it correctly" (Hayes & Reddy, 1983, p. 233). The occurrence of the abovementioned obstacles are common, but human-to-human interaction is robust because of people's abilities to detect and remedy such troubles in the communication (Hayes & Reddy, 1983, p. 269). Consequently, conversations between people rarely break down completely. Hayes and Reddy (1983, p. 235) do not find robustness to stem from elimination or minimization of the abovementioned errors or obstacles, but rather characterize robustness as comprising the techniques we use to detect, recover from, and correct errors. They mention four techniques to serve this purpose: *implicit acknowledgments, explicit indications of incomprehension, echoing* and *fragmentary recognition*. If graceful interaction is to happen, the machine has to be able to both use and react correctly to these techniques.

Implicit acknowledgment: If the receiver fails to produce a reply to a message, it indicates that the message has been lost. Hence, graceful interaction entails producing responses quickly. If the system needs time to process data, it should indicate this through time-filling utterances like "hold on just one moment". The system should enable the user to do

the same and provide an appropriate response if the situation occurs. In instances where the user fails to respond promptly, the system should attempt to reestablish connection ("can you hear me?").

Explicit indications of incomprehension: the receiver should explicitly state incomprehension. This can be done in an uninformative way, e.g. "I beg your pardon", which does not indicate to the sender exactly what was not understood. Rather, the receiver should indicate precisely what they did not understand, as it increases the likelihood of a successful repair. It is important that the system implies what it did not understand, but also enables the user to clarify in a way that the system will be able to interpret. Hence, if the system failed to understand a specific word, it is unwise to ask the user to explain if it is incapable of interpreting the explanation. Instead, the system should try to interpret what the user meant based on the sentence and reply with a question pertaining to the users' wishes. They provide the following example; if the system did not understand the word 'extension' in the utterance "what is the extension for Jim Smith?" it can ask "Do you want Jim Smith's number or his address?".

Additionally, Hayes and Reddy (1983) establish *fragmentary recognition* and *echoing* as strategies for achieving graceful interaction. They base the former on the fact that not all words in an utterance are important and propose that the system should identify the words in the utterance that are most significant and echo a response that incorporates those words. The system may have been unable to identify all words in the following utterance but understood the place and the date and deemed them important for the objective at hand: "I am interested in going to Pittsburg on the 17th of May". The place and date are used to create a response: "at what time on the 17th of May did you want to go to Pittsburg?" Similarly, the latter is a strategy for eliminating the tediousness of clarifying questions by the system repeating or guessing the parts it is unsure about: "*what is the number for (?)ter Smith?" - "Walter Smith .... His number is 5592".* Hence, the conversation has a higher likelihood of progression and the user can correct the system if it misinterpreted their intent.

2.3.2 Robustness in interaction design

We proceed by describing some definitions of the concept robustness. Robust can be defined in general terms as something being "strong and unlikely to break or fail"

("ROBUST | meaning in the Cambridge English Dictionary," 2019). There are numerous definitions of what it entails to *be* robust. Within the field of computer science, we relate robustness to different aspects regarding the technology, e.g. hardware, software, and design. One way of understanding robustness is related to the system's ability to cope with errors and to handle erroneous input ("Robustness (computer science)," 2018). Pertaining to interaction design, we can understand the concept of robustness as how Dix, Finlay, Abowd, and Beale (2004, pp. 260–273) present it. They find robustness to pertain to the level of support provided to the user in order for them to successfully achieve their goal. They go on to describe four principles they find to affect robustness, *observability, recoverability, responsiveness* and *task conformance*.

- *Observability*: The ability of the user to evaluate the internal state of the system from its perceivable representation in the interface.

- *Recoverability*: The ability of the user to take corrective action once an error has been recognized. The degree to which the user knows which strategies for repair to use and how to use them is essential for a robust design.

- *Responsiveness*: How the user perceives the rate of communication with the system. Instantaneous response times is desirable. If this cannot be achieved, it is important to indicate to the user that the system has received a response and is working on creating a reply.

- *Task conformance*: The degree to which the system services support all of the tasks the user wishes to perform and in a way that ensures that the user understands them (Dix et al., 2004, pp. 270–273).

The lack of robustness in any system will most likely lead to a breakdown in the interaction. In relation to robustness, we contend that an understanding of what instigates breakdown is essential, as we have already explained that communication is often fragile and prone to trouble and misunderstanding. Thus, we dedicate the next section of this chapter to different perspectives on breakdown within the fields of technology and design.

2.4 Understanding breakdown

Breakdowns are well-known phenomena generally describing situations where technologies fail to work as intended. In every design endeavor, it is important to remember that breakdown is an inevitability of prolonged use – it is impossible to avoid breakdowns all together as ideal situations of use rarely manifest. Consequently, aiming to design to eliminate the occurrence of errors is futile. Rather, designers should put their effort into exploring what it means for a technology to fall apart gracefully so that both users and system can easily recover from glitches in interaction. Next, we provide an overview of some prevailing perspectives on the topic of breakdowns within the domain of technology and design. This is done in order to support the concluding discussion of this thesis in which we discuss the merits and faults of making breakdowns the starting point for understanding technology in use.

2.4.1 "Broken world thinking" – foregrounding breakdown and repair

In the article "Rethinking Repair", Steven Jackson (2014) contends that "the world is always breaking", and attempts to instigate a move towards a repair-centric view within the domain of technology and media. He suggests a perspective named *broken world thinking* in which he proposes to make disintegration the starting point for development, thus urging us to recognize the value of breakdown and repair as engines of invention. Jackson's point of departure originates from what he calls a "productivist bias", a bias that downplays and renders restorative activities invisible, therefore discarding their intrinsic value in the life cycle of objects and artifacts. Currently, we position innovation as both a primary activity as well as being of superior value. Jackson criticizes this ordering and states that innovation can, in fact, be born from breakdown. Moreover, he proposes that *broken world thinking* gives us the ability to study how objects are a part of the social world as well as how "ongoing forms of labor, power, and interest" affect the endurance of things (2014, p. 230). For instance, if repair is not beneficial to the producers and manufacturers of technological devices, its design might reflect this and perpetuate the position that forefronts innovation and the new as being more valuable than repairing and maintaining the old. Consequently, this perspective contributes to discern the inherent value of breakdowns as revelatory moments within the lifespan of

technological objects – it positions breakdown as the moment of origins for something new as well as an inevitability of use.

2.4.2 Breakdowns as revelatory moments in technology use

Can breakdowns be a good starting point for eliciting insights regarding technologies and technology use? There are many kinds of seeing in design and many strategies to make different aspects of a system or a technological artifact visible. To provide a few examples, Gaver (2003) has explored the role *ambiguity* can play as an element in design by provoking reflection. Furthermore, Schön's *reflective practitioner* has a central role in the development of design theory as he asserts that design situations inhabit "back-talk" in which "the designer reflects-in-action on the construction of the problem" (Schön, 1983, p. 79). Similarly, within this vast domain, there are different ideas concerning what we might learn in or from moments of breakdown. For instance, can breakdowns serve as an instance of back-talk? Jackson (2014, p. 229) asks if the fixer or repairer on account of their vantage point is party to insights and knowledge that lies beyond the reach of designers and technologists. If this is the case, breakdowns are meaningful because they contribute to uncovering the inner dynamics of artifacts, processes, and infrastructures – they herald the unwrapping of the black box.

Heidegger, a German philosopher of the 20th century who provided contributions within an extensive range of domains, is amongst the antecedents of perspectives on breakdowns as opportunities for seeing. He was primarily engaged with uncovering what it means to be *in* the world (Susi & Ziemke, 2005). Within the field of technology and design, his views on object perception and tool use are widely applied. Heidegger believed that functionality was a defining characteristic of equipment, but that functionality only manifests in use, or in the context of meaningful activity. Hence, we should always consider artifacts in relation to their "equipmental whole", meaning that their function is always connected to other human and non-human actants. His prime example was that a hammer is a hammer only in relation to the activity of hammering and in relation to the nail, which in turn holds other objects together. Concerning breakdowns, Heidegger's notion of how tools are in the world as either "ready-at-hand" or "present-at-hand" supports the thought that breakdowns are revelatory moments in which the object itself becomes visible because a functionality fails to appear as expected. When technologies work as intended it is "ready-at-hand" and we do not contemplate the thing explicitly

when using it – rather we focus on the work we are doing. However, when something fails to work as expected it becomes "present-at-hand", the qualities it possesses as a physical thing is emphasized and become the object of examination. Breakdowns precipitate transitions between "readiness-at-hand" and "presence-at-hand" and thus instigate reflection on the thing itself (Susi & Ziemke, 2005, pp. 8–11).

One possible application of Heidegger's reasoning is implemented in Star and Ruhleder's (1996) investigations of the interconnectivity of human and non-human entities embedded in infrastructures. They define infrastructures as relational and emergent of *in situ* practices. Therefore, infrastructures manifest differently to different people. They argue that an inherent quality of infrastructures is its "visibility upon breakdown" as failures occasion transitions to "presence-at-hand"; the repair work one must carry out on different entities of the infrastructure highlights the connections, actants, standards, procedures, processes, and artifacts of which it is constituted. Hence, this perspective highlights how breakdowns provide opportunities for gaining a deeper understanding of the technology in use and supports Jacksons (2014) claim that the repairer is privy to a unique outlook on the technology as well as the interconnected contexts in which they are situated.

2.4.3 Difficulties of repair in human-computer interaction

Inconsistencies between user intention and machine interpretation often cause breakdowns, which precipitates repair work. Thus, repair is an essential aspect of human-computer interaction. Suchman (2007) accentuated this in her work "Plans and situated actions" in which she argued the futility of modeling technology on inadequate conceptions of reality; plans are not prescriptive of action as actions are situated and dependent on the current context of use. She utilized concepts from conversation analysis as a means to study interactions at the interface of a machine and conveyed the profound difficulties in interactive interface design stemming from asymmetries between humans and machines as interlocutors. Suchman contends that the ability to conduct communicative repair is essential in human-to-human interaction, but that machines lack adequate abilities to detect the situational cues that signal the need for such action (2007, pp. 11–12). Machines have minimal sensory input and access to the peripheral world in which they are embedded (2007, p. 37), thus there is a discrepancy between the human

and the computer's ability to access and make use of contextual and situational information in order to detect and initiate repair.

The act of planning action before it occurs is referred to as an "imaginative and discursive practice" (Suchman, 2007, p. 13), thus, designing to prevent or recover from breakdowns is in many ways a predictive activity as designers must imagine what might go wrong in any particular situation. Fischer (1994, p. 222) stated that "a cooperative agent needs to understand the nature of open problems, the intentions of the problem solver, and the fact that goals are modified during the problem-solving processes". This indicates that the process of recovering from breakdowns is cooperative but also unpredictable and that the machine should be able to comprehend the nature of the breakdown as well as the user's strategies and changing intentions while solving it. As mentioned, this is precisely the ability Suchman believes to be lacking in machines; the resources available to the parties in the process are decisively different, thus creating asymmetries in human-computer interactions. Thus, we argue that there is a need to understand how to address these asymmetries in order to design robust dialogue for CAs. Primarily, we want people to experience the CA as a proficient interlocutor that can help them achieve their objectives.

2.5 Summary – key theoretical concepts

The importance of the conversation as an instrument in social life is indisputable and it enables us to fulfill many different needs, ranging from the pragmatic to the social. Providing that CAs possess adequate conversational competencies, they might take on various roles in its user's lives. In the second edition of "Plans and situated actions", Suchman questions the conversation as an effective metaphor for the interaction between people and technologies (Suchman, 2007, p. 23). We find that the application of conversational technologies negates this claim, as the conversation becomes the paradigm for how we organize interaction between users and computers.

To conclude this chapter, we would like to summarize the conceptual framework of this thesis. Thus, we reiterate the six main theoretical perspectives and concepts that we applied in our research. They are the following: the cooperative principle of conversation, conversation analysis, repair, graceful interaction, robustness and broken world thinking.

The cooperative principle in conversation

It is important to bear in mind that conversations are a cooperative endeavor, meaning that in order for communication to be efficient, all participants must contribute to the mutual intelligibility of dialogue by adhering to the maxims om *quantity*, *relation*, *quality* and *manner* (Grice, 1982). *Quantity* is contributing statements that are as informative as the situation requires. *Quality* pertains to truthfulness and involves transparency concerning one's motives and objectives. *Relation* concerns being relevant to the purpose of the conversation. *Manner* connects to being specific, brief and unequivocal in one's statements.

Conversation analysis

Conversation analysis is a research domain that focus on the study of spoken social interaction, and it has contributed several descriptive concepts related to the organization of ordinary conversation (Richards & Schmidt, 1983) We have chosen to focus on three central concepts. Firstly, the *turn-taking* rational of conversations describe how most conversations are organized and identify the allocation mechanisms that effectively distribute turns amongst the participants in a conversation. Secondly, we explained that spoken human interaction is organized in *sequences*; universal patterns that consist of a group of turns related to the same course of action. Lastly, we described *the adjacency pair*; a flexible sequence pattern that in its most basic form consists of two turns where the second turn is responsive to the first and where different speakers compose each turn. We found that these concepts aided our comprehension of the design material because it allowed us to recognize how actants create and sustain mutual understanding in communication.

Repair

Suchman (2007) emphasized the central activity of repair in human-computer interaction and as mentioned, repair is an essential concept in ordinary conversation as well. Communicative repair refers to the work the participants in a conversation perform in order to recover from troubles during an exchange, and consists of three components: identifying the trouble source, repair initiation, and repair solution (Richards & Schmidt, 1983). When designing CAs, communicative repair, and repair in human-computer interaction concerns the same thing, namely addressing the complications that arise from the fragility inherent in contingent interaction. Designing robust CAs is a venture that

entails predicting the unpredictable, as both machine and user should be able to adapt to breakdowns by identifying the need for repair and applying the appropriate strategies for action in coordination. Hence, we applied this concept in order to understand the issues that complicate communicative repair processes in child-CA interaction.

Graceful interaction

We borrow the concept of graceful interaction from Hayes and Reddy (1983) who argue that interaction between people and computers must be as humanlike as possible in order to be graceful. They base their argumentation on an understanding of *robust communication* as "the set of strategies needed to ensure that a listener receives a speaker's utterance, and interprets in correctly" (1983, p. 233). However, there are good and bad ways to repair communication, and incessantly asking the user to repeat and confirm pieces of information would probably create poor user experiences. They propose that in order to design for graceful interaction, the system has to use and react correctly to the techniques we commonly use when initiating and implementing repair work.

Robustness

In this thesis, we aim to construct an understanding of what robustness entails in design of CAs for children. Our point of departure in this endeavor is Dix et al.'s (2004, pp. 270–273) definition of robustness as pertaining to the level of support the systems provides the user in order for them to successfully achieve their goals. They present four main principles that robustness includes: (1) *observability* is the ability of the user to evaluate the internal state of the system from the observable representation in the interface, (2) *recoverability* refers to the ability of the user to take corrective action once an error has occurred and been recognized, (3) *responsiveness* concerns how the user perceives the speed of communication with the system, and (4) *task conformance* is the degree to which the system supports all the tasks the user wishes to perform.

Broken world thinking

Broken world thinking (Jackson, 2014) is a perspective on technological development that foregrounds repair and maintenance as activities that offer opportunities for innovation. We live in a dual world, simultaneously destructive and restorative. Thus, breakdowns are an inevitability of life, and of technology use. We will apply broken world thinking as a perspective because placing breakdowns and disintegration at the core

of our research might provide insights into what makes interactions between children and CAs fragile and enable us to understand how one can remedy or mediate this fragility. Hence, we will also contribute our thoughts on the potential merits and faults of utilizing this perspective in a slightly different way than suggested by its originator, namely as both the starting-point and as a perpetual focus during the course of a design process.

3 Research methodology

In this chapter, we will first present our philosophical perspective before we describe the case study as a research methodology and the design approach. Next, we describe the children's role in the process, before we introduce the techniques used for collecting and analyzing the empirical data. Finally, we present the methodological challenges we have encountered, and the ethical considerations we have considered before and during this research project.

3.1 Philosophical perspective

Myers (1997) states that we can divide qualitative research into three underlying philosophical assumptions; positivist, interpretive and critical research. Positivists generally assume that reality is objectively given and that there exists one truth independent from the researcher. Conversely, interpretive researchers believe that reality is socially constructed and they try to understand a phenomenon through the meanings people assign to them and focus on the complexity of human sense-making in the given context (Myers, 1997). Conversations provide the researcher with access to people's thoughts and enable discussion. The philosophical grounding for this research project is the interpretive paradigm. We believe that our findings are dependent upon us being the researchers, as well as our participants being who they are. We as researchers interpret knowledge that we have jointly constructed along with the participants.

3.2 Case study

At the beginning of this research project, we sought to gain an understanding of a phenomenon: children's interaction with conversational technology. We found it suitable to investigate this by conducting a case study. Stake (2005) identifies three types of case studies: intrinsic, instrumental and multiple (collective) case study. If one seeks to gain a better understanding of a particular case, because the case itself is of interest, the case is intrinsic. If one chooses a case to provide insight into a general issue, the case is instrumental. An instrumental case study can teach us something about a phenomenon, but the case itself is not of interest. A researcher can also conduct several instrumental case studies jointly, and these will constitute a multiple (collective) case study. The cases

are chosen because we believe that understanding them will lead to a better comprehension of a still larger collection of cases (Stake, 2005). The case we have chosen is in itself not what we seek to understand, but it is the phenomenon. Hence, we define our study as an interpretive *instrumental* case study. We believed the case could teach us about the general phenomenon of child-CA interaction, as we perceive the case to be of a general character where nothing particular about the cases would take away from the study of the phenomenon.

Stake (2005) claims that qualitative case studies call for examinations of complexities pertaining to the context of the case. In our case study, we found the need to understand the context in which we situated the technology. We wanted to learn how a CA could function as a pedagogical tool in a kindergarten setting. Through the methods we applied in the case study, we attempted to gain an understanding of pedagogical practices in the kindergarten in order to inspire the development of the CA.

3.3 User-centered design

After the case study, we conducted an empirical study founded on our findings from the case study by employing a user-centered design (UCD) process. We concluded that we needed to do a UCD process after we established conversational technology as the technology to be explored as well as preschool children as the target group. We made this decision because we were unable to find a CA for children based on spoken input and output in Norwegian to employ in the exploration. We were also motivated to explore dialogue as design material, and we concluded that we would be better equipped to explore child-CA interaction by creating a CA of our own because we would be able to customize it to the target group.

Figure 3 depicts a version of a UCD-process model (Pagliari, 2007, p. 6). The core value of UCD is to include the users throughout the process to ensure that the product in development realizes their needs (Lowdermilk, 2013). The process encompasses activities through which one seek to gain insight and understand the context, specify requirements, produce solutions and evaluate them. One conducts these activities in iterations and work towards a solution that meets the set requirements.

Our process was somewhat different from the model shown in Figure 3. In our research, the step where we work to understand and specify the context of use was conducted in the case study, during which time we conducted observations and interviews for this purpose. We also identified the need to conduct a design process involving the users prior to the case study. In the UCD process, we worked in four iterations where we produced design solutions and gained insights through several evaluations. Our goal was not to establish and specify requirements and make a system that satisfied said requirements. Rather, the prototype was an artifact that allowed us to explore the research question and gain more knowledge of how we could design robust speech interfaces for the intended target group.

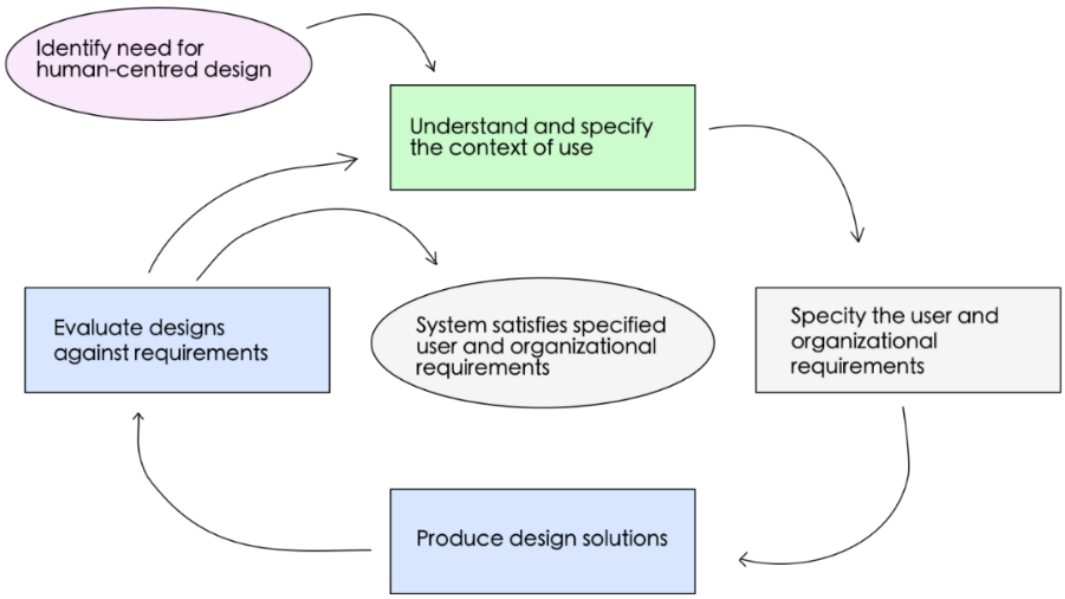


Figure 3 The ISO 13407 standard for human-centered design processes for interactive systems

3.4 The children's role in the study

Allison Druin (2002) establish and describe four distinct roles that children can inhabit during a design process; *user*, *tester, informant* and *design partner*. *Users* simply use released technology whilst being observed or tested for skills. The researcher seeks to understand the impact the technology has on the user so that future technologies can be changed for the better. Druin explains that in the role of a *tester*, the children test the prototype whilst the designers/researchers observe and ask for comments regarding their

experience. In the role of *informant,* the child is a part of the design process at various stages, such as the ideation phase, the prototyping or when the product is finished. During this time, they might be asked for input on sketches, or be observed with existing technology or with the finished product. In the role of *design partner,* the child is considered an equal stakeholder throughout the process and is as much an integral part of the team as the designer or researcher is (Druin, 2002). We aimed for the participants to inhabit the role of informants to enable them to influence the making of a technology where they are the intended target users. Through selected strategies, we sought to make the children competent actors by enabling and supporting participation and having them be mindful of the interconnected activities of a design process. In retrospect, we found that throughout the process the children alternated between the roles of user, tester and informant. We will elaborate on this in Chapter 7.

3.5 Methods of inquiry

We chose the methods of inquiry in the design process to obtain data that would allow us to gain knowledge, not only of how to design for robustness but also of child-CA interaction for research purposes. Our objective in choosing methods was to better understand important aspects of a specific target group, technology, and context. Thus, we were interested in learning about people's thoughts and viewpoints as well as learn from their expertise. Due to our philosophical lens and the fact that this is an interpretive study, our data collection methods have been qualitative because this approach permits a thorough and deep understanding of the subject at hand. The methods used in the case study were participant observation, interviews, and focus groups, and we conducted a workshop and several evaluations in the design process, as well as one interview. We have conducted all data collection in Norwegian. Consequently, we translated all quotes with the aim of making it as verbatim as possible.

In the coming sections, we describe the methods of inquiry and provide an explanation of how we have used them in our research. Table 1 gives an overview of our data gathering process, in which we conducted the first four activities as a part of the case study and the latter five as a part of the design process.

Source	*Method*	*Time of Execution*	*Participants*	*Section*
Children and educators	Participant observation	Week 38, 39 & 41	Children and educators	4.1
Educator	Interview	Week 39	1 educator	4.2
Researchers at OsloMet	Interview	Week 40	2 researchers	4.3
Children	Focus group	Week 41	7 groups, total of 17 children	4.5
Children	Workshop	Week 45	7 groups, total of 23 children	5.2.2
Children	User testing	Week 47	8 groups, total of 17 children	5.3.3
Children	User testing	Week 49	3 groups, total of 7 children	5.4.3
Children	User testing	Week 1 (2019)	4 groups, total of 9 children	5.5.3
Children and educators	User testing through indirect observation & interview	Week 5 (2019)	6 groups, total of 15 children 2 educators	5.6.2

Table 1 Overview of our data collection

3.5.1 Observation

Observations can provide insights into a context that could otherwise be lost. Participant's accounts do not always provide a full picture, as they may forget or simply be unaware of certain aspects. Thus, we can use observations to supplement interviews. During observations, the researcher can be either a passive or a participant observer. A passive observer does not take part in what is happening, while a participant observer joins in on the activities (Preece, Rogers, & Sharp, 2015). Observations can be either direct or indirect. The observation is direct when the researcher is present in the context and observe what is happening first hand. The observation is indirect if the observer is not present, and what is of interest can be caught on camera or tape and the researcher observes through this material, or the researcher can use logs or diaries to gain an understanding of what occurred (Preece et al., 2015). These are just some techniques for doing indirect observation.

In our research, we have done both direct and indirect observation. During the initial observations, as well as throughout all of our visits to the kindergartens, we have immersed ourselves in the context and participated in activities arranged by the educators and engaged in play with the children. Hence, the observations were participant and direct. In the last evaluation, we had the participant's make audio recordings of the sessions in which we were not present, which made the observation of this last evaluation indirect.

3.5.2 Interview

Walsham (1995) states that interviews are the primary data source within interpretive studies as it is through interviews that the researcher can access people's interpretations and viewpoints. Interviews can be structured, semi-structured, or unstructured (Preece et al., 2015). In structured interviews, the questions are predetermined; the interviewer asks them in a set order and follows a plan. Unstructured interviews have some pre-defined themes and the conversation unfolds naturally around these. Semi-structured interviews are a mix between the former two; the interviewer has prepared questions or themes but remains open to following new topics or themes if these arise (Preece et al., 2015). Walsham (1995) expresses the importance of balancing between being too passive and too directive when conducting interviews. If the interviewer directs the interview too closely, Walsham warns that the data obtained from the interview will lack the richness that interpretive studies seek. The researcher should also abstain from being too passive, meaning not following new directions introduced by the participant or not offering their own input to the conversation (Walsham, 1995).

The interviews conducted for this thesis have been semi-structured. The interviewees have been domain experts within pedagogy, and we chose them as primary informers due to our curiosity and need for insight within the field of preschool education. The questions asked were of an open character and invited the participant to share experiences and knowledge. We were also open to deviate from the original plan if other interesting topics arose. Prior to the last interview, we gave the participants an elicitation diary to record their thoughts and experiences during the activities in the field trial (see Appendix N). This diary was structured as a form and prompted the participants to provide brief answers concerning their experiences and thoughts regarding use. We used the rating

scale *semantic differential scale*, were one can express one's opinion by choosing a point on a scale, which often range from positive to negative. This as a way of having them quickly rate their perception of use. We used the answers the participants provided as a basis for the concluding interview after the field trial.

3.5.3 Focus group

Focus groups are often described as a group interview. The researcher invites several people within the target group (normally 3-10) to participate and the method could possibly provide a wide range of viewpoints and insights effectively (Lazar, Feng, & Hochheiser, 2010). There is an assumption that individuals develop opinions within a social context by talking with others, and as focus groups allow for discussion amongst the participants this method could get the participants to form opinions they would not have had or been aware of in an individual interview (Preece et al., 2015). The researcher will take on the role of facilitator rather than as interviewer. There will be a preset agenda but openness to follow unanticipated issues are important. It is the facilitator's job to guide the discussion. The more active participants must be moderated to avoid them becoming too dominant, and the quieter ones must be encouraged to partake (Preece et al., 2015).

To make sure the children felt comfortable in the activity of talking to Siri, we tried to set an informal tone for the sessions. For that reason, we made the decision to do focus groups instead of interviews, to include the children in groups, not one and one. We also wanted to facilitate discussion regarding the use of conversational agents and enable the children to play off each other's input.

3.5.4 Workshop

In workshops, the facilitators provide the participants with tasks, and the participants actively partake in their learning about a topic. According to the Merriam-Webster dictionary, a workshop is "a class or series of classes in which a small group of people learns the methods and skills used in doing something" ("Definition of workshop," 2019).

We used workshops at one instance during the project. In the workshop, we engaged the children in making physical representations of robots as a way of widening their

perspective regarding robots and their looks and functionality. This was also the first activity in the design process and a strategy for motivating the children to partake in the development of the CA. The objective of this workshop was to make the children open-minded as to what a robot could be. Moreover, we aspired for the children to be aware of the process they were a part of, and we considered this a fun and "easy" start onwards to the following activities.

3.5.5 User-based testing

Lazar et al. (2010) describe user-based testing as "a group of representative users attempting a set of representative tasks" (Lazar et al., 2010, p. 260). User-based testing can take place anywhere in the development process. Early inclusion of users is preferable, as it enables the designer to gain insight before major decisions have been made. Hence, changes are easier to implement. User-based testing is important as it can be difficult for the designer to envision the real end-users needs. The purpose of these tests is to improve the product by discovering flaws that would cause problems for the users and identify what works well (Lazar et al., 2010).

We conducted user-based testing at four instances to evaluate our prototype. The focus in these evaluations was to gain an understanding of the prototype's limitations, uncover challenges during use, as well as to identify precursors to breakdowns. The underlying goal was to increase our knowledge about child-CA interaction in order to investigate what robustness entails. The last evaluation also included the educators, as we also consider them users in this context. We wanted to understand their experience and thoughts about the use of this technology in a kindergarten setting.

3.6 Prototyping tool

In this section, we will present the software tool used for developing the CA prototype (application) for this project. We will elaborate on the tool's functionality, opportunities and limitations. We will also describe the hardware on which we ran the application: a Google Home device.

3.6.1 Software: Google Dialogflow

When prototyping the CA for this project, we used Googles *Dialogflow*. This is a developing tool for technology based on natural language conversations. The tool is available to anyone who wants to create conversational user interfaces for private or commercial use. The applications built in Dialogflow runs on Google Cloud Platform, which enables hundreds of millions of users to access the content when released.

Dialogflow is powered by Google's machine learning. They provide a natural language understanding engine to process and understand natural language input (Google, 2019c). This makes the system able to understand different ways of uttering a request, which again enables a more natural way of interaction. The Google Assistant was released in Norwegian on the 11th of September 2018. As the Natural Language Processing improves by using data provided by the users, this recent release means that its proficiency of the Norwegian language during our project was still limited but this will improve with time.

We will now explain how we construct an application in Dialogflow, focusing on the functionality we used when we developed our prototype. In Dialogflow, the developer must first create an *agent*, which is the application (the CA). The agent consists of intents, which is defined as a task or a goal of the user (e.g. ordering a flight ticket, getting information about the weather forecast, finding the nearest coffee place, playing a game, etc.). The intents are triggered by user utterances and the developer design the output and program the action which fulfills the user's intention (see Figure 4 (Google, 2019c)). Each intent has a set of training phrases that comprise of different ways the user might communicate what they want the system to do. The more training phrases you add, the more likely is it that the system manages to match the user's utterance to the correct intent. The training phrases can be added manually by predicting how the user would phrase their requests, or by applying phrases taken from the training data. The training data is a log of the interaction between the user and the agent. The developer can review how the system interpreted the input and assign this input to the correct intent if this did not happen during the interaction. The input will then be added as a training phrase for that intent, and the next time the user says the phrase, the correct intent will be triggered. If the system is not able to match the input with any intent, the fallback intent is triggered. This is a default intent in Dialogflow that the developer can customize. The standard output is, "I didn't understand you. Try saying it one more time!".

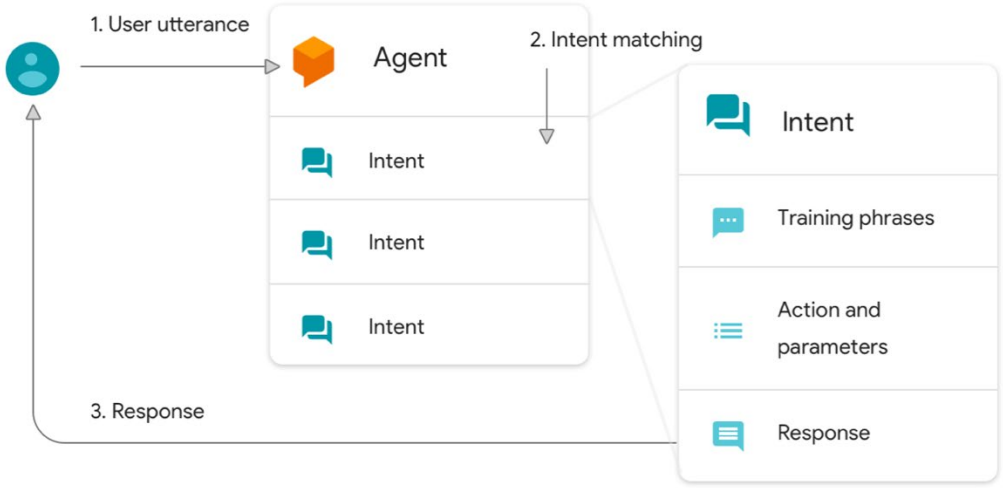


Figure 4 Visualization of the content of an agent

The dialogue we design is the output, implemented as text, which the agent presents to the user as a spoken utterance. If the agent is used on a device with a screen, the agent can present the output visually (the text on a screen). One can add an unlimited amount of responses, and if the agent is created without implementing code, the agent provides these responses randomly (if there is more than one response). The response can also be implemented using code, which allows more flexibility as to what the agent can do. The code is then deployed through a web service. Google provides Firebase, a mobile platform to contain code for this purpose. When the agent is in use, Dialogflow sends an HTTP POST request containing information input to Firebase. The web-hook returns the output based on the information it received. Figure 5 illustrates this (Google, 2019b).

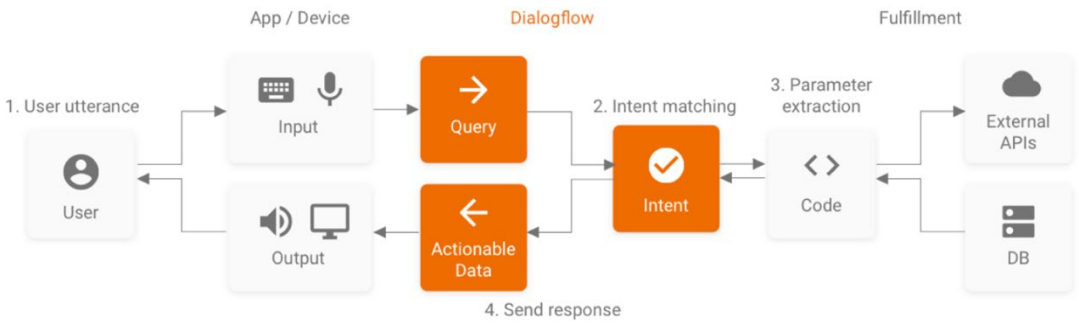


Figure 5 Visualization of the data flow using external APIs

To control the flow between intents, the developer must implement *contexts*. Contexts connect intents and restrain entry into intents. Hence, the purpose of contexts is to control the dialogue flow. This is done through activation and deactivation of contexts on entry

and exit from specific intents. To provide an example: the agent should not be able to enter an intent called “answer”, which provides the answer to a quiz question, without first having entered the intent “question”, which provides the question. This is being done by activating the context that allows the user to trigger the “answer” intent at entry of the intent “question”.

As previously mentioned, all interactions with the agent will be logged and available to the developer through their account on Dialogflow. The developer can train the agent based on this data or just observe the input and output from an interaction. This will give the developer insight into what the system interprets the input to be and compare this to what the user actually said, and thus provide a basis to understand the ability of the natural language processing.

To summarize, Dialogflow is a tool for developing CAs that is easy to use if you want to create a basic application. It is not necessary to implement code to create an agent that can receive input and provide simple text-to-speech output. As Dialogflow supports the implementation of code and actions, there are vast opportunities for creating custom-made outputs. However, as we comment on in this thesis, implementing an agent that is able to partake in conversations that are more open requires both extensive programing skills as well as comprehensive contents production.

3.6.2 Hardware: Google Home

To run the application, we used a Google Home Mini. Through the Google Home, the user can access applications from a vast range of third-party providers. Applications must be launched to be accessed by anyone, but self-made, un-launched applications can be accessed by the maker directly through the Google Home when connected to one’s user profile. To enter applications from third-party providers, the user says “OK Google” to activate the Google Assistant application, and then says the activation phrase for the desired application. To enable the system to detect the activation phrase “OK Google”, the microphone has to be on, but no sound is logged before the activation phrase opens the application.

As for the physical appearance of the Google Home Mini, the design is rather simple, as shown in picture 1 (“Google Home,” 2019). The artifact looks similar to a speaker, and we can adjust the sound by tapping on the sides. There are lights on the top that indicate the status of the system – if it is providing output or waiting for input. While providing an output, the lights pulsate and while waiting for input, the lights are static. There is one button on the device, which turns the microphone on or off.

Picture 1 Google Home Mini

3.7 Methods for data analysis

The data acquired through the abovementioned methods have been analyzed using techniques inspired by thematic analysis. Thematic analysis was introduced by Virginia Braun and Victoria Clarke (2006) as a distinct approach to conducting qualitative analysis The objective was to provide a basic and flexible approach applicable either in relation to theory or independent from theory (Johannessen, Rafoss, & Rasmussen, 2018, p. 279). When using thematic analysis, the data material is grouped into categories – in themes – to order the data material and identify relations.

Braun and Clarkes (2006) approach consists of six phases: (1) get to know the data, (2) generate initial codes, (3) searching for themes, (4) reviewing themes, (5) defining and naming themes, and (6) producing the report. The first phase consists of transcribing, reading and re-reading the data material and writing down initial ideas. In the second phase, the researcher codes the data systematically. A code is explained as a phrase that describes what is said as objectively and specifically as possible. In phase three, the interpretive analysis of the data occurs, and the codes are ordered into potential themes. In phase four, the themes are reviewed and refined; some might be removed due to lack of

codes, some might collapse into each other, or some might need to be divided in two. In phase five, the researcher has to identify the "essence" of each theme. The theme is well refined if they can describe the scope and content in a couple of sentences. Phase six involves writing up the themes (Braun & Clarke, 2006). Braun and Clarke (2006) emphasize the need to include good examples or excerpts which captures the essence of the point made. The researcher must use the data to make an argument related to the research question at hand.

After the focus groups, we had collected considerable data material in the form of transcripts. We were unsure of what this material could teach us, and we found it necessary to do a thorough analysis to gain an understanding of the material and the relations within it. Hence, we followed the abovementioned steps rather strictly. In the remainder of the research process, however, we did not follow the steps this rigidly but drew inspiration from thematic analysis. Our data material has mainly consisted of transcripts, which we have thoroughly familiarized ourselves with and discussed before establishing themes of interest. After the focus groups, the emphasis has been on breakdowns including the trouble source, precursor, and repair strategies that occurred during the interactions.

When transcribing the audio recordings of each activity, we highlighted and commented on several things we considered important in order to discern the reasons why the interaction either broke down or succeeded. These were the following: pronunciation, intonation, volume, phonetics, perceived feelings, mood, and interplay between children in interaction (finishing each other's sentences, interruptions, talking aloud, reacting on output, etc.). This was important in order to understand what actually happened during the interaction.

Use of concepts from conversation analysis

To understand the structure and logic of conversations as well as the difficulties that arise in child-CA interaction, we used the following descriptive concepts and principles from conversation analysis: turn-taking, sequences, adjacency pairs, communicative repair, and the cooperative principle. We described these in further detail in Chapter 2. These concepts allowed us to investigate how the complexities of natural conversations complicate conversations between humans and machines. Moreover, it has helped us identify how conversations with CAs are different from human-to-human conversations

and to understand the intricacies of human-computer communicative repair work. When the restrictions in human-CA interaction became apparent, the concepts from conversation analysis aided us in structuring our understanding and focusing our analysis of child-CA interaction.

A framework for the analysis of transcripts

To structure the analysis of transcriptions in the design process, we made a framework intended to accentuate trouble sources, precursors, and repair strategies. The term "trouble source" is described in Chapter 2 as a term referring to the point during talk where the need for repair occurs. In our analysis, we use it as a category to group specific precursors to breakdowns. Placing extracts of transcriptions into this framework made these two factors highly visible and enabled us to clearly see the precursors for each trouble source as well as what repair strategies the children applied when the communication broke down.

Speaker	Verbal Exchange	Trouble Source	Repair Strategy

Table 2 Framework for analysis

3.8 Methodological challenges

3.8.1 Gaining access

The first challenge we had to handle was to establish cooperation with kindergartens. Our aim was to work with participants from two kindergartens, as we found just one to be vulnerable, and three to be too extensive in consideration of the time-limitations we had to work within. We wanted to be fully invested in the kindergartens we cooperated with and have the time and opportunity to show them that we were both competent and professional. To gain access, we started by casting our net wide. We reached out through e-mail to about 10 managers of different kindergartens that we found suitable, meaning that they had departments for children ages 4-6 and was located in the Oslo area. Additionally, we used our personal networks and contacted kindergartens with which we had connections. Only two managers answered our request, both from kindergartens with whom we had a personal affiliation. They both asked for a meeting where we could

explain what we had planned to do. In this meeting, we were able to clarify what our intentions were and how we wished to include both the children and adults in the process. Both managers were optimistic, and they invited us to conduct our fieldwork in their kindergartens.

3.8.2 Recruitment of participants for each activity

After the initial observations, we were familiar with the children and tried to establish groups we thought would work well together. In one of the kindergartens, this did not work. When we convened the children for the activities, several children eagerly asked to join us, all from different groups. We considered it best to include the children who asked at the same time, as we did not want to make anyone feel bad about being excluded. In the other kindergarten, we managed to keep the groups consistent during the whole design process. This was possible because the children resided in several departments and we could gather the children from one department at a time. Thus, seven children became permanent participants for the workshops and evaluations in the design process. When recruiting the children, we were always careful not to interrupt or intrude on those who seemed busy, and we did not offer any incentives for them to join.

In both kindergartens, several children did not participate in the project, due to either being too young or because the parents did not sign the consent form. We did experience incidents were these children wanted to join the activities. We handled this by explaining that only the oldest children could join, or by clarifying that their parents had not signed off the participation form. The children found this somewhat disappointing but seemed to accept it rather quickly.

3.8.3 Managing expectations

It was important for us to try to set the expectations for our project at a reasonable level for both adults and children. We explained our objectives for the project to all stakeholders and described how we would include them during the process. We clearly stated that we would not provide a finished product but develop a rather simple prototype. As for the children, we started out by calling the conversational agent a “talking robot”, as we deemed this an adequately descriptive and a comprehensible term. We wanted to decrease the expectations of a robot as something that could move around. However, the

term “talking robot” quickly became just “robot”, and it became evident that the children had some preconceived notions about what a robot should be able to do, such as move its body and face and inhabit other human-like capabilities. Thus, we experienced that some children did not want to acknowledge the conversational agent as a robot (“it’s not a *real* robot”), as it was neither mobile nor able to answer all of their inquiries. This suggests that we failed to manage the children’s expectations for the final prototype properly.

3.8.4 Our role as researchers

During our fieldwork, we wanted to establish trust to create a safe environment for creativity and discussion. This led to some children wanting to tell us about their personal lives, and it seemed to be difficult for the children to distinguish between the adult employees and us as researchers. This is something we discussed with the managers in each kindergarten, and we reached the conclusion that if difficult situations arose, we should kindly refer the children to one of the employees, and in cases where children argue, or fights develop, we should notify someone and let the employees handle the situation. Although these are good guidelines to go by, it did sometimes prove difficult as situations develop very quickly, and the children looked to us for guidance or help to resolve these situations. We could not simply just walk away from children fighting and had to interfere in some instances. We tried to do this in a kind and mediating way to maintain a good relationship, but at the same time not appear tolerant of bad behavior.

3.9 Ethical considerations

The Norwegian Centre for Research Data has approved this research project. An application consisting of the research purpose, planned methods, interview guides, and informed consent forms (for researchers, educators, and the children’s legal guardians) was disclosed prior to the data collection process.

3.9.1 Informed consent

Complete voluntary participation was essential to uphold in this project. Children are not able to give “formal” consent of participation, and consequently, we needed their legal guardians to sign a written consent form. However, in addition to the adult’s written consent, it was important that the choice to participate in activities was entirely up to each

child and that the children perceived that opting out of an activity was without consequences. Furthermore, the children should know and feel that it was ok to withdraw from an activity. Under no circumstances did we let our need to finish an activity supersede the children's freedom of choice. All adults participating in the study was also provided with informed consent forms, which explained their role in the project, and their rights related to the use of the data material they provide.

3.9.2 Working with young children

Working with children this age introduced several ethical considerations that would not necessarily be applicable when working with other age groups. In relation to children, adults are ahead in their physical and cognitive development and it is crucial to be aware of this imbalance of power when conducting our research. Moreover, it was important to us that the children never felt pressured to participate. Under no circumstances did we force completion of an activity if we observed that the children were uncomfortable or reluctant. Some of the activities we conducted were of less interest to some of the children, which caused them to express boredom and ask if we could do something else. This posed the dilemma of how we would get the activity done but at the same time uphold the ethics of voluntary participation. In these situations, we had to diverge from the planned approach. Sometimes, we followed the children's suggestions for doing things a little differently. Other times, we reached an agreement with the children and played board games at the end of the session, played music and danced to give them a break, or simply gave them the option of discontinuing the activity. This way, we were able to balance everyone's needs and both children and adults had a good experience.

The degradation of air quality is a serious subject. During our process, we were conscious of how we presented the topic to avoid scaring the children with facts regarding negative effects on the human body. Our focus was on raising awareness about air as a substance and on how it important for all life on earth. We tried to draw parallels towards littering and described poor air quality as "dirty air", which we found to be a comprehensible metaphor.

3.9.3 Protection of privacy

In a kindergarten, private concerns are frequently discussed between parents and employees. We consistently used a separate room for the activities in order to avoid recording such conversations, as well as to ensure that we only included and recorded conversations with children whose parents had signed the consent form. We never used a recorder in the common areas. Moreover, we also considered the Google Home a recorder as it detects speech, interprets it, and stores it in the system. Thus, we refrained from placing the device in the common areas. Additionally, we had to adjust to the children talking about private affairs during the activities. What started as a conversation about technology or air quality could quickly take a turn towards the more personal and intimate, and many children were not able to filter appropriate personal information. When this happened during an activity we recorded, we refrained from transcribing that section of the conversation and deleted the recording after finishing the transcription.

Lastly, before we began the fieldwork, we signed a non-disclosure agreement with the kindergartens and procured a police certificate of conduct. All the names, both of people and kindergartens, are fictional in this thesis and we have refrained from providing descriptions that could lead to the identification of both people and places. An exception from this is the researchers at Oslo Metropolitan University. They did not want us to refer to them by name but approved the use of identifying information. We provide this information in the thesis as we found their area of work to be of importance related to why we chose them as informants.

4 Preliminary case study

In the early fall of 2018, we decided that the core of our research would be young children in interaction with conversational technology. We could not find existing conversational agents developed for this specific target group, and thus recognized the need to create and employ a prototype in order to explore communication between these actants.

We decided to conduct a preliminary study to refine the focus area for the design- and research process. This exploration took on the form of an instrumental case study, where the phenomenon we sought to investigate was children's behavior in interaction with conversational technology. As we had never before observed children using CAs, we decided to begin the project by attempting to gain a deeper understanding of what Löwgren and Stolterman (2004) call *the design situation*. This design situation simultaneously refers to the reason behind the design process and the context within which the design work is done. A designer's understanding of this situation encompasses the technology, the actants involved and the issues they experience. The understanding of the design situation is continuously evolving as we learn more about it, thus the problem space also changes (Löwgren & Stolterman, 2004, pp. 6–9). Our inexperience with the technology in relation to the target group demanded us to do groundwork in order to inform and develop our understanding of the challenges we were facing. Thus, we decided to begin to explore the phenomenon by observing the children in interactions with a *proxy* (Apples Siri). This way, we could identify challenges in child-CA interaction and establish a better foundation upon which to structure the subsequent design process. Additionally, we identified a need to study and learn about the specific context of use. Stake (2005) states that an examination of the complexities of the context in the case is central to understanding the case itself. We intended the children and educators to use the prototype we developed for an educational purpose. Hence, we also attempted to understand how a CA could function as a pedagogical tool in a kindergarten setting. We established two research questions for the preliminary case study, which guided our explorations during this part of the research process:

RQ 1: How do children behave as users of conversational technology?

RQ 2: In what ways can a CA function as a pedagogical tool in a kindergarten context?

Figure 6 depicts the structure of the case study. In this chapter, we explain how the insights from two interviews, participant observations and seven focus groups culminated in a research question that we addressed through a subsequent design process. Through the case study, we gained insights into how a CA could fit into the kindergarten context. Furthermore, we recognized a fragility in the interactions we observed between the CA (Apples Siri) and the participants, which led us to focus on robustness in design of CAs for children. We ended the preliminary case study by establishing the main research question for this thesis: *What does the concept robustness entail in design of conversational agents for preschool children*?

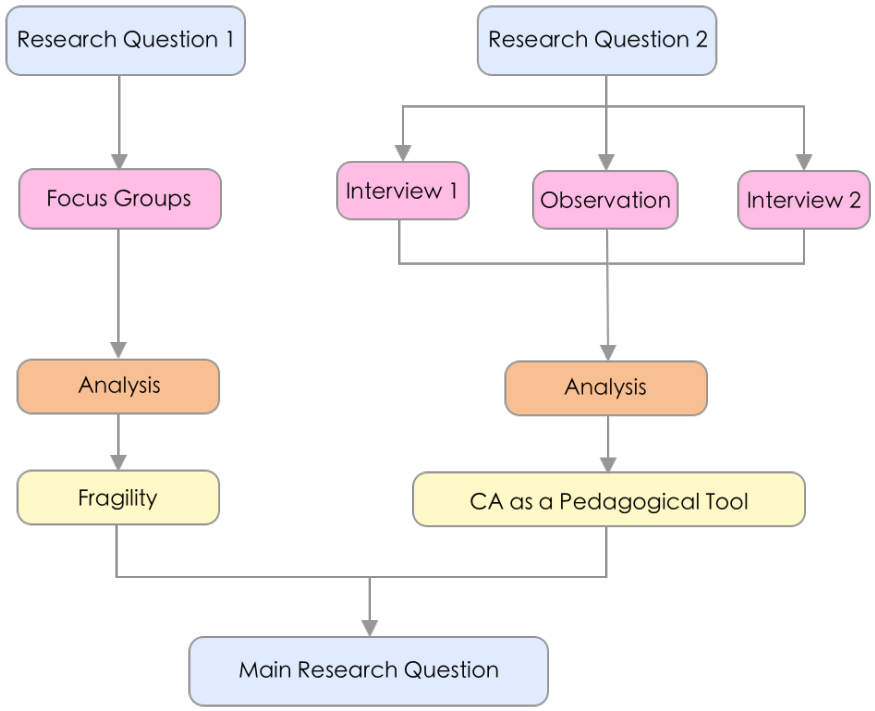


Figure 6 Visualization of the case study

During August and early September of 2018, we established collaborations with two kindergartens. We will call them Blueberry kindergarten and Raspberry kindergarten. Children age 4-6 and some of the educators partook in the case study and the design process.

4.1 Participant observation in a kindergarten context

We conducted participant observation over the course of six full days in the end of September and beginning of October of 2018. The objective was to learn more about how the CA could function as a pedagogical tool in this context. Additionally, we wanted to take the time to get to know the children and adults as they would participate and advise us during the design process. In this introductory stage, we took on the role of "kindergarten assistants" and participated in everyday activities alongside the children and adults.

During the participant observation, we observed children in many different kinds of unrestricted and organized play. We joined them in their explorations and followed their whims and impulses. We also participated in several pedagogic group sessions. At both kindergartens, these sessions had a personal or interpersonal focus, and the educators applied different techniques to teach the children about the current themes and issues. They often used books and fairytales to illustrate the point they were trying to make and asked the children if they had experiences that resembled the situations conveyed in the stories. Our observations of their use of artifacts, storytelling, and reflection regarding personal experiences and feelings inspired us when making the content of our prototype. We will elaborate on this in Section 4.4.

4.2 Interview with educator

During the period of participant observation, we conducted an interview with the pedagogic leader of the department for the oldest children at Blueberry kindergarten. We will call her "Guri". We spoke of how they approached the subject of environment and sustainability, as well as her thoughts on how a CA could be useful to both educators and children in a pedagogic context. We inquired into what use she might make of a CA in her practice. She told us that on field trips they considered it valuable to convey to the children that adults do not have all the answers, and rather than providing answers to all their questions, they deemed it more important to be engaged in the children's musings and reflections. She suggested that they could use the CA as a tool for asking questions before or after the field trip, as the children often asked intriguing questions when engaged in different activities and settings. Thus, the educator could partake in "the wondering" along with the children, whilst the CA could provide answers to more

specific questions afterward. Regarding techniques for engaging children in pedagogic conversations, Guri put emphasis on the importance of using "concretes" such as table theatre, books or other artifacts in order to contextualize the topics; "that way, it isn't just talking".

4.3 Interview with researchers at Oslo Metropolitan University

At the beginning of October 2018, we conducted an interview with an associate professor ("Tina") and a research fellow ("Nora") at the Faculty for Education and International Studies at Oslo Metropolitan University (OsloMet). Their research group is concerned with children and kindergartens in a sustainable future. Tina's work mainly revolves around how adults in kindergartens use technology in their daily practice. Nora had been involved in studies examining children's esthetic exploration of recycled materials, as well as of the role of sustainability and the environment in kindergartens in general. The objective for this interview was to learn about technology use in kindergartens today; what is the purpose of use, what is the educator's role and what should we consider when introducing a novel technology in this context. Additionally, we had questions concerning how educators can breach the subject of sustainability to young listeners. Lastly, we asked what we should think about when including children in a construction process, as we thought this could provide important insights concerning how to involve the children successfully in our design process.

Exploration supported by adults

We learned that when educating small children, providing information is considered secondary to facilitating exploration on the children's own terms. Children in kindergartens learn through active and explorative participation. Hence, they advised us to make use of concrete experiences and everyday surroundings to instigate learning and reflection. They continued by emphasizing that the educator's role is essential when conveying complex subjects such as sustainability. It is their job to put the small things and everyday activities into a larger context and thus bridge the gap between mundane, habitual activities and surroundings and the larger context of the environment and sustainability. Based on these arguments, they underlined the importance of including the educators as an active participant when using CAs. They warned us that if we failed to

either include the educators or make evident the positive contributions a CA might provide, it was at risk of becoming "just another thing, that's sort of fun, but seldom or randomly used by the children".

Technology practice in kindergartens

When asked about technology practice in kindergartens today, Tina answered that the main challenge is developing and sharing knowledge, as well as creating confidence regarding use and exploration of technology. Furthermore, there is a vast difference between technology-use in different kindergartens. Some kindergartens have bought iPads but have failed to develop an understanding of how it can function as a tool in their day-to-day practice. Others are in the forefront by utilizing tools for simple programming or applying other digital equipment to include the children in different construction processes. However, they are presently part of a marginal group. Tina and Nora continued by explaining that introducing new technology in an expedient fashion requires support by explaining and demonstrating use to the educators.

Considering the above, Tina stressed that there is an ongoing discourse of when, how and why one should utilize technology in kindergartens. The technological approach departs from traditional pedagogy, which has primarily concerned itself with how to develop interpersonal competence and human-to-human relations. In society today, technology is omnipresence and children are surrounded by technology in every aspect of their lives. The challenge within the pedagogic field is raising awareness amongst kindergarten employees regarding how that influences the children's childhood.

> *"For the children, a smartphone or a tablet is a natural thing, contrary to the reality of our childhood. This massive shift requires us to keep up with the children's state of mind and the group of people working in kindergartens are not, traditionally, the most technologically competent"*

To conclude, Nora told us that her experiences suggest that technology quickly becomes an individual exercise and is thus at risk of producing isolating practices. One example is children using laptops or tablets, which quickly turns into a one-person activity. Thus, trying to make the technology part of a group situation can instigate dialogue and accentuate the communal aspects of technology use; if designed well it can make the children both more independent and engaged as well as support collective learning processes by including both children and educators.

Children as associative beings

Lastly, Tina and Nora emphasized that unlike adults, children do not consistently stay within the context of a conversation. They are associative in nature. For instance, although the topic of conversation is air quality, some children may associate certain words that are spoken with something entirely different. Hence, the conversation can transgress into something else, and children might share quite personal stories from private environments. They were concerned with what happens to these stories if a system connected to the CA logs them. This is something we need to consider and that we have commented on in Chapter 3.

Children in construction processes

Lastly, they told us that when participating in activities, children often "move in and out of situations", and that people tend to interpret this as disinterest. However, it is not certain that this is the case. Children need to follow their impulses, play out their thoughts, or they just need space to process what is happening away from the situation. Often, they return after a while and should be allowed to do so. Thus, we should take this way of being into consideration when implementing planned activities during the process. We tried to be mindful of this in our work and have commented on how we dealt with similar issues in Chapter 7.

4.4 Key insights: CA as a pedagogical tool

During the observations and interviews, we were inspired regarding how the CA could function as a pedagogic tool. In this section, we briefly discuss how we envisioned a CA as one of many actants in a process of knowledge construction.

Facilitating reflection – from the familiar to the complex

Through our observations in the kindergartens, we found that in conversations with the children the educators seldom provided answers or information as a first resort. This was interesting in consideration of how to construct a pedagogic conversation between a CA and children. We observed that the educators often instigating a group session by introducing something concrete such as a story or an object and asked the children to simply repeat what they had seen or heard. They continued by associating the theme to the children's personal experiences. Only after this, did they ask the more difficult

questions that required reflection and thinking, or provide new information or explanations. Hence, they constantly shifted between the familiar and the complex by meticulously broadening the contextual limits of the conversation.

Exploration as the primary approach to learning

Regarding children and learning, we found that simply providing information was considered secondary to enabling exploration. We should not treat children as passive recipients of information but provide them with experiences with which to connect new knowledge. Moreover, our experiences from the participant observation in the kindergartens suggested that children's questions arise in the context of activities, meaning that hands-on-experiences trigger musings and reflections. The desire to learn does not appear out of nothingness. Guri's story of how the children often ask clever questions whilst on field trips illustrates this point.

Embedding the CA in a situation with educators as mediators

Lastly, we found that although the kindergarten was the overarching context of use for the CA, within this environment it could contribute in many different situations. We could envision it as an active provider of tasks and missions before or during field trips, as a companion or "concrete" during group sessions held by the educators, or a "talking encyclopedia" answering questions the children asked during activities. However, we found there to be a consensus amongst the experts and practitioners within the field, that the educators should mediate the use of the CA. They based this on the complexity of the subject matter we were to convey and underlined the importance of including the educator in contextualizing the experiences, questions, or information the children might have.

4.5 Focus groups: conversations with Siri

We concluded the case study by conducting a series of focus groups with children from both kindergartens in which they tried to interact with Apple's digital assistant *Siri*. We wanted to observe them in interaction with a technological *proxy* to gain a deeper understanding of how they would communicate with a CA and attempted to identify the challenges that can arise when young children become users of this technology. A proxy refers to "devices and applications that incorporate as much as possible similar functionalities and characteristics as the future media technology" (Pierson et al., 2006, p.

30). Choosing Siri as the proxy was a matter of accessibility. At that time, Siri was the only CA with a voice interface available in Norwegian. Consequently, we chose to use this as an exemplar of a conversational technology, although this digital personal assistant is not developed specifically for a target group comprising children. Siri is embedded in several of Apple's devices. You can ask Siri to perform specific tasks, like setting an alarm, make a phone call, search for information, but "she" is also able to engage in fun interactions and banter as well as provide some follow-up questions to its users.

Additionally, we found it interesting to observe what kind of conversations the children instigated with the CA. Lastly, the children were to become informants in the design process and we considered it valuable for the continuation of the process that the children gained some experience with a technology similar to what we were to develop.

We conducted seven focus groups with 2-3 children at a time. The focus groups took place in a room separate from the common areas in both kindergartens and lasted between five to twenty minutes. During the focus group, we gave the children free reigns to explore the technology. If they found this difficult, we guided the children in their interaction by demonstrating what we could ask Siri to do. We gave every child in the group the opportunity to interact with Siri and asked questions about whom or what they thought Siri was afterward in order to instigate a conversation about their experiences of the technology.

Only two of the children participating in the focus groups said they had previous experience with Siri. The other 15 children told us it was their first time using a CA, and we observed mixed reactions to their first "encounter". Some children were very eager to start interacting. They talked loudly, in long sentences, and nearly pulled the phone out of our hands. Others did not want to say anything and seemed apprehensive and skeptical of engaging in interaction with Siri.

4.6 Key insights: children as users of CAs

Our main insights from the focus groups concerns factors that instigated communicative breakdowns in child-CA interaction. The different breakdowns that unfolded during the conversations between Siri and the children provided significant insights upon which we derived much of the subsequent research and design. In the following section, we want to

describe some of these incidents. Preceding this, we must provide clarification of what we define as a *communicative breakdown* in this context. We find that communication has failed to achieve its purpose if there is no shared understanding of the context matter between the actants participating in the dialogue. Based on this we defined a communicative breakdown as situations in which the actants fail to achieve a mutual understanding of one another's intents and responses. During the focus groups, we identified three interactional consequences that stemmed from communicative breakdowns: (1) Siri ended the interaction due to inability to match intent with function, (2) Siri provided a response that was unrelated or incorrect, or (3) Siri provided a response that was not intelligible to the users.

The insights gained from the focus groups underlined the fragile nature of communication between children and CAs, and we proceed by describing this fragility by detailing several identified precursors to breakdowns. We have categorized the precursors into three main trouble sources: children's behavior, unrealistic expectations of capabilities, and technological inexperience.

4.6.1 Trouble sources in child-CA interaction

Children's behavior

The trouble source we identified as *children's behavior* relates to how certain types of conduct influenced the interaction with Siri. Throughout the focus groups, there were several instances of communicative breakdowns, even when Siri was not part of the exchange. The following excerpt illustrates this point:

> ***Interviewer:*** *Who do you think made Siri?*
> ***Sindre:*** *I did not make Siri.*
> ***Interviewer:*** *No. But do you think that somebody made her?*
> ***Sindre:*** *I don't think so.*
> ***Andrea:*** *I really don't think so because I've seen a different church!*
> ***Interviewer:*** *What did you say, Andrea? What did you mean?*
> ***Andrea:*** *I've seen a different church. One where I was not baptized.*
> ***Interviewer:*** *... oh.*
> ***Sindre:*** *It is Daidalos [Childrens TV-inventor] who has*
> ***Andrea:*** *It has a blue house on the roof!*

This conversation exemplifies what we recognize as the *associative nature* of children and illustrates how it confounds cooperative meaning-making in communication. During our interview at OsloMet, the researchers told us that children often think associatively. In the focus groups, we experienced that this could lead to sudden changes of topic, difficulties in keeping the conversation on-theme, and abrupt backtracking to previous conversational contexts. Considering Grice's (1982) principle of *relation*, which concerns staying relevant to the purpose of the conversation and being context-aware, this incident represents how children might relate topics differently than adults and how this causes confusion. If they make a connection to something familiar, they might bring this up because it is relevant to them, but for the listener, it is out of bounds regarding the context of the conversation. Hence, it can be a source of a breakdown. Moreover, clarifying this misunderstanding would probably require extensive communicative repair work in the form of elicitation and explanations. These insights are in line with Ninio and Snows (1996) comments on how keeping to one subject and developing this topic throughout a conversation is a difficult skill that it takes time for children to acquire. Thus, conversations with children often require extensive adult support in order to progress in a meaningful way.

Furthermore, we often observed that the children applied a make-it-up-while-you-go approach, which created trouble because Siri was unable to discern their intents. The children did not always seem to have a specific purpose when they addressed Siri and consequently, their inquiries or statements were inefficient, disorganized, and often contained phrasal breaks, restarts, and hesitations. This inefficiency is in breach of Grice's (1982) maxim of *manner,* which primarily concerns requirements to be specific and brief in one's communication. It might also reflect the fact that introducing topics and themes in a conversation is a conversational skill that it takes time to master (Ninio & Snow, 1996).

Unsurprisingly, we observed that the children were inclined towards playful and silly behavior. In many cases, this resulted in non sequitur conversations in which every participant talked about different things. The following passage illustrates such incoherent dialogue as Siri interprets the child's statement to mean something else entirely:

Interviewer: *Is there anything else you would like to ask Siri, or are you done?*
Louise: *I'm not done! Why do we draw, what happens if we draw on a sheet? Why is that possible?*
- Siri does not respond-
Louise: *What name do you think if …*
Siri: *What affiliation are you looking for?*
Louise: *Excuse me, I'm not done talking! I said what happens if a snowman melts and becomes an ice-warmer?*
Siri: *I'm sorry to hear that. If you feel threatened, you must seek help!*

These instances often occurred because the children asked nonsensical questions like the above, or "do you have a pen that is a unicorn like a very long word?" We perceived these kinds of utterances as a playful approach to the technology; we do not see it as she sought an answer to this question, but rather that she wanted to elicit some sort of response. This finding is comparable to Druga et al. (2017) observations pertaining to how children often tested the limits of the technology in a playful manner.

Unrealistic expectations of capabilities

It eventually became evident to us that the children expected Siri to be capable of engaging in complex, give-and-take conversations similar to human-to-human communication. One of the most frequently used approaches when initiating a conversation with Siri was for the children to try to get to know her. They inquired about personal information and offered facts about themselves in return, which is similar to what Druga et al. (2017) found in their study, where children probed into the identity of the agents. The inquiries we observed the children make eventually caused breakdowns because Siri was unable to handle these types of reciprocal interactions adequately.

One of the main objectives of the focus group was to understand what kind of conversations the children instigated with Siri. We found that they seemed to be socially motivated when initiating contact and providing topics and activities around which to center the conversations. We frequently observed the participants telling Siri stories, telling Siri about the surroundings, clarifying perceived misunderstandings, offering information about themselves and so on. This might indicate that children this age are more interested in engaging in social interaction rather than communicating with CAs for pragmatic purposes. To illustrate one instance in which this led to breakdown we provide

an example where a child asked for information about Siri and offered information about herself in return:

> ***Louise**: Can you ask if…. Can you ask what happens with (incomprehensible from audio) where you are, and which world do you live in and bla-bla-bla?*
> ***Siri**: As it says on the box, I was designed by Apple in California*
> ***Louise**: Do you live in California? We live in Norway.*
> ***Siri**: I found this online about "the belief yes we live in Norway".*

Lastly, we want to note that we observed several initiations for repair or elaboration. In some instances, the children prompted Siri to elaborate on her answers when they found it intriguing. In other situations, they asked her to explain when they did not understand her responses. In all these cases, the conversations broke down because Siri is not proficient enough to engage in conversations of this complexity or to perform intricate repair work. As described in Chapter 2, Shedroff and Noessel (Luger & Sellen, 2016; 2012) suggested that framing a system in an anthropomorphic way will raise the user expectations, which is what we may have seen an example of in this activity. When asking about who or what Siri was, several observations indicated that many of the children thought Siri was a human, thus suggesting that they perceived the technology as anthropomorphic.

Technological inexperience

We found that some precursors of breakdowns were connected to the children's inexperience with the technology. As previously mentioned, most of the participants in the focus groups had never interacted with voice-based technology before. Hence, it proved difficult for them to understand that Siri was not always able to detect what they said at any given time. They often failed to recognize that Siri had to be "activated" before they could talk to her, which led to Siri only detecting fractions of their responses or inquiries. Furthermore, Luger and Sellen (2016) present findings that suggests that users with limited technological experience have trouble adjusting their mental models in accordance with the systems actual capabilities. We believe that this was evident during the focus groups, as the children did not seem to lower their expectations of neither intelligence nor capability despite continuous breakdowns.

4.6.2 Summary – key insights

We present a summary of the trouble sources and precursors in Table 3.

Trouble Source	Precursor
Children's behavior	• Associative: unpredictable behavior and random acts. Violate the contextual limits of the conversation. • Playful: play and silliness confound communication. • Inefficient speech patterns: fractured speak, stops, hesitations, restarts.
Unrealistic expectations of capabilities	• Expectations of humanlike capacity in interactions. However, CA is incapable of engaging in complex exchanges
Technological inexperience	• Children not aware of when the technology can detect their spoken input

Table 3 Trouble sources with precursors established in the case study

4.7 Establishing a research question – towards a concept of robustness

Hayes and Reddy (1983) defined *robust communication* as the interlocutors' abilities to respond and recover from unexpected situations. In accordance with this definition, we would describe the interaction between the children and Siri as far from robust. The technology was not equipped to handle the children's behavior, nor adequately manage their expectations regarding capabilities and sociality respectively. Conversely, the children were unequipped to handle or understand the requirements for effective communication posed by the technology with which they were to interact. They did not attempt to uncover the capabilities of their conversational partner, and neither did they adjust their behavior in accordance with the failures they experienced. Hence, the mutual inability of the children and the agent to adapt to each other precipitated continuous breakdowns in interaction.

The findings from the case study made evident the profound fragility of communication and cooperative meaning-making in interactions involving children and conversational technologies. One can argue that it is impossible to expect a CA to be able to account for all eventualities and uncertainties of a situation. Consequently, breakdowns are inevitable as the technology is limited in ways we cannot yet address. However, we argue that we can strive for knowledge on what types of breakdowns are predictable and perhaps preventable. More importantly, we contend that understanding which breakdowns the

technology is unable to prevent or mediate will inform us on how to design so that the interaction falls apart gracefully, thus enabling positive experiences even in situations where the CA fails to understand the meaning or intents in the user's utterances. One of our main concerns in the succeeding design process was addressing the inherent vulnerability in communication between children and CAs, and we applied *broken world thinking* as a perspective for this purpose. The main research question for this thesis was established:

> *What does the concept robustness entail in design of conversational agents for preschool children?*

5 The design process

Following the formation of the research questions for our project, we conducted an empirical study integrating a user-centered design (UCD) process as an approach to developing a functional prototype. During this process, we worked in iterations with dialogue as our design material, aiming to understand how we could prevent and remedy conversational breakdown between the users and the CA. We evaluated and developed the prototype founded on an analysis of transcripts. In addition to informing our design decisions during prototyping, the transcripts from each activity served as research material, which enabled us to generate new knowledge about the interplay between the users and the technology. Hence, the prototype served as a tool that allowed us to explore our research question concerning the conceptualization of robustness in relation to conversational technology with young children as target users. In this chapter, we describe how our focus on addressing identified precursors to breakdowns led to a prototype with a strict structural design that provided little opportunity for improvisation outside the scripted flow of the conversation, thus failing to account for the main characteristics we observed in the user group. We discuss this further in Chapter 6.

An overview of the design process is provided in Figure 7. The design process lasted three months, during which time we carried out five activities with the children. We organized the activities within three distinct phases: phase 1: “engaging the children”, phase 2: “designing dialogue for a CA”, and phase 3: “field trial”, each with their own main objective. The iterations all began with a prototyping activity, as this was the basis for our exploration, and continued with an evaluation and a concluding analysis that in turn informed the next iteration. We conducted the first four evaluations several times with different participants in each kindergarten, and the fifth evaluation was a field trial in which we left the prototype in one kindergarten for them to use over the course of a week.

In this chapter, we describe the objectives of each of the overarching phases as well as the activities and outcomes of the related iterations in the design process. Our goal concerning the prototype itself was to make it good enough so that the children and educators could use it without requiring our presence in the field trial. In each iteration, we emphasize how breakdowns have been the origins of several realizations and design decisions along the way. The insight we gained during the process enabled us to discuss

our main research question on how to understand the concept of robustness in relation to this technology, user group, and context.

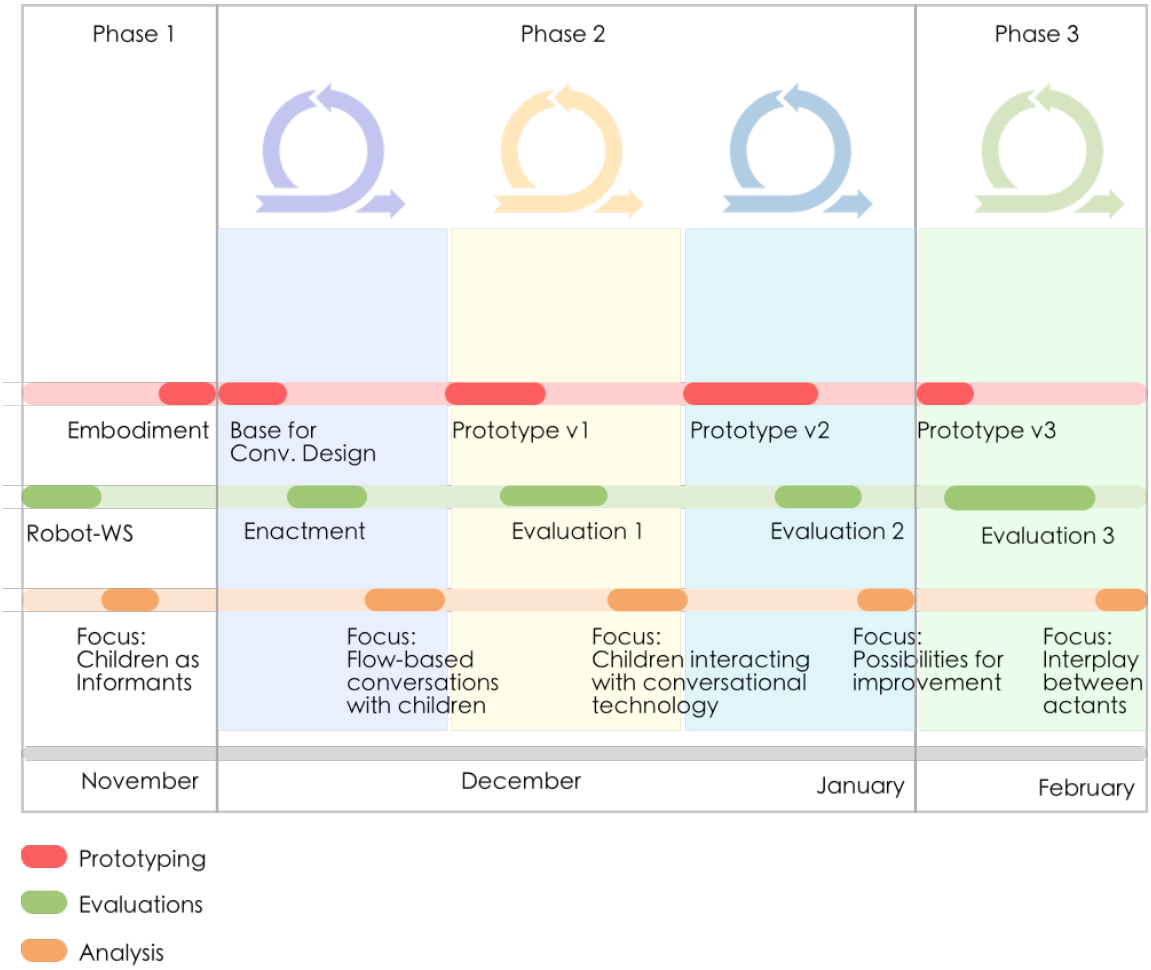


Figure 7 Visualization of the design process

5.1 CA as a pedagogical tool – general guidelines

When developing a pedagogic CA, we would have to keep the rationale of education in mind. We established three guidelines for the prototype that we founded on previous insights pertaining to how a CA could work as a pedagogical tool. We sought to explore how a CA could function as an initiator for contemplative conversations, and thus support the educator in their pedagogic objectives. The guidelines were as follows:

1. **The educators should be included in the use of the CA**

   The educators are essential in children's knowledge construction processes. The intricacies of explaining and contextualizing a complex subject such as sustainability were not feasible to manifest in a CA alone.

2. **The CA should be situated in a group situation, such as a pedagogic assembly**

   Avoid solitary use and encourage collective reflection, discussions, and learning.

3. **Make it "concrete"/embodied**

   Based on the findings from observation and the interviews, we found it important to be both specific and tangible when working with children. The educators always brought a thing related to the topic at hand, as a way of grounding the conversation and helping the children focus. Moreover, in the focus groups, we observed that many of the participants were apprehensive to interact with Siri and it was hard for the children to answer when we asked whom or what Siri was. We proposed that one way of mediating this uncertainty was to give the CA a physical form (other than the phone).

Phase 1: engaging the children

As explained in Chapter 3, including and supporting the children as competent actants in the process was an important motivating factor for us as interaction designers. We wanted the children to be an integral part of the process, and we constantly worked to elicit and understand their reactions and thoughts on the prototype we developed. Thus, the main objective in the first phase of the process was to enroll and engage the children that would subsequently act as our informants. Figure 8 shows the activities in phase 1.

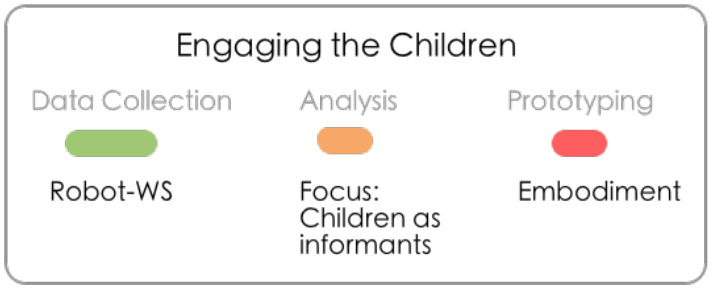


Figure 8 Activities in phase 1

5.2 Strategies for engagement

We applied three strategies to achieve the abovementioned purpose, "the Inventor's Book", a "robot-workshop", and a "concrete" that acted as a physical representation of the prototype in the duration of the process.

5.2.1 "The Inventor's Book" – connecting the design activities

"The Inventor's Book" was a tangible tool for visualizing and explaining the design process. We wanted the children to understand how the separate activities in the process was fractions of one overarching and continuous design endeavor. This was important in order to empower them as informants in the process. In this regard, we wanted to make an artifact that could incite comprehension of the process and highlight how the feedback the children gave during one activity influenced the prototype we developed. We intended it as a means for looking back at what we had done, as well as give the participants an idea of what was to come. Additionally, we hoped that by underlining how their contributions

informed the design, we could also motivate their participation in the succeeding activities.

Description of the artifact

The Inventor's Book had a front cover (Appendix A) with an illustration of the children and us as a team. The back cover (Appendix B and Figure 9) showed a pathway with four signposts that represented the activities we planned to carry out. The fifth post portrayed a rainbow symbolizing the very last activity – the field trial. In addition, we made a cutout of a drawing of three children, which we used to illustrate progress as we moved it along the path in accordance with the progression in the design process (Figure 9).

Figure 9 The Inventor's Book - back cover with figures

Introducing analogies - "inventor" and "talking-robot"

We introduced the term "inventor" as an analogy for "designer". We assumed that this term would be easier to understand, as many of the children had been talking about "Daidalos", a popular inventor on Norwegian children's television. There are also several other inventors in the cartoon world that they might be familiar with. When first introducing the Inventor's Book, we explained what it means to be an inventor and said that we as a group were going to invent a "talking-robot" similar to Siri. We explained the importance of documenting our thoughts and ideas as the reason for using the Inventor's Book. We briefly introduced each signpost on the roadmap and explained that they represented different activities related to creating our robot. We also explained that when

we reached the rainbow at the end, the talking robot would be “finished”, and they could test it along with the adults in the kindergarten.

Description of its use

In the first workshop, we gave each group one book and wrote their names on the cover. We added a new chapter to the book in each workshop. In the chapters, we would collect and document feedback from the children, and visualize each child’s contribution through pictures or writing. We wanted to emphasize each child’s inputs by writing their name next to their feedback. At the beginning of each new activity, we held a short “retrospective” in which we read the chapter we made the last time, and repeated and discussed everyone’s contributions in relation to the changes in the prototype. The techniques we used to elicit feedback from the children were not always successful. We describe difficulties related to obtaining comprehensible feedback in Chapter 7.

Did the Inventor’s Book work as we intended?

During the first introduction in the “robot-workshop”, the Inventor’s Book did seem to engage and motivate the children. The participants were attentive and showed great interest in the book at the initial introduction, and in later activities. Several children picked it up and started talking about it. They repeated which signposts we were presently at and discussed what we were going to do in the remaining activities. They repeated that when we reached the rainbow we would be finished. Some of the children were already looking forward to the next activity and expressed concern that they would miss it for some reason. We observed similar behavior in all proceeding activities. Consequently, we contend that the book to some degree succeeded in creating a sense of understanding and continuation in the design process.

Figure 10 The children looking at The Inventor's Book

5.2.2 "The robot-workshop" – what is a robot?

The second strategy we applied as a way of engaging the children in the process was "the robot-workshop". The objective of this workshop was twofold. Firstly, it was a starting point for the design process. The aim was to prepare the children of their role as informants, as well as engage and motivate future participation. We also wanted to widen their perspective regarding what a robot *is* and what they can do, as a way of managing expectations of what we were going to make. Secondly, as explained in Chapter 4, our findings indicated that a physical representation of the CA could be beneficial. We assumed that a physical form could make the remaining workshops easier as the participants would have something tangible to anchor them to the process. Thus, the main objective of the workshop was making robots and having fun. As the embodiment of the prototype does not address our research objective directly, we will refrain from describing details from this activity and the products the children made in these workshops.

Execution and analysis

We conducted this workshop over a period of two days and spent one day in each kindergarten. In Raspberry kindergarten, we conducted three workshops with seven children participating; in Blueberry kindergarten, we conducted four workshops with sixteen children. The workshops lasted approximately one hour each. We started by introducing the process and the overarching goal using the Inventor's Book as a supporting prop. The workshop consisted of two activities in which we wanted to challenge and expand the children's notion of what a robot can be or look like. We made 2-dimensional robots from a "robot-kit" we had prepared, that consisted of several laminated shapes and robots. We also made 3-dimensional robots from different arts and crafts materials.

Figure 11 Making of 2- and 3-dimentional robots

To prepare for how we should plan and execute the succeeding activities with the children, we founded our reflection on three main questions; what went well, what was challenging and what considerations did we have to deliberate going forward. We will refrain from describing how we implemented our strategies in the following phases and iterations of the design process. However, we will discuss our reflections and experiences concerning this topic further in Chapter 7.

5.2.3 "The concrete" – making an embodied representation of the CA

Our third strategy in phase 1 was making a physical representation of the CA. This so that the children would have something enduring to engage with during the evaluation of the dialogue design. We incorporated features from the robots the children created into this representation, as we wanted to accentuate the value of the children's input. We hoped this could increase their enthusiasm for the process and contribute to making their participation a positive experience. We wanted the look of the robot to reflect its area of expertise: air.

Figure 12 The final embodied representation with Google Home Mini

Phase 2: designing dialogue for a CA

Our design-objective in phase 2 was to develop a prototype that the children and educators could use without requiring our presence as facilitators of the interaction. During this phase, we developed and evaluated the dialogue design through which we tried to address the fragility in cooperative meaning-making between the children and the CA. The overarching research objective was to expand our understanding of robustness related to the interplay between children and CA. Figure 13 accentuates phase 2 and the iterations it consists of in the design process. We proceed by describing these three iterations, including design-choices, prototyping, evaluations, and findings.

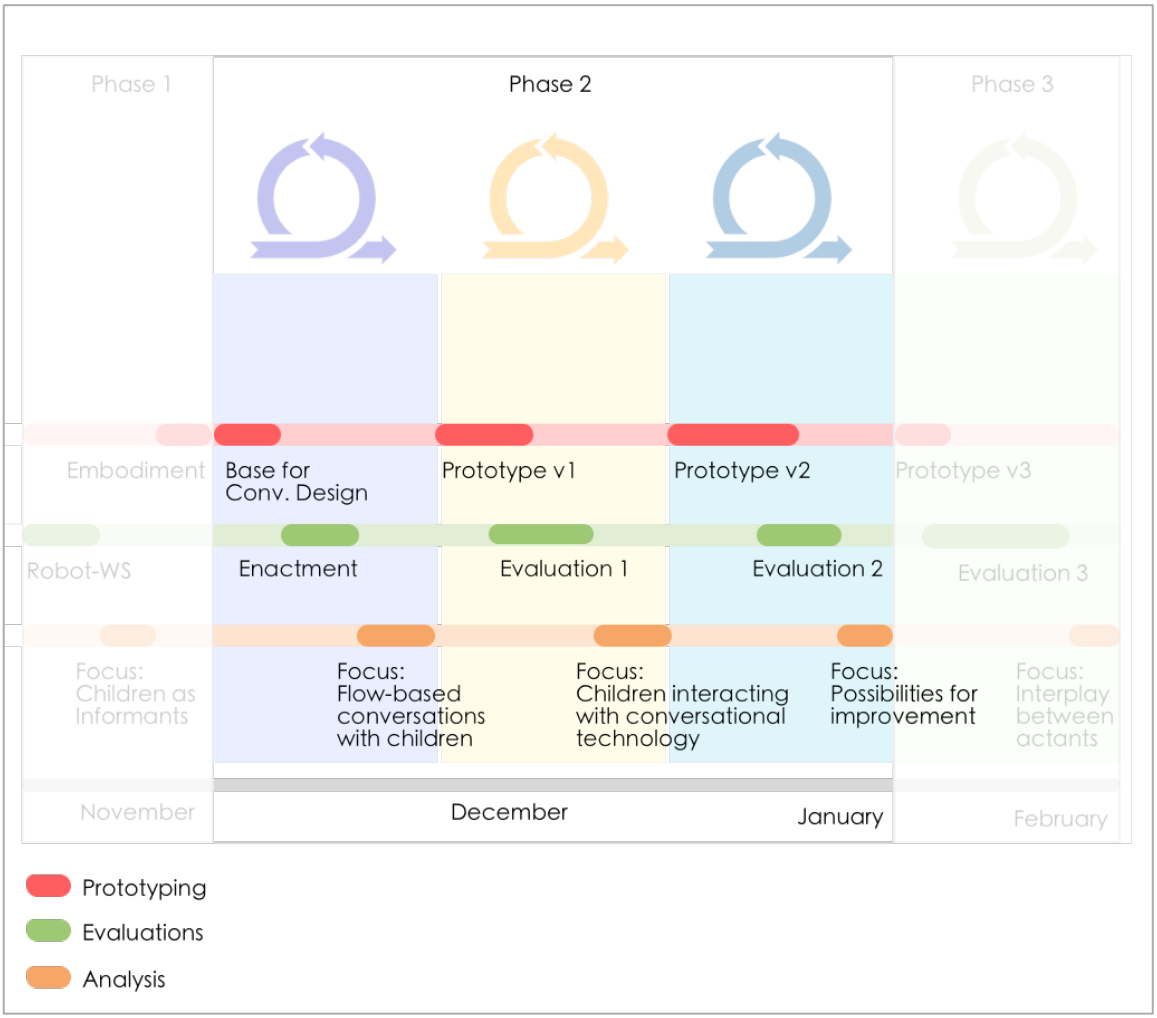


Figure 13 Visualization of phase 2

5.3 Iteration 1: enacting examples for dialogue design

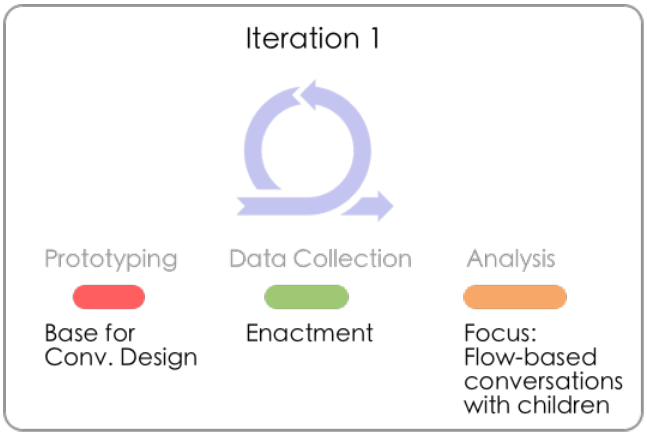


Figure 14 The activities in iteration 1

As a starting point for implementing the prototype, we began by conducting a role-playing workshop where we played the part of the CA. We intended this activity to generate representations of conversations (transcripts) that would serve as research material to inform our understanding pertaining to the research question, but also inform the development of the next iteration of dialogue design.

Enacting refers to methods where the participants try out prototypes or act out scenarios in settings that resemble the places where these activities would commonly unfold (Bratteteig, Bødker, Dittrich, Mogensen, & Simonsen, 2013, p. 164). Conversations are transient and will always be completely context dependent. Thus, role-playing could be a method for envisioning dialogue that resembled how children and CAs might interact in real life. Exploring both open and scripted inquiries with the children was an objective of this workshop, as well as identifying other potential breakdowns. In addition, we sought to observe the children's questions, replies, and phrasings so we could develop an adequate range of utterances to which the prototype would be able to react.

5.3.1 Low-fidelity prototyping of dialogue-flows

To make the interactions between the "CA" and the participants more similar to how such conversations would naturally progress, we began by prototyping simple dialog-flows (Appendix C) for a few "use-cases" (Table 4) which we wanted to enact along with the children. We intended the scripted dialogue to limit the range of possible responses to the

children's inquiries during the enactments as well as control the inputs and outputs of the conversation to some extent.

5.3.2 Three activities for dialogue design: mission, story, and quiz

Founded on the experiences, observations, and conversations we had during the case study, we established activities around which to design the dialogue. We determined three approaches we wanted to implement:

1. *Mission:* we decided that the CA should provide the children with short missions as a way of gaining concrete experiences upon which adults could base the succeeding reflective, pedagogic conversations.

2. *Story:* we based this approach on observations of how the adults used stories, books, and fairytales as a source for discussion and reflection on different topics of interest.

3. *Quiz:* the quiz originated from observing that the adults most often asked questions of the children before they explained or informed them about something.

We will not discuss the learning potential of these particular approaches further in this thesis, as it is not relevant to our research objectives.

5.3.3 Execution of the role-playing activity

Over the course of two days, we conducted three workshops at Raspberry kindergarten and six at Blueberry kindergarten. 19 children between the ages of 3-5 participated in total. 2-3 children attended each session. We recorded audio in each session, and the observer took notes of the participant's physical reactions and behavior. We executed the activity as follows: one researcher hid behind a curtain and pretend to be the CA whilst the other facilitated the interaction between the CA and the participants, observed, took notes, and asked questions when needed. The observer gave the children tasks (Table 4) that they had to solve by talking to the "robot". We had previously observed that it was difficult to find conversational topics when not provided with a point of departure. Consequently, we established the tasks as a means to make the interaction easier for the

children. We used a task-based form for note taking during the observation and intended to ask questions immediately after each task in order to assess the children's understanding of the content, questions, and words that occurred during the interactions. However, we discontinued this after a few tries, as it turned out to be difficult for the children to provide answers. In addition, the continuous questions disrupted the role-playing activity.

The purpose of the tasks we provided the children was to observe how they would try to elicit information about the capabilities and properties of the CA, as well as observe how they would initiate contact with the CA. For each task, we specified what to look for in advance, so we had a clear focus during the observation and could acquire valuable data upon which to build the first prototype. These sessions were also the first evaluations of the three activities; mission, story, and quiz, and would inform the way we structured the dialogue in the following prototype.

Task	What to look for
Find out what the robot can do	- Information load: do the children remember what the CA said? - How do they phrase their inquiry?
Ask the robot for a mission	- Do they understand the directions given by the CA? - How much do they lean on the adult/observer? - How do they respond when the robot elicits reflections on the activity? - How do they react to a long message? - Do they remember the information given by the CA?
Get the robot to tell you a story	- How do they react to the story? - Can they repeat any of the information given?
Get the robot to play a quiz	- How do they react when they answer incorrectly? - Do they talk about the question/answer? - Do they want more facts or more questions?

Table 4 Tasks for the enactment

Figure 15 The children's first meeting with the embodied representation of the CA

5.3.4 Insights from the role-playing activity

Recurring precursors to breakdowns

Our analysis reaffirmed many of our earlier conjectures regarding the many challenges of designing a CA for children. We noted several examples of associative thinking and ineffective communication as observed during the case study. We also observed several instances of playful behavior such as the children talking to the CA in a made-up language or making funny noises.

The problem with turn-taking and multiple responders

The main insight from this workshop was the identification of a new trouble source for breakdowns: the group situation. This poses a challenge related to the distribution of turns between the CA and the group of children. The practice of turn-taking is very complex, and as previously mentioned, the distribution of turns in a conversation is conducted either by the speaker overtly handing the next turn to someone, by someone self-selecting or by the speaker continuing to talk. In a dialogue involving users and a CA, the progression of the conversation necessitates the CA occupying every other turn. This does not emulate how conversations in groups are naturally organized. When the CA is finished talking and directs its question to the group of children, knowing or deciding

who should answer was a challenge which proved to demand some form of communication as the CA is unable to address anyone specifically.

During the role-playing, the lack of cues to indicate who was to speak next resulted in the children talking at the same time or repeating each other's answers in quick succession. Consequently, even simple questions such as "how old are you?" would naturally result in all the children answering because the inquiry was directed at the group and not one child in particular. Conversely, in some instances, none of the children would answer because no one was sure of whom the CA addressed. In a real setting, this would probably cause the conversation to fail, as a CA is usually unable to manage answers from multiple responders.

Debating responses

We identified an additional precursor related to the group situation, namely the occurrence of discussions as a way of determining a response. When the CA gave the children alternatives to choose from, there were instances where the children gave conflicting responses as they disagreed on which options to choose. Moreover, asking questions that required contemplation delayed the children's responses, during which time the children often talked amongst themselves, provided comments, or discussed what they should answer. We also observed many of the children thinking aloud, something that the CA might mistake for a response. The technology is currently unable to handle this input and would most likely misinterpreting these discussions as responses.

Silence: I'm thinking

Closely related to the need for debating answers is the need for contemplation when deciding upon a response. In interactions with a CA, an active microphone means that it is susceptible to any utterance or noise – increasing the possibility of breakdowns. The children often needed time to think about what the CA said, thus creating periods of silence whilst the microphone was active. If other participants spoke during this time, there is a possibility that the CA could detect this and interpret it as a response.

Reaction to output

We identified a new precursor pertaining to the reaction to output, which we relate to the trouble source "behavior". We observed that the children frequently interrupted the "CA"

because they reacted to something it said or answered a question before it was finished talking. In these situations, the person playing the CA would stop and listen to what the child said, and either answer or continue with the planned script. A real CA would be unable to do this. Consequently, it would only detect fractions of what the child said, possibly failing to identify the user's intentions and failing to respond correctly.

Participants did not perform general capability checks

One of the tasks we gave the children was to try to find out what the CA could do. They approached this task by being very specific. Instead of merely asking the CA "what can you do", they asked if it could do specific things, like "can you jump?", "can you breathe?", etc. Consequently, the CA was never given the opportunity to explain its capabilities and purpose.

Extensive need for adult intervention

Our experiences during the activity, as well as the analysis of the transcripts, revealed the extent of adult intervention and support the participants needed during their interaction with the "CA". In several instances, the observer had to interfere because the children did not respond to questions, responded incorrectly, or became distracted during the conversation.

Evaluating our first attempt at dialogue design: inefficient and cumbersome

We made some important findings pertaining to our attempt at dialogue design. Our analysis made clear that the initial dialogue flows were inefficient and added unnecessary complexity to the conversation. We identified two poor design-choices that caused this inefficiency: (1) we had failed to include a welcome message that conveyed the specific capabilities of the CA efficiently and (2) the dialog-flow inflicted the user with several redundant questions before reaching the actual pedagogic content.

Guiding the user by limiting possible inputs

We identified some positive aspects of the design. Questions with yes-or-no alternatives induced the correct responses from the children making the conversation progress as planned. Similarly, providing a limited set of predefined options seemed to work well. The children repeated the trigger words the "CA" used, they seldom made their sentences

too long, and mainly kept to the context of the conversation. Hence, we asserted that providing clear alternatives increased the predictability of input.

5.3.5 Summary – key insights from iteration 1

We include additional findings from this iteration in Table 5, which provides an overview of trouble sources and the precursors of breakdowns that we had identified this far. A complete list of trouble sources found in the design process can be found in Appendix L.

Trouble Source	Precursor
Children's behavior	• Interrupting: Responds to what the CA says before it is done talking
Unrealistic expectations of capabilities	No new additions
Technological inexperience	No new additions
The group setting	• Multiple responders: simultaneous talk, succeeding responses or conflicting answers. • No response: uncertainty of who should talk results in no one talking. • Talking amongst themselves. Debating answers means the CA can detect talk not intended as responses.

Table 5 Trouble sources with precursors established in iteration 1

The following findings were not direct precursors to breakdown, but were nonetheless important because they informed subsequent design decisions:

- The children did not perform general capability checks but asked specific questions such as "Can you do … "
- The need for adult facilitation was extensive. The children often looked to the observed if they needed help or was uncertain of something during the interaction.
- Inefficient communication increased susceptibility to breakdowns. The low-fidelity prototype of the scripted dialogue flow was inefficient regarding communicating possibilities for action as well as in progressing the conversation.

5.4 Iteration 2: first interactions between children and the CA

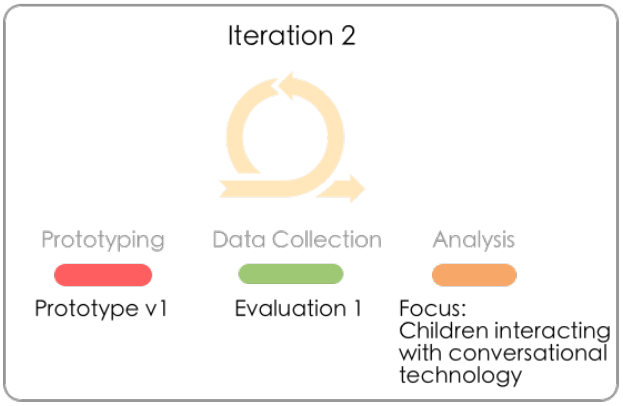


Figure 16 The activities in iteration 2

In the second iteration, we used the insights gained from the previous role-playing activity and created the first prototype in Dialogflow. Thus, we could observe how a conversation between the children and a customized CA would actually unfold. We found it important to assess how the technology handled children's voices and behavior in conversation. The evaluation was conducted with participants from Raspberry kindergarten. We intended to examine three main questions:

1. *How well would the CA understand the children*?
   We had identified the group situation as a trouble source. We were curious about whether this would be an actual issue, or if the technology would manage to interpret the children's responses regardless. We also wanted to see whether pronunciation would be a problem, as we expected that it would.

2. *How would the children behave in interaction with a CA designed for them?*
   We wanted to see if the children followed the scripted dialogue-flow, or to what degree they would diverge from the restrictions in the design. Moreover, we would examine how the children phrased their responses, e.g. would they answer with complete sentences, or repeat the trigger words offered by the CA. Lastly, we would examine the repair strategies applied by the children during the interactions.

3. *How efficient were our design decisions as steps towards a more robust prototype?*
   We wanted to assess the efficiency of the implemented design strategies for addressing the precursors for breakdowns we had previously identified.

5.4.1 Content and dialogue-flows in the first prototype

The first prototype was rather simple. We neither wrote code nor made use of custom-made contexts. We built the prototype using the intents in Dialogflow. To limit complexity, we separated the three main activities (quiz, story, and mission) into three agents. We found that the activities were well received by the children during the role-playing activity in iteration 1 and we decided to continue developing the dialogue based on this content. The dialogue was based on the pattern of basic question-answer type adjacency pairs (Schegloff, 2007), and at this time it was not possible for the user to expand the conversation e.g. by asking for repetition or paraphrasing, due to the necessity to implement code to enable this functionality. In the following sections, we present an example of a dialog-flow for each of the three activities. Flowcharts of structure and content can be found in Appendix D, E and F.

The quiz

The structure of the quiz was as follows: the CA asked the user if they wanted to play a game. If they answered "yes", they received the first question. The CA confirmed the user's response and prompted another question. If the user declined, the CA closed the conversation with a farewell message. If the user confirmed, it proceeded to ask a new question. The CA could ask up to three questions before closing the interaction. The following is an example of the scripted dialogue:

> ***CA**: «Hi, my name is Luftolino? Do you want to play a game?*
> ***User**: Yes!*
> ***CA**: Great! I'm thinking of a number between one and ten. Guess which one!*
> ***User**: four?*
> ***CA**: Wow! That's right. Do you want another question?*
> ***User**: No*
> ***CA**: That's ok! We can play at another time. Bye.*

The story

In the story-activity, the children contributed by making decisions during the course of the narration. The story had four possible outcomes. The decision of including the children during the storytelling was made based on observations from the *enactment*, where almost all participants had trouble concentrating during the narration. Consequently, we divided the story into shorter utterances and offered two choices of action for the character at the end of each utterance (see Appendix E for flow and content). The two options functioned as trigger words to enter the next section of the story. We provide an example:

> ***CA**: Hi, I'm Luftolino. I make stories together with children. Do you want to make a story?*
> ***User**: Yes!*
> ***CA**: The story is about Josephine. She is walking through a small forest. In front of her, the path splits into two. One path goes further into the woods, and the other goes into the city. Should she go to the woods or to the city?*
> ***User**: To the woods!*
> ***CA**: Josephine walks further into the woods. […] Should Josephine help Ola to get rid of the garbage?*
> ***User**: Yes*
> *Interaction continues*

The mission

The mission-activity entailed breathing through a straw for five seconds as a way of experiencing what it feels like to be deprived of air, thus acting as a starting point for discussing why air is important. We divided the mission into four instructional sections: (1) get the straw, (2) put the straw in your mouth, (3) breathe through the straw for five seconds and (4) discuss how it felt to breathe through a straw. When the children had finished one instruction, they had to let the CA know that they were ready to proceed. An example:

> ***CA**: Hi, I'm Luftolino. I give missions to children. Do you want a mission?*
> ***User**: Yes!*
> ***CA**: Great. First, you have to find a straw. Let me know when you're done!*
> ***User**: I'm done.*
> ***CA**: Super! Now, put the straw in your mouth. Tell me when you're done.*
> ***User**: We're done!*
> ***CA**: Now, I want you to breathe through the straw for five seconds. Let me know when you are done.*

***User**: We're done.*
***CA**: Great job! I want you to talk to each other and describe how it felt to breathe through a straw. See you later!*

5.4.2 Design-decisions in the first version of the prototype (v1)

Based on our findings, we made four design decisions before we began creating dialog-flows for each of the tree activities (mission, story, and quiz): convey capabilities efficiently, provide structure, decrease the number of turns, and support understanding of strategies for repair.

Efficiently conveying capabilities

Our findings from the *enactment* indicated the importance of offering the user an option quickly and explicitly as the children did not themselves initiate capability checks. In all three activities, the interaction began with the prototype greeting the user, identifying itself, and immediately prompting a yes/no answer regarding whether the user would like to proceed with the main activity (see excerpt below and full dialogue in the next section). The self-identification creates an expectation of capability, and the question lets the user take control of the conversation as well as handing the preceding turn to the user (Google, 2019a). This is also in line with *introduction* – one of Halls (2018) four key moments in which the CA should attempt to establish a relationship with the users by introducing its identity and make clear the potential value of continued interaction.

CA: *Hi, I'm Luftolino! I can give you a quiz/mission/story Do you want to get a quiz/ a mission/a story?*
User: *Yes/No.*

Structure and decrease the number of turns

As a strategy to decrease the probability of breakdowns, we decided to remove redundant utterances and provide clear alternatives for interaction at the end of each turn. We would not ask the children questions that did not contribute to the progression of the conversations. The content in Table 6 illustrate an example of this.

Previous prototype	Prototype v1 (current)
CA: I can tell you about air, I can give you a quiz about air and I can give you a mission. Does any of this sound fun? User: Yes CA: Cool! Let me know when you want to do any of the activities. Bye!	CA: I can give you a quiz/story/mission. Do you want a quiz/story/mission? User: quiz/story/mission CA: I will give you a quiz/story/mission. (provides quiz/story/mission)

Table 6 Structure of the opening phrase in previous prototype and the current prototype

Provide strategies for repair

In addition to efficiency and structure, we made one additional attempt at addressing breakdowns by way of dialogue design. The CA triggers a fallback-intent when it is unable to discern meaning from a user's response. In these situations, we designed the dialogue to let the user know that the CA did not understand by prompting a repeat of the response that necessitated repair. Additionally, we wanted to provide the user with guidance to identify appropriate strategies for repair. As described in Chapter 2, Hall (2018) describes guidance as contingency messages that can help the users back on track after a breakdown. The fallback intent would randomly provide the user with a prompt for repetition as well as offer general guidance on strategies the children could try to repair the interaction. The following are two examples of dialogue designed for the fallback intents:

> **CA**: *"I think you're being silly. Maybe you have to talk a bit louder for me to understand you?"*
> **CA**: *"I think you're joking around. I didn't quite understand you. Remember to talk one at a time"*

Summary of design decisions

Table 7 provide a summary of the design decisions we made before developing the first prototype.

Design decisions	Originated from
Immediately convey specific abilities in welcome message	The children did not perform general capability checks
More structure: Offer and repeat the options for interaction using trigger-words	The need for adult facilitation is extensive Increase predictability of input
Decrease the number of turns in the flow by removing redundant utterances and questions	Inefficient communication increase susceptibility to breakdowns
Provide repair strategies	Mediate breakdown by providing strategies for communicative repair

Table 7 Design decisions made in iteration 2

5.4.3 Execution: evaluating the first prototype

We conducted the first evaluation of the prototype at the beginning of December at Raspberry kindergarten. Seven children participated (groups of 3, 2 and 2) and each group session lasted between 15 to 25 minutes. We made audio recordings of each session and allowed time between to make notes and discuss our observations. We conducted the evaluation as a user test where the children were to try all three activities. First, we provided a short explanation of the prototype and its ability to provide activities. We decided in advance to refrain from teaching the children how to use the CA explicitly, as we wanted to see how they interacted with it without too much tutoring. We proceeded to let them try the story and mission. Due to technical difficulties, we were unable to evaluate the quiz.

5.4.4 Insights: evaluation of the first prototype

The evaluation of this prototype confirmed several assumptions: mispronunciation was a problem and the group situation was a challenge because it resulted in erroneous input. We were able to identify frequently used strategies for repair as well as observe the children's reactions to a fragile technology that repeatedly broke down. We also observed the children having difficulties assessing system status especially regarding when the system was able to receive input. In this section, we elaborate on these findings and exemplify some of them with extracts from our transcripts. These examples are presented

in the analytical framework we describe in Chapter 3, which is intended to highlight the trouble sources and the repair strategies that occurred during the interactions.

Trouble with interpretation of the children's responses

Concerning the CAs ability to understand the children, we identified three main issues: mispronunciation, flawed Natural Language Processing (NLP), and limited training. As mentioned, we expected mispronunciation to be an issue but did not want to establish this as a trouble source before evaluating an actual prototype. We found that the system struggled to interpret the words correctly when the children mispronounced the letter "r" and "s". We noted that the NLP was limited as it also misinterpreted input the children pronounced perfectly. Some of these issues might have been related to inadequate training with users from the target group. As the NLP, which is the systems foundation, is in continuous development, it is difficult to predict what the system will understand.

Multiple responders

As mentioned earlier, we expected that the system would have trouble interpreting input from multiple responders. We confirmed this as the CA was usually unable to handle simultaneous talk, succeeding responses, and contrary answers. Although, we also noted some instances in which the prototype interpreted this type of input correctly. The extracts shown in Table 8, 9 and 10 illustrate this challenge.

Poorly timed responses

When observing the children in interaction with Siri, we noticed that the participants had trouble interpreting system status leading to premature or overdue responses that the system was not able to detect. During this evaluation, we observed similar problems. After the CA has finished its turn, it takes approximately one second to activate the microphone so it can receive input. The Google Home device provides feedback about system status in the form of white lights transitioning from pulsating to static. However, as the children were not aware of this, they responded immediately after the CA had finished talking. This caused several breakdowns due to the system not detecting the input or only detecting partial input, resulting in failure to discern intent. The extracts presented in Table 9 and 10 show instances of premature responses.

Strategies for repair

We observed that the children attempted to repair by applying several strategies; repetition, increasing the volume of voice, talking more slowly and punctuating words in sentences. We provide an example in Table 8. The excerpt begins at a point where the participants, after some initial troubles, began creating the story. Eventually, they encountered a breakdown induced by contrasting responses. "Philip" became visibly frustrated at this point, taking deep breaths before shouting "to the woods". Shouting is a counter effective strategy, often leading the CA to trigger the fallback intent. However, during the next part of the repair-sequence, the CA manages to interpret the input correctly, despite both children yelling different answers. This illustrates that it is hard to predict when the NLP will succeed or fail – which makes it difficult for the children to understand whether shouting is an efficient or inefficient strategy.

Speaker	Verbal Exchange	Trouble Source: Precursor	Repair Strategy
CA	The story is about Josephine. She enters a small forest. […] Which path should she follow?		
Sophie	To the town		
Philip	To the woods	Conflicting responses	
CA	Now I think you're joking around. I did not understand you. Try talking one at a time		Guidance
Philip	"TO THE WOODS!"	Unintelligible response	Increase voice volume
CA	I think you're messing with me. Maybe you have to talk a bit louder for me to understand you?		Guidance
Philip	TO! THE! WOODS!		Increase volume Decrease tempo Punctuating words (augmentation)
Sophie	TO! THE! TOWN!	Conflicting response	Increase volume Decrease tempo Punctuating words (augmentation)

CA	Josephine walks into the woods […]. Should Josephine help Ola clean up or should she walk away?		
Sophie	Walk away		
Philip	Walk away	Succeeding responses	
CA	Josephine walks away because she does not want to help the bunny, Ola. […] This is the end of our story about Josephine.		

Table 8 Extract 1

Another extract, shown below in Table 9, exemplifies how the system had difficulties interpreting repetitive answers. We also observed how the children answered too quickly. In this case, the system was unable to discern meaning from the partial response. This extract also exemplifies how the children repeating things they heard the researcher say as they started using the word "Google". The fact that Philip says "yes" many times over may indicate frustration or be a strategy for augmenting his intent.

Speaker	Verbal Exchange	Trouble Source: Precursor	Repair Strategy
Researcher	Hi, Google. Talk to our robot		
CA	Hi. I make stories together with children. Do you want to make a story with me?		
Sophie	Yes, Google		
Philip	Yes, yes, Google	Succeeding response	
CA	You joke a lot! Or maybe I didn't understand you? Try saying it one more time.		Initiates repeat
Philip	YES, GOOGLE!	Premature response	Implements Repeat Increase volume of voice

CA	You're silly. I didn't understand you. Maybe you have to say it one more time.		Initiates repeat
Researcher	Yes		Adult Intervention
Philip	Yes yes yes yes yes yes yes yes yes yes yeeees	Succeeding response Repetitive answer	Multiple Repeats (augmentation?)
CA	Sorry. I cannot help you.		Close interaction

Table 9 Extract 2

Poorly timed advice

The evaluation made evident the failure of the guidance offered via the fallback intent as we observed incidents where the strategies for repair offered by the CA was unfitting to the situation. In one instance, the CA advised a participant to talk louder after triggering the fallback intent. However, the previous response was spoken in an extremely loud volume leading to the participant screaming the following response. This led us to think that these kinds of random advice would fail to serve its purpose and could even be detrimental to communicative repair.

Adult intervention to ensure progression

Despite not wanting to interfere extensively during the interaction between the participants and the CA, there were instances in which we inserted ourselves in the exchange in order to ensure the progressions of the conversation. We usually did this when the children had already attempted to make themselves understood repeatedly. We took over the interaction when the following situations unfolded: the participant's responses were premature or the participants shouted, and we predicted probable failure. The excerpt below shows one such event.

Speaker	Verbal Exchange	Trouble Source: Precursor	Repair Strategy
CA	The story is about Josephine. She goes through a small forest.[…] Which path should she follow?		

Philip	TO THE CITY!		Loud volume
CA	You joke a lot. Maybe I did not underst.. The story is about Josephine. She enters a small forest. […] Which path should she follow?	Technical failure	
Philip	THE CITY!	Premature response	Loud volume
Researcher	The City		Adult intervention
CA	Josephine starts to go towards the city […] Should Josephine ask Ola what contaminated means or should she just go home?		
Sophie	Ask		
Philip	Ask	Succeeding response	
CA	You joke a lot. Maybe I didn't understand you. Try saying it one more time		Initiates repeat
Philip	*Addresses researcher while CA is talking*: Can you do it?		Asks for help
Researcher	Ask		Adult intervention
CA	That the air is contaminated means […] This was the end of our story about Josephine.		

Table 10 Extract 3

Clear choices increased the predictability of input

We observed that the children mainly stayed within the restrictions of the dialog-flow. Hence, we assumed that presenting clear interactional options increased the predictability in interaction. However, although the dialogue made obvious the choices within the flow, it did not always offer explicit trigger words, which sometimes resulted in erroneous input and unpredictable responses, e.g. "sånn" (a Norwegian signal for readiness/completion) when we expected "ready", and "can we do that" when we expected "yes".

Low tolerance for failure and repetition

The children struggled immensely in the interactions with the CA, and the constant need to conduct repair resulted in frequent repetition of output. This clearly frustrated the participants, as they tried to remedy the communication, usually unsuccessfully. We

observed them asking us for support, questioning what went wrong, rolling their eyes, sighing, as well as outbursts of shouting at the CA. These observations are in contrast with what Cheng et al. (2018) found in their study, where the children rarely showed frustration when failure occurred. In our observations, the children quickly became discouraged and exhausted and asked us to discontinue the interaction. We comment on how we handled this frustration in Chapter 7.

5.4.5 Summary – key insights from iterations 2

We present new findings regarding precursors and trouble sources in Table 11. A complete list of trouble sources identified in the design process can be found in Appendix L.

Trouble Source	Precursor
Children's behavior	No new additions
Unrealistic expectations of capabilities	No new additions
Technological inexperience	No new additions
The group setting	No new additions
Linguistic skills	• Mispronunciation
Technological weakness	• Flawed Natural Language Processing • Limited training

Table 11 Trouble sources with precursors established in iteration 2

The evaluation provided the following key insights:

- We confirmed that the group situation caused difficulties regarding the distribution of turns between CA and children. Consequently, the system was frequently unable to interpret the user's intent in instances of succeeding responses, overlapping responses or contrasting responses.

- There were extensive issues regarding the interpretation of system status, which led to poorly timed responses that the CA was either unable to detect or only able to partially detect.

- Providing random advice via the fallback intent lead to misplaced and misleading advice and were in some instances detrimental to communicative repair.

- The children usually tried to repair the conversation by repeating their utterances, or augmenting their response by increasing the volume of their voice, punctuating the words, or slowing down the tempo.
- The frequent occurrence of repetitive repair sequences strained the interaction and frustrated the users.
- The children mainly stayed within the restrictions of the dialog-flow. Hence, we assumed that structure with clear interactional options increased predictability in interaction.
- The troubles experienced through interaction sometimes necessitated adult intervention to avoid breakdown.

5.5 Iteration 3: identifying remaining opportunities for correction

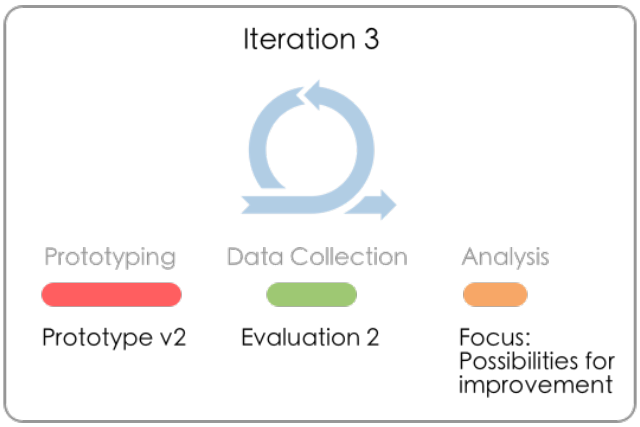


Figure 17 The activities in iteration 3

Preceding the concluding evaluation, we wanted to complete an assessment of the CA by having the children talk to it and test the design and functionalities we had implemented this far. We would evaluate the success of our design- and prototyping endeavors by identifying remaining weaknesses and opportunities for corrections. Our main objective was to improve the prototype as much as possible in advance of the field evaluation. We also wanted to simulate a more realistic use context by evaluating the prototype with larger groups of children.

5.5.1 Content and dialogue-flows in the second prototype

In order to develop a more sophisticated prototype, we had to implement code. Since we were novices as JavaScript-programmers, we spent about a month creating this version of the prototype (code for the final prototype is to be found in Appendix K). Implementing code gave us more flexibility regarding the structuring of content as well as more control over the complexity. Additionally, we could create intents that were triggered anywhere in the flow as a way of repairing the conversation (repeat, paraphrase), as well as apply Speech Synthesis Markup Language (SSML). This is a markup language that allows for a more humanlike output. We used it to add pauses, slowing down the tempo of output, making it more pleasant to listen to and easier to understand. We also added sound effects to capture the user's attention, give more life to the narration in the story-activity, and add a playful feel. We restructured the flow by adding all three activities into one agent. Figure 18 show a flowchart of the main intents as well as the users and the CA's actions in each respectively. Flowchart "story" can be found in Appendix I.

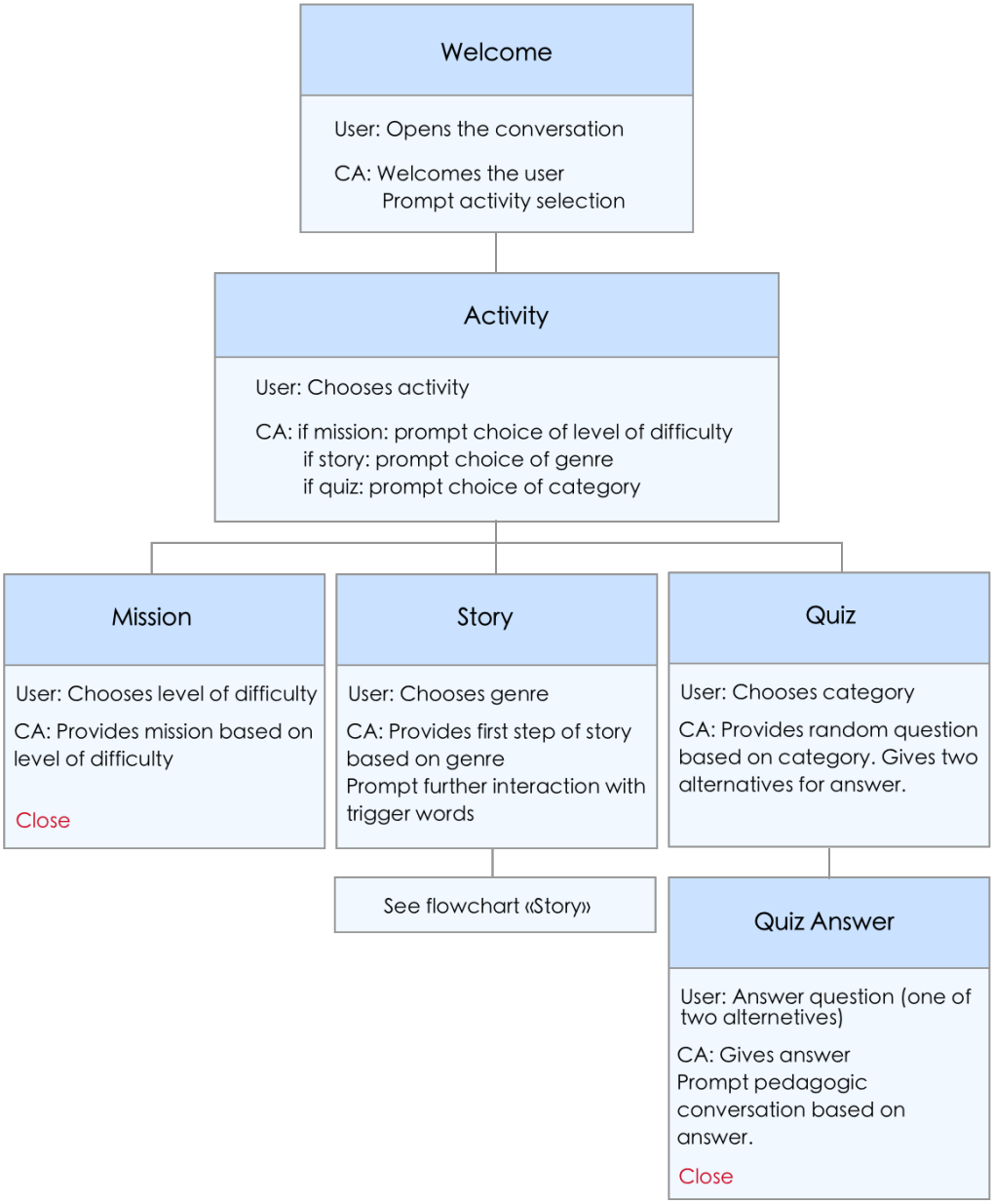


Figure 18 Flow of main intents v2

The quiz

First, the user had to choose a question either about nature or about air (see Appendix G for content). The CA asked the user a question from the chosen category and offered two alternative answers. The interaction ended with the CA providing the user with information about whether they answered right or wrong, as well as some additional information about the topic. The CA prompted a discussion about the topic before the system closed the interaction. The excerpt below illustrates the changes made since prototype v1.

Prototype v1 (previous)	Prototype v2 (current)
CA: «Hi, my name is Luftolino? Do you want to play a game? User: Yes! CA: Great! I'm thinking of a number between one and ten. Guess which one! User: four? CA: Wow! That's right. Do you want another question? User: No CA: That's ok! We can play at another time. Bye Bye!	CA: Hi. I can give you mission, play a quiz or make a story. Do you choose the mission, the quiz or the story? User: Quiz CA: I can give you a quiz about air or nature. Do you choose air or nature? User: Air CA: Get ready for a question about air. [...] Is it bikes or cars? User: Cars CA: Correct! [...] Talk to each other about what we can do to make the air less dirty.

Table 12 Changes from v1 to v2 for the quiz

The story

First, the user had to choose between making a "cool" or "lame" story. If the user chose "cool", the CA triggered the first part of the story and prompted for selection of two new trigger words. The user made three decisions before the story was over (see Appendix H for content and Appendix I for flow). There were eight possible endings to the story as opposed to four in the previous prototype. If the user chose "lame", the CA responded with "I can only make cool stories". The excerpt below illustrates the changes and intended flow of the first part of this activity.

Prototype v1 (previous)	Prototype v2 (current)
CA: Hi, I'm Luftolino. I make stories together with children. Do you want to make a story? User: Yes! CA: The story is about Josephine. She is [⋯] Which path should she take? User: To the woods! CA: Josephine walks further into the woods.	CA: Hi. I can give you a mission, play a quiz or make a story. Do you choose the mission, the quiz or the story? User: Story CA: Let's make a story that is cool or lame. Do you want to make a cool or lame story? User: Cool CA: We'll make a cool story. Once upon a

[…] Should Josephine help Ola to get rid of the garbage? User: Yes *Interaction continues*	time […] Should she go to the left or to the right? *Interaction continues*

Table 13 Changes from v1 to v2 for the story

The mission

First, the user had to choose to receive an "easy" or a "difficult" mission. The CA proceeded to give the instructions in one turn, removing the need for user-input between each segment of the instruction (see Appendix J for content). We added pauses in between each segment enabling the users to get ready for the next instruction. The excerpt below illustrates the intended flow for this activity.

Prototype v1 (previous)	Prototype v2 (current)
CA: Hi, I'm Luftolino. I give missions to children. Do you want a mission? User: Yes! CA: Great. First, you have to find a straw. Let me know when you're done! User: I'm done. CA: Super! Now, put the straw in your mouth. Tell me when you're done. User: Ok! CA: Now, I want you to breathe through the straw for five seconds. Let me know when you are done. User: ok. CA: Great job! I want you to talk to each other and describe how it felt to breathe through a straw. See you later!	CA: Hi. I can give you missions, play a quiz or make a story. Do you choose the mission, the quiz or the story? User: Mission CA: I'll give you a mission that is either easy or difficult. Do you want an easy or difficult mission? User: Difficult CA: You will get a difficult mission. […] Talk to each other about how it felt when the body needed more air when you did an activity that made you exhausted. Bye!

Table 14 Changes from v1 to v2 for the mission

5.5.2 Design decisions in the second version of the prototype (v2)

Concrete guidance using explicit trigger words

In the previous prototype, we guided the user by telling them what their options were but did not always have the CA articulate specific trigger words. In some instances, this resulted in responses that deviated from the desired input. In order to increase the predictability of input, we thought it important to guide the users by offering clear and specific trigger words when necessary. This was also in line with Halls (2018) description of *orientation* as a key moment in interaction with a CA; the CA must offer the right choices at the right time.

Generic feedback in the fallback intent

The CA triggers the fallback intent if it is unable to match the input to available options. However, the technology is unable to detect the exact reason for failure (e.g. the user shouted, responded too quickly, multiple responders, the CA is lacking in capability). Our observations indicated that it could be unfortunate to offer misleading advice, and we sought to avoid feedback that could potentially misidentify the trouble source that caused the need for repair. Consequently, we re-designed the dialogue and structure of the repair-sequences in the fallback intent. In this prototype, the CA initiated repair by providing the user with four attempts to make themselves understood. The output was as follows:

> Attempt 1: *"I did not understand. Try saying it one more time".*
>
> Attempt 2: *"I did not understand what you said. Try saying it again".*
>
> Attempt 3: *"Now I am having trouble understanding you. Let's start over. I can give you a mission, make a story or do a quiz"*
>
> Attempt 4: *"Tell me if you want a mission, make a story or do a quiz"*
>
> Close interaction: *"I'm sorry, I'm having trouble understanding you. Let's try again later. Bye!"*

The first two outputs initiate a repeat. These are generic instructions that do not offer information as to why the CA cannot respond. Usually, the default fallback intent offers the user two retries and closes the interaction after the third failed attempt. Our experience showed that the children occasionally needed additional tries to make themselves understood, which is why we added two more attempts before closing the interaction. However, for the third try, we wanted a shift in the interaction. Therefore, this

output would prompt the user to start over, asking them to choose an activity. The fourth re-prompt clarified what the user's options was and this was the user's last chance to make themselves understood. If the system still failed to understand, it discontinued the interaction.

Tutorial

We identified poorly timed responses and multiple responders as prominent issues in the interaction and wanted to explore strategies to address these problem areas. We decided to implement a tutorial that could support the children's understanding of the technology. The tutorial would communicate important rules for correct interaction with the CA: hold your response until the lights have stopped blinking and talk one at a time. The dialogue below illustrates the tutorial.

> ***CA**: I can help you understand how to talk to robots […] What do you have to wait for before talking to me?*
> ***User**: Wait for the lights to stop blinking*
> ***CA**: Good. You've got it. It can be hard for me to understand when […] How many can talk to me at a time?*
> ***User**: One*
> ***CA**: Hurray! Now you're ready to talk to me. Say "hi" to find out what we can do together.*

Repetition, paraphrasing and handling silence

Repetition and substitution are important strategies for repair in a conversation (Golinkoff, 1986). It was important for the CA to inhabit these abilities in cases where the children initiated repair during the conversation. Moreover, we observed that the children sometimes refrained from responding. We decided to implement three functions that addressed these issues:

- *Functionality to support the user's request for repetition:* the system will repeat the CAs last utterance at the user's request. The system triggers the intent by input like "what?", "I did not hear that", "what did you say" etc. A prefix is added before the repetition ("what I said was", "I'm sorry, I said", or "let me say that again").

- *Functionality to support the user's request for an explanation (paraphrasing):* the CA rephrases the latest output. For each output, we designed an alternative phrasing of the utterance. The system triggers the paraphrase intent if the user implies that he or she does not understand: "I did not understand", "what do you mean" etc.

- *Functionality to prompt the user in the case of no response:* the system notices when the system does not receive any input, meaning that the microphone does not detect a response. The CA triggers the intent after about five seconds of silence and prompts the user to provide a response. The user gets two more tries before the CA closes the conversation after the third time the intent is triggered.

Confirm, explain and ask

We attempted to increase robustness in the interaction by carefully structuring the output. First, the CA would confirm that it understood the request from the user ("you will get a mission…"), and then list the options the user had ("…that is either easy or difficult…") before it explicitly asked the user which of the options they chose ("…do you want an easy or difficult mission?"). Our intention was to ensure the user that the CA understood them correctly. Providing the options twice gave them more time to process the output and make a decision regarding the question. Ending with an explicit question signaled the transition of turns to a new speaker and established an expectation of a reply. We intended the output to clarify that the system expected explicit trigger words.

Summary of design decisions

Design decisions	Originated from
Implement tutorial	Remedy extensive issues regarding the interpretation of system status and multiple responders
Generic feedback in the fallback Intent	Avoid misplaced and misguiding advice
Offer explicit trigger words	Increase predictability of input
Confirm, explain and ask	Attempt to ensure mutual understanding through confirmation and repetition, and mark end of turn

Table 15 Design decisions made in iteration 3

5.5.3 Execution: evaluating the second prototype

We conducted the evaluation at Blueberry kindergarten at the beginning of January. Three groups tested the CA, and eight children ages 3-5 participated in total. This was the first time the children in this kindergarten used the prototype. Our objective was to start each session with the tutorial. However, none of the groups managed to get through this interaction, and we provided a short explanation of how one should talk to a CA before continuing the session. Next, we let the participants decide if they wanted to do the mission, story or quiz. We encouraged them to try all the activities during the session. We sought to avoid facilitating the progression of the conversation because we wanted to see whether the participants would manage to adapt their strategies correctly without our interference. However, this proved difficult. Not once during the four sessions were an activity completed without our interference. The children became visibly frustrated when the CA constantly told them it did not understand and prompted for repeats. In order to avoid further exasperation, we found it best to intervene in the interaction in some instances.

5.5.4 Insights: evaluation of the second prototype

Recurring trouble sources: multiple responders

We continued to note occurrences of previously identified precursors and trouble sources. Multiple responders was the primary precursor of breakdowns during the participant's interaction with the CA. In nearly all the interactions, several children responded, either in succession or simultaneously, most often leading to a breakdown. The following excerpt shown in Table 16, illustrates several precursors for the trouble source related to the group situation. Firstly, we observe how the children were quiet for some time before Caroline self-selected to take the next turn. However, she is unsure of what to choose. The children did not have the time to deliberate on their answer, nor did they have a lot of time to decide. Hannah ultimately made the decision, but the response was overdue. The failure led to frustration, and inexpedient strategies for repair (screaming), again from multiple responders.

Speaker	Verbal Exchange	Trouble Source	Repair Strategy
CA	[…] should Elisa go to the woods or to the city?		
	Silence	No response Thinking	
Caroline	Ehm..theee..	Hesitation	
Hannah	City!	Overdue response	
CA	I did not understand. Try saying it one more time.		Initiating repeat
Hannah	THE CITY!		Increase volume
Caroline	CITY! We're saying THE CITY GOOGLE!	Multiple responders	Increase volume explanation (augmentation)

Table 16 Extract 4

Recurring trouble sources: poorly timed responses

Testing the prototype with both larger groups of children as well as one-on-one, emphasized the profound issues of poorly timed responses. We were aware of this issue before this evaluation and had tried to mediate the trouble source by adding the tutorial. However, the tutorial did not work as intended. The children did not pay attention to the information provided by the CA and seemed to forget to wait for the lights to stop blinking during the interaction. In one session, Hannah tried the CA on her own. The following excerpt shown in Table 17, is a transcript of her first attempt at interaction.

Speaker	Verbal Exchange	Trouble Source	Repair Strategy
CA	(Pling-Sound) Hi! I can give you missions, tell a story, or play a quiz. Do you choose the, story or quiz?		
Hannah	… Mission!	Overdue response	
CA	Hello? Are you there? You have to answer me.		Prompts response

Hannah	Mission!		
CA	I'll give you a mission. Do you want an easy or difficult mission?		
Hannah	Easy	Premature response	
Researcher	(whispers) Try again!		Adult guidance
Hannah	Easy	Unknown – possible misinterpretation.	
CA	I did not understand. Try saying it one more time		Initiates repeat
Hannah	Easy	Premature response	
Researcher	Easy!		Adult Facilitation
CA	I'll give you an easy mission […]		

Table 17 Extract 5

Hannah proceeded to try the story activity, and we observed similar impediments during this sequence as her responses were often either premature or overdue. The CA continuously tried to repair the interaction by initiating repeats. However, the constant need for clarification or repetition made the participants noticeably irritated. Thus, we confirmed that the visibility of system status was a significant problem.

Reactions to a breach of the maxim of relevance

In one instance, the CA triggered the wrong intent and diverged from the trajectory of the story. However, Hannah responded to this output and the participants neither reacted nor commented on the change in activity. Afterward, the conversation took an unexpected turn when the children all reacted to a word used by Caroline, which in turn exemplifies the associative and playful nature of children.

Speaker	Verbal Exchange	Trouble Source	Repair Strategy
Caroline	CITY! We're saying THE CITY GOOGLE!	Multiple responders	Increase volume Explanation (augmentation)
CA	Ohoh, you have to choose easy or difficult.	Wrong intent activated	

Hannah	Easy!		
Caroline	Easy as pepps!	Multiple responders	
Ellen	Pepsi Max!	Multiple responders	
Caroline	Pepsi Max!	Multiple responders	
CA	I did not understand. Try saying it one more time.		Initiates repeat

Table 18 Extract 6

Specificity in repair attempts – what did I do wrong?

The evaluation accentuated the importance of specific responses in case of erroneous or out-of-bounds inputs. We observed that failing to clarify the distinction between telling the user "I didn't catch that. Try again!" and "I don't know how to do this yet!" resulted in prolonged and unsuccessful repair-sequences. During the evaluation with the first group, "Ellen" requested the CA to make a "robot-sound". The CA told her it did not understand and asked her to try again. However, in this instance, it was a case of not possessing the capability of fulfilling her request, and not one of understanding. Caroline, believing that the CA did not understand her request, continued to ask for "robot-sounds" instead of moving the conversation forward. The attempt at repair failed to induce a shared understanding of the problem, thus causing the conversation to fail.

Speaker	Verbal Exchange	Trouble Source	Repair Strategy
CA	(Pling–sound) Hi. I can give you mission, make a story or play a quiz. Do you choose the, story or quiz?	The children are talking with a low voice to each other – unrelated talk	
Researcher	How do we talk to robots?		
CA	I can help you understand how to talk to robots [..] What do you have to wait for before you answer?		
Ellen	Hello, Google can you make a robot sound?		
CA	Pling! Hi. I can give you mission, make a story or play a quiz. Do you choose the, story or quiz?	Triggers welcome intent because of the word "Hello"	

Ellen	Hey, Google, can you make a…		Repeats request
Hannah	Can you make a mission for us?	Interruption Conflicting request	
Ellen	No, hey, Google, can you make a robot sound	Multiple responders Conflicting request	
CA	(Pling-Sound) Hi. I can give you mission, make a story or play a quiz. Do you choose the, story or quiz?	Triggers welcome intent because of the word "Hey"	
Hannah	YES!		
Ellen	Hey, Google, can you help…	Multiple responders Conflicting request	
CA	I did not understand. Try saying it one more time.		Initiates repeat
Ellen	Hi, Google, can you help us find which robot sound we should have?		Repeats request by paraphrasing

Table 19 Extract 7

Experiences with larger groups

Larger groups led to more "noise" during the interactions with the CA, as well as increased difficulties regarding distributing turns between the CA and the participants. Another effect was less time to interact per child, creating long periods of inactivity, which in turn seemed to affect their concentration and motivation to participate. They often forgot or failed to detect the trigger words, diverged from the script and had more difficulties in staying attentive to what the CA said. As previously mentioned, we tried to mediate these challenges by intervening directly to repair the interaction. Furthermore, we offered guidance about what to say and when to say it, either while the CA was talking or before we activated it.

5.5.5 Summary – key insights from iteration 3

- The tutorial failed to ease the troubles we had regarding multiple responders and poorly timed responses. We assume this was largely due to the children not paying attention to the CAs advice, or simply forgetting whilst in interaction. Perhaps this suggests that such guidance should be immediate to the situation, and not provided as a separate element.

- Despite obvious breaches of the relevance-maxim (Grice, 1982) in instances where the CA triggered the wrong intent, the participants did not seem thrown off or frustrated by the inconsistencies. They simply carried on with the interaction without commenting on it.

- Unspecific repair strategies can be misleading and fail to initiate appropriate repair. This became obvious when the CA did not distinguish between "I did not understand" and "I can't do that", resulting in unsuccessful repair-attempts.

- Larger groups seemed to lead to added trouble in interaction and necessitated an increased degree of adult interference through guidance or interventions by direct repair.

- The participants diverged from the script to a larger degree. They would ignore the questions, refrain from using the proposed trigger-words, or set their own agenda as to conversational topics.

Phase 3: understanding the interplay between the actants in use

The objective of the last phase was to explore the interplay of the actants involved in use: the educator, the children and the CA. We conducted a concluding field trial of the prototype in which two educators and children at Raspberry kindergarten used the CA over a period of one week. The final activity was a group interview with the educators who participated, as well as a retrospective with the children where we talked about the project and gave them an "inventors-badge" signifying their warranted status as "robot-inventors".

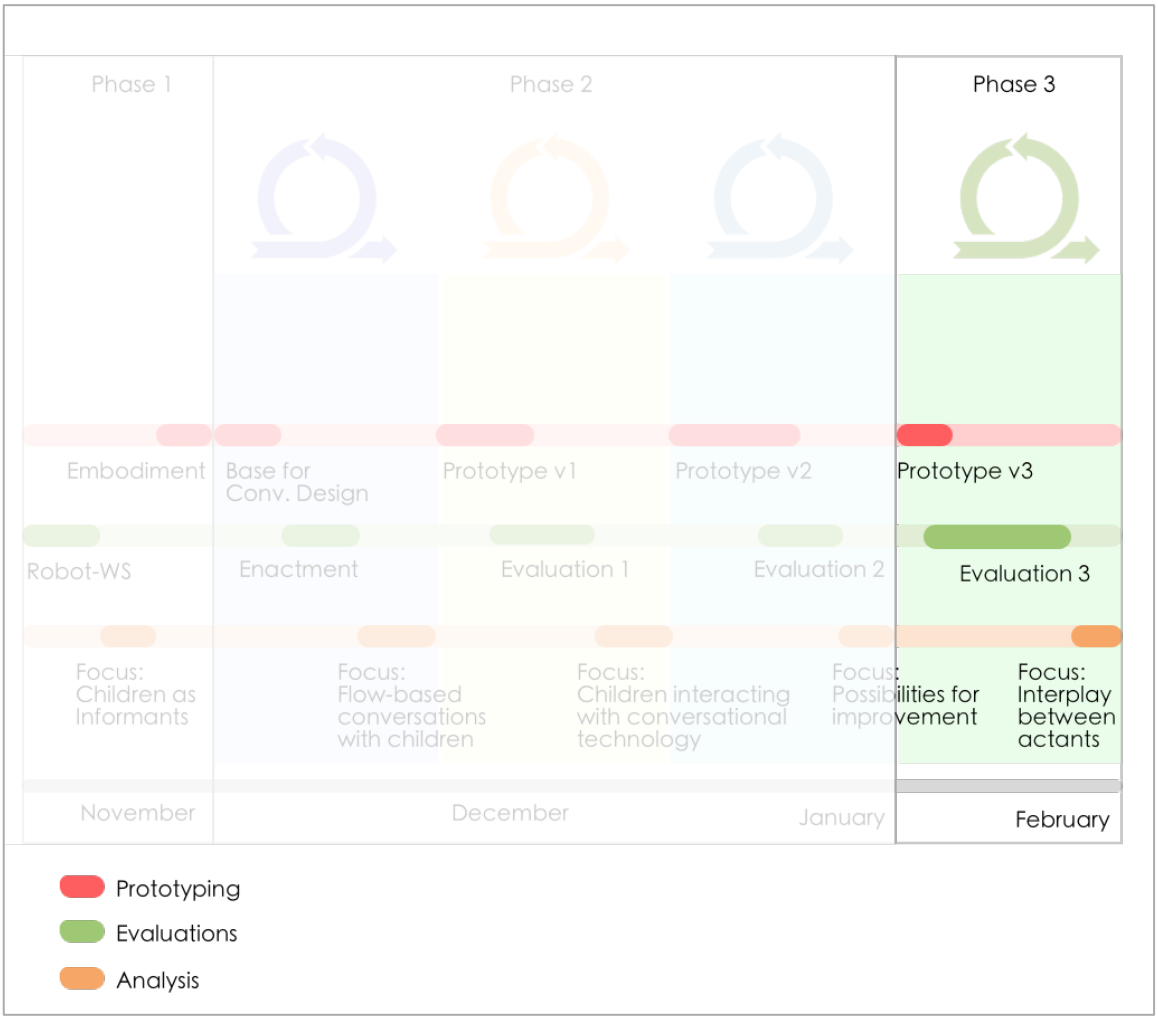


Figure 19 Visualization of phase 3

5.6 Iteration 4: field evaluation of final prototype

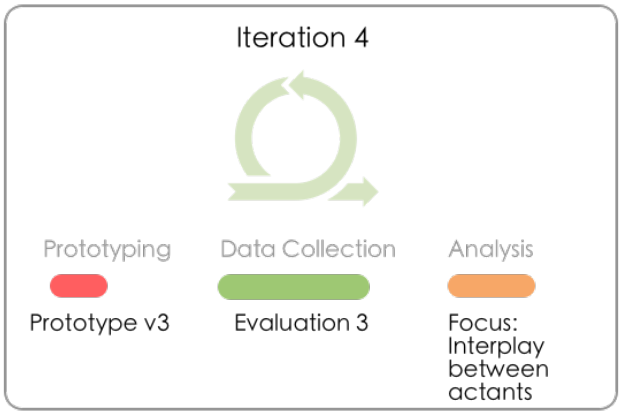


Figure 20 The activities in iteration 4

5.6.1 Design decisions in the final version of the prototype (v3)

We attempted to improve the prototype by making some minor changes before the field trial. These changes were the following:

Increase efficiency

In the previous evaluation, we found that the dialogue lacked adequate efficiency. The stage in which the CA asked the children to choose either level of difficulty, story genre or quiz category, usually ended in breakdowns or prolonged repair sequences, which meant that they seldom reached the actual content of the chosen activity. If they did, they had grown distracted or discouraged by this time. Consequently, we decided to remove this intermediary section of the interaction. Figure 21 shows an updated version of the flow and Appendix K contains the code.

Improve phrasings in tutorial

As stated, the tutorial did not work as intended. None of the groups payed attention or were able to answer the CA during the tutorial without adult facilitation. However, we wanted to try to evaluate this functionality one more time, as we were curious of its potential. Hence, we decided to rewrite the output trying to be more concise, minimizing the information load, hoping this would remedy some prevalent trouble sources regarding timing and turn-taking.

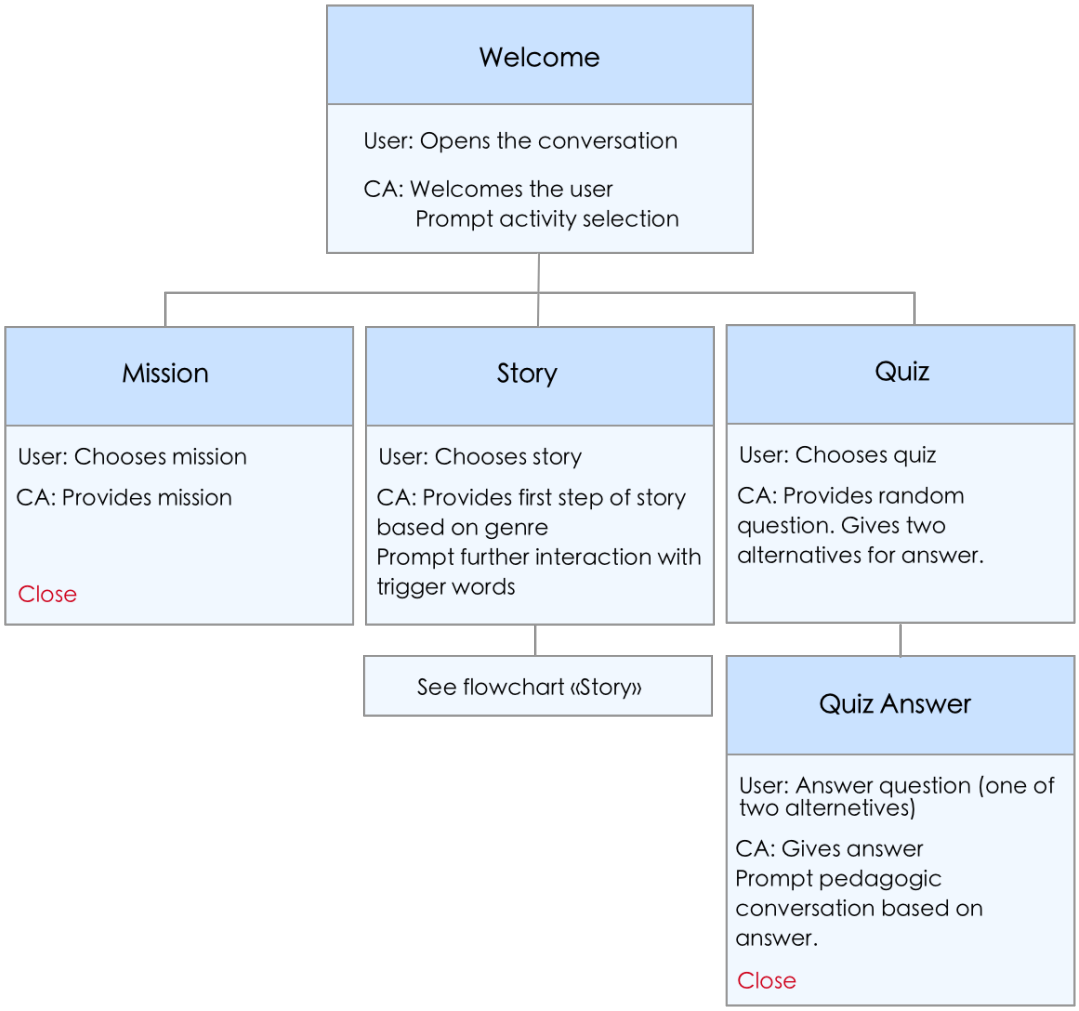


Figure 21 Flow of main intents v3

Improve phrasing when prompting response

During the evaluation of the second prototype, we observed that the CA would trigger the intent for no-input (the system did not detect a response) in instances where the children had answered before the microphone was activated. Thus, the output (see Table 20) caused confusion as the children had already provided an answer. Hence, we decided to change the phrasing so it would initiate repair but avoid confusion if the children responded without the CA detecting it.

Output v2 (previous)	Output v3 (current)
Hello? Anybody there? You have to answer me.	Hello? I didn't hear what you said. Remember to wait till my lights have stopped blinking

Table 20 Changes in the output in the no-input intent

5.6.2 Field evaluation of the prototype

Prior to field trial, we found two educators who wanted to participate in the activities. We arranged for them to use the prototype with at least three groups each. 14 children from Raspberry kindergarten participate in the project, which would allow six groups of 2 or 3. We told them that each group should try the tutorial as well as all three activities (mission, story and quiz). The educators recorded each session, which allowed us to observe indirectly by listening to and transcribing the audio recordings.

We wanted the educators to document their experiences in an elicitation diary (Appendix N), which was meant to be the foundation upon which to conduct a concluding interview. This diary had four sections: we intended part 1 to elicit general feedback about their experiences from using the prototype, whilst the other three contained questions connected to the three activities. The elicitation diary mainly consisted of questionnaires, with yes/no questions and room to elaborate using key words or short sentences. We wanted to know if they were able to get through the activities, if the children payed attention, if the children had any troubles making themselves understood, and how the educator handled breakdown situations. In part 1, we included a semantic differential scale regarding the children's attitudes to the CA (positive-negative), and the degree to which the children understood the CA and vice versa. We were interested in how much they felt they had to interfere to help the children during the interaction, and in what ways they did this. We also added space for the educator to comment on any other notable observation or experiences.

Handing over the prototype

When delivering the prototype, we held a meeting with the two educators in which we explained the plan regarding number of groups, what the activities involved, as well as how to document the sessions. We decided to tell them about the two most prominent causes of error: the children talking before the microphone was on and multiple responders. This way, they would better be able to understand some of the breakdowns that would most likely occur. We ended the brief with a demonstration and let them try the CA themselves. To help them remember everything, as well as provide them with a visual representation of the content, we made a pamphlet (Appendix M) that contained information such as a checklist for the execution of each session, the most prominent causes of error, as well as the flows of each activity. We also added our contact

information in case they had trouble with the technology. We urged them to reach out at any time if they had any questions or experienced troubles.

Execution

Each session lasted between 20 to 45 minutes. All groups completed all activities (story, mission and quiz) in the same session. The groups consisted of 2-3 children and one adult. All groups forgot to try the tutorial and it is thus not a part of the evaluation. When delivering the prototype, we failed to specify to the educators that they were the ones who should activate the CA, which is done by saying "hey Google, talk to our robot". This led them to instruct the children to do it themselves. We had experienced that this phrase was very sensitive to mispronunciations and poor timing. Consequently, it led to major difficulties when the children were the ones who tried to activate the CA. This tedious struggle clearly tired both the children and the educators. Upon returning for the prototype, we conducted a semi-structured interview with the educators based on the elicitation diary. The interview concerned their experiences and thoughts regarding the potential of introducing this technology in a kindergarten setting.

5.6.3 Insights from the field trial

Our findings pertain to the *facilitation* of communication done by the educator before, during and after interaction. We also comment on how the educators supported repair in instances of breakdown – what are their *repair strategies*? First, we will explain how we differentiate these two concepts:

- We define facilitation as what is being done to avoid breakdown (e.g. decide beforehand what activity to choose, remind the children of what the alternatives are when they are to answer, or help the children time their response).

- We define repair strategies as what is done when a breakdown has already occurred, trying to establish a comprehension in order to proceed the interaction.

The incessant breakdowns in the interactions between the children and the CA required the educator to intervene extensively. The field trial emphasized their indispensable supporting role as repairer and facilitator during use. The educators compensated for the limitations of the technology by coordinating the interaction through distributing turns

and assisting correct timing, scaffolding cognition by offering partial repeats and encouraging attention, correcting and guiding in instances of incorrect input or behavior. The extent of their interference highlighted the troubles the children experienced during the interaction, and the communication primarily transpired between the children and the educator rather than the children and the CA. Next, we present the findings related to what happened before, during and after interaction.

Before interaction

We found three distinct ways the educator facilitated before the interaction started:

1. *Distributing turns:* the educator prepared the children of who was to talk ("we have to talk in turns, and X can start") or assigned the task of "speaker" to a specific child. As the CA cannot indicate whom is to respond, the educator took on this function during the interaction.
2. *Activity selection:* the educator initiated choice of activity before activation of the CA, either through discussion ("what should we do?") or by making the decision ("let's do story next").
3. *Preparation for selection of specific trigger words:* the educator initiated choice of trigger words in the story-activity if they were aware of what was to come (e.g. "we will choose story and then we will choose left, ok?").

During the interaction

The educators were constantly involved in conversation with the children during use in order to do direct repair, initiate or aid repair, repeat content the children had forgotten, distribute turns, guide in instances of detrimental behavior, assist correct timing of responses and make the children focus. This transpired in the following ways:

1. *Direct repair*: the educator interacted directly with the CA in order to clarify the children's intent. This usually occurred when the children responded to early, had repeatedly tried to repair unsuccessfully, or when they had to start an activity over because the CA failed to understand them.
2. *Initiate or aid repair*: the educators pointed out instances where the children had to repeat words or helped them understand what they had to correct, "don't say the word so many times or it won't understand". Sometimes, they would help them choose the correct trigger words, "try saying *save*", or "try the other one, *left*". If the children failed to respond, the educator would prompt for selection of answers

by repeating the alternatives “do you choose left or right?” The educators frequently repeated content in instances where the children forgot what the CA had asked them. If they forgot specific trigger words, the educators would repeat these in a whisper, “left or right?”, “what doesn’t belong in the forest: garbage or birds?” etc.

3. *Distribute turns*: during the interaction, the educator would sometimes appoint speakers while the CA spoke: “you can answer next”, “now it’s Emily’s turn”.
4. *Guiding*: in instances where the children behaved in ways that were causing breakdowns, the educators would offer corrections or advice such as “you don’t need to talk so loudly”, “talk louder”, “you have to say it properly or it won’t understand”, or “don’t talk with your hand in front of your mouth”.
5. *Assist correct timing*: the educators would provide physical cues (point) and whisper “now, now!” to the children who had difficulties with timing their responses.
6. *Encourage focus*: throughout the sessions, the educators had to motivate and make the children focus on what the CA said; “you have to answer the robot”, “did you hear what the robot said”, etc.

After interaction

Before closure, the CA prompted the users to instigate a pedagogic conversation. Even though what happens next is not a part of the interaction with the CA, we find it interesting to comment on this since an important part of the CA’s pedagogic objective is initiating discussion and reflection.

1. *Seeks affirmation of the children’s understanding:* in instances where the educator suspects the children were not paying attention, they asked if they heard what it said: “did you hear what it said? What was the end of the story?”
2. *Ask opening question:* in nearly all instances, the educator asked an opening question to start the conversation the CA instigated: “what can people do so that animals do not get stuck?” or “how can we avoid garbage in the nature?”
3. *Repeat the output:* when the children had not payed attention, the educator saw the need to repeat parts of, or all, the output. In some cases, this was done in combination with an opening question: “now it said that we had to talk together about how it felt to breathe when we jumped up and down. How do you breathe?”

The children's behavior and the educator's thoughts

We found that the children depended on the educator to a large degree, and that there were several instances of collaboration between the children. For example, when they had trouble getting the CA to understand them, they asked other children for help, or children would suggest trigger-words their friend had forgotten. The participants seemed to make an effort to communicate with the CA, and there were few instances where they diverged from the script. Regarding breakdowns, we noted that the usual difficulties arose in the interaction, which seemed to result in the children leaning heavily on the educator to help them through the interaction. In some instances, the child who spoke to the CA would ask if someone else could take over ("I'm not able to do this. Can someone else try?") or ask the educator for help when they could not remember or did not understand ("what did he say again?", or "can I say this?"). The CA's capabilities regarding repair are invisible to the children. Hence, they never initiated repair by prompting repeat. The CA could repeat its last utterance, but unaware of this, the children addressed the educator instead. Additionally, we saw that they usually asked about specific part of the utterance "what was the first thing he said", or "what was the last choice?".

In the interview, the educators expressed that even though the children seemed happy and willing to participate, they had trouble concentrating. It seemed like the children found it hard to understand and hear what the CA said, and the monotonous voice did not help concerning their ability to stay attentive. They experienced that the children had troubles paying attention to new information as they became preoccupied with what had just happened.

5.7 Concluding the data gathering process

The field trial was the last activity we implemented along with participants from the kindergartens, and it marked the end of the data gathering process of our project. At this point, we had conducted a case study and a design process constituting three phases and four design-iterations during which time we had constructed an understanding of the issues that complicate child-CA interaction. Some of this understanding manifested along the way as we conducted the activities and analyzed the data material after each evaluation. However, some insights also emerged in the aftermath of the process, because

we had the time to reflect upon the information thoroughly and review the project in its entirety. In the succeeding discussion, we present our full insight in relation to relevant theory presented in chapter 2.

6 Discussion

In this chapter, we attempt to answer our main research question:

What does the concept robustness entail in design of CAs for preschool children?

We will first outline some of the main realizations that stemmed from applying broken world thinking as a perspective in this research project and design process. We find that it accentuated communicative repair work as an integral part of the interaction between users and CAs and emphasized the main challenges that derive from the complexity of cooperative meaning-making. However, we also comment on our thoughts regarding what this perspective ultimately obscured during the design process. We argue that primarily, it led to a design that was severely restricted in its rigidity. Finally, we offer a discussion on why we argue that designing robust CAs for young children entails supporting the user in understanding the constraints that limit possible behavior in the conversation, understanding and supporting the collaborative nature of communication, and lastly understanding what constitutes an effective conversation.

6.1 Broken world thinking as a perspective in a design process

We begin by examining the insights and design results brought about by making communicative breakdowns our starting point. We suggest that making breakdown our main investigative lens aided our understanding of the challenges inherent in the interaction between children and a CA, and thus enabled us to recognize what robustness entails regarding our chosen technology, users, and context. However, we assert that in some ways this focus became unfavorable to the design endeavor as it obscured the value of certain design choices resulting in the elimination of all elements that we assumed increased the fragility of communication between the children and the CA.

6.1.1 A repair-centric view on human-CA interaction

In Chapter 2, we defined breakdowns within the field of technology and design as a phenomenon generally describing situations where technologies fail to work as intended. In this regard, we introduced Jacksons (2014) thoughts on breakdowns as moments in

which the activity of repair makes formerly hidden aspects of technologies visible, open to inspection and available to change. He asserted that breakdowns are not just moments of disintegration, but that they can be vehicles of innovation and provide the repairer with a unique vantage point because it allows them to "learn, see and engage our technology in new and surprising ways" (Jackson, 2014, p. 230). In this project, we have attempted to apply this perspective in a slightly different way than what Jacksons initially suggested. He positions himself within technology and media studies and put his focus on technological artifacts and systems by proposing that broken world thinking can reframe our outlook towards more sustainable models of production and use. We, on the other hand, brought this perspective into a design process by making breakdown our starting point when assessing communication between young children and conversational agents (CA). We echoed Jackson in this regard, "what happens when we take erosion, breakdown, and decay as our starting point in thinking about the nature, use, and effects of information technology?" (2014, p. 221).

In relation to this project, broken world thinking had us position communicative repair at the center of our design process. We have defined a communicative breakdown as incidents where the actants fail to achieve a mutual understanding of one another's intents and responses, ultimately resulting in unsuccessful interaction. The act of conversing is a cooperative and coordinative practice often exposed to moments of breakdown, which demands detection of trouble sources as well as initiation and implementation of communicative repair work. Thus, conversations, like the world, are continuously breaking (Jackson, 2014, p. 223). Our research has made evident that this is unquestionably true in interactions that transpire between young children and CAs – perhaps indicating the need for a repair-centric view when designing such technologies.

6.1.2 Communicative repair as an integral part of a conversation

Let us reiterate the definition of communicative repair. Conversational or communicative repair is loosely defined as the work done by speaker, hearer, or both, to correct an unsuccessful part of the communication (Richards & Schmidt, 1983, p. 147). Making breakdown the main lens through which we tried to understand child-CA interaction revealed the necessity for repair as well as the embedded nature of communicative repair work. Jackson talks about repair or maintenance of artifacts as distinct activities related to the production and use of physical things in the world; they are thought of, designed,

produced, used, discarded and perhaps repaired or maintained sometime within this period. Thus, regarding physical objects, repair work is to a certain extent outside of, or separate from, regular use. Conversely, communicative repair is always conducted within the conversation, and if needed must be implemented during use. Consequently, we argue that the way in which communicative repair is as an integral and embedded part of use increases the need for designers to understand the different precursors to breakdown and implement effective strategies for repair in conversational technologies. When considering the high likeliness of breakdown in the interaction between children and CAs, this rings even truer.

Jackson suggests that there are differences to be found in the world of repair (2014, p. 229). Although this comment refers to cultural variations and differences in what he called repair ecologies, we believed that it was equally interesting to examine how traditional types of repair work is similar or distinct from communicative repair. This comparison enabled us to understand how the activity of communicative repair fit into the overall process of human-CA interaction. There are several important distinctions between breakdown of physical artifacts and breakdowns in communication, which in turn affects the activity of repair. We believe the main difference relates to their nature, or more precisely their way of being in the world; a physical artifact is concrete, enduring, and made manifest through materials, whereas a conversation is immaterial, transient, situational and made manifest through spoken words and shared meanings. We suggest that these distinctions influence the *immediacy of repair*, which we define as the level of urgency connected to the implementation of repair work. We can inspect a physical object as it is now; its qualities are primarily tied to tangible or perceivable attributes, which in turn affects its functionality. We can examine the artifact and detect the trouble source of breakdown after it has occurred. By example, we can study an error message and take time to discern its meaning. We then decide upon the most efficient repair strategy and the most convenient time for doing repair work. Communicative repair, however, requires immediacy; participants in a conversation should carry out repair work in close proximity to the occurrence of the trouble source in order to avoid cognitive or organizational challenges regarding the alignment of topical and referential coordination (Hutchby & Wooffitt, 1998). That means the users and the technology operate under rather strict time constraints; communicative repair must happen instantly upon breakdown, and the efficiency of repair strategies will only become apparent in relation to the construction of

mutual understanding. Equally, other forms of human-computer interactions also necessitate error handling similar to communicative repair. However, in spoken human-CA conversations the degree of immediacy surpasses that of responding to a message presented on a screen. Thus, designing for robustness would entail supporting the user in quickly assessing the cause of the breakdown and implementing appropriate action. We elaborate on this later on in the discussion.

6.1.3 Obscuring valuable design choices

Establishing breakdowns as the primary focus in the design process led to the discovery of a range of trouble sources, some of which we attempted to address in our design and some that we were unable to remedy due to technological limitations and other constraints. Although a broken world perspective might have helped our research objective, it also constricted us in the design process by obscuring the fact that rigidity in the form of structure and increased predictability of input does not necessarily equate robustness. Consequently, we failed to create a design that addressed the playful and social nature observed in the children, which led us to question whether we can characterize a CA that is unable to consider or handle this behavior as robust.

There were several reasons why designing for breakdowns ultimately became detrimental to the design outcome. Primarily, our attempts to prevent breakdowns from happening led us to eliminate what we now consider valuable design choices and narrowed our focus in exploring different ways of designing for robustness. During the development of the prototype, we decided to remove several sequences in the conversation. These decisions originated from a design paradox, which suggested that in order to decrease the opportunities for errors we also had to reduce the length of the conversation. Hence, we removed all parts of the conversation that we considered redundant to the objective of the task. By removing these stages in the dialogue-flow, we left little room in the conversation for the user to occupy. To provide an example, in the final version of the prototype, the only input required of the user during the mission was saying one single word – “mission”. Yes, we did minimize the possibility of breakdown when limiting the children’s role in the interaction with the CA, but we also diminished – perhaps completely removed – enjoyable interaction that could contribute to creating natural and engaging conversations. This had us wondering; did we indeed design the CA to be

conversational or was the prototype merely a machine that offered basic interaction based on limited voice input.

Furthermore, the tediousness in our way of organizing the dialogue was counterproductive to creating a natural conversation. The structure of confirmation – information – repetition proved unnatural, ("Elisa goes to the left. […] (confirmation). She can go to the city or to the woods (information). Should she go to the city or to the woods? (repetition)"). Structuring the output this way provided information to the user as to what their alternatives were for interaction and increased the predictability of input by explicitly offering the correct trigger words. We suggest that this way of designing the output perhaps indicated distrust in the user's ability to conduct conversations. Furthermore, it might have led both children and educators to consider the possibilities for interaction as being more limited than they actually were, as the output seemed to give the impression that there was only one correct way of providing input: saying one specific word. Hence, this dialogue design did not exactly set the stage for interaction similar to the way we conduct ordinary conversations – and thus discarded Hayes and Reddy's (1983) principle of *graceful* interaction. We now find that designing a robust CA does not require it to impose strict control over the conversation in this way. Rather, our focus should have been on supporting the users in the event of a breakdown, and on developing graceful strategies for repair e.g. suggesting the correct trigger words *after* a breakdown. We will elaborate on what we find this to entail later on in the discussion.

In hindsight, we saw that the dialogue design allowed little leeway for the children to express themselves naturally, counteracting the principle of graceful interaction. This made us question if the dedicated focus we put on breakdowns as a resource for understanding robust dialogue design confined us during the design activities and led us to produce a contrived and constricting prototype. The positive aspect is that it helped us realize that excessive attention to breakdown can in fact devalue the importance of graceful interaction in conversations between children and CAs. We find it important to keep in mind that robust communication is not exclusively the minimization or elimination of obstacles, but rather the ability to recover from misunderstandings and communicative bumps in the road.

6.1.4 What becomes visible in moments of communicative breakdown

During the concluding sessions with the children, we asked them if they remembered what the CA could do, and one child answered, “it knows how to not understand us!”. We cannot overestimate the importance of recognizing what robustness entails in design of CAs for children, as well as knowing how to design in order to not only prevent but also support the user in recovering from breakdown without exhausting them. Our observations suggest that the children usually understood what the CA wanted them to say or do. However, poor timing, difficulties in handling turn-taking, and flawed Natural Language Processing (NLP) were prevalent trouble sources that precipitated breakdowns throughout the children’s interaction with the CA. The efforts of getting the CA to understand them was strenuous to the point that the actual content of the dialogue was ignored, or interaction was terminated before completion. This made us question if the breakdowns that resulted from these technological constraints prevented us from evaluating the actual dialogue design. We contend that the answer to this is yes. We will elaborate on this in Chapter 7.

Heidegger proposed that we relate to equipment primarily through its functionality during use, but that its specific attributes might become visible in moments of breakdown (Susi & Ziemke, 2005). Thus, the hammer as an object of inspection become our focus when it ceases to perform its expected function. To us as researchers and designers, the retrospective analysis of the breakdowns in conversation provided us with insights into the challenges inherent in the communication as well as poor design choices. Thus, during the field trial, it became evident that our attempts at minimizing and supporting recovery from breakdown had been unsuccessful. We saw that the foci in these sessions was saying the trigger words correctly, at the right time, in the right way, by the right participant. Consequently, we believe that from the user’s vantage point, the *lack of competency* is what became most noticeable when the CA failed to work as intended. Our experiences suggest that this can significantly affect their evaluation of the CAs capabilities since the consequences of experiencing continuous breakdown are obviously detrimental to the user’s assessment of the technology. This substantiates the importance of enhancing the CAs ability to handle and prevent breakdown because a highly fallible CA will not perform its intended function - the activity of learning through conversation never becomes the primary focus in use.

6.2 What does robustness entail?

In Chapter 2, we presented different ways to understand robustness. We applied a definition presented by Dix et al. (2004) pertaining to the domain of interaction design, stating that robustness concerns the level of support the user is given to successfully achieve their goal. They argue that robustness affects a system's usability and establish four characteristics that they believe the concept to comprise of *observability*, *recoverability*, *responsiveness* and *task conformance*. In this chapter, we discuss these principles in relation to the insights from our research in order to understand what robustness entails in design of CAs for preschool children and perhaps for others as well.

6.2.1 *Observability*: understanding constraints in the interaction and correctly assessing capabilities

Observability concerns whether the users are able to discern enough information about the internal state of the system in pursuance of task completion. Take online shopping as an example. In this situation, the user's goal is to find an item they want and buy it. Hence, it is important that the user can look at the screen and quickly recognize which stage of the purchase process they find themselves. Moreover, the site should provide filtering-options in order to find the right items or perhaps suggest possible words in a search. Lastly, it is important that the user can navigate through observable states, for instance by enabling effortless navigation between the different stages in the purchase-process. In sum, these qualities support observability so that the user can easily achieve their objective.

Participation in a conversation requires the children to receive a message correctly and remember and process the message so they can engage with the sender in a comprehensible manner. As mutual understanding is an important prerequisite for reaching our goals in interaction, we argue that observability entails both sending perceivable and understandable messages and supporting the children in understanding how to provide perceivable and intelligible responses in return. Observability in verbal interaction differs from traditional visual interfaces, partly because of the persistence of the presented information (Dix et al., 2004, p. 271). This necessitate awareness of the cognitive load the output places on the user. The output should be coherent and appropriate to the age group, and not burden the children with too much information.

Furthermore, the CA has to support coordination and provide mechanisms that enable the children to balance irregularities in the conversation (Moore, 2018). When using the prototype we designed, the children were required to either follow instructions, answer a question, or pay attention to a story in order to make choices or provide answers. We observed that the prototype did not support them adequately in these deceivingly simple tasks, because they had trouble understanding when to talk, understanding what the CA could help them with, and detecting what type of responses the CA expected them to provide. Hence, these trouble sources hindered the progress of the conversation and prevented them from achieving their goals. Next, we discuss these issues and the implications in more detail.

When can you hear me, and who should talk?

What elements of the internal system state should be readily available to the children for them to complete the task of providing satisfactory responses? Our research showed that the children had to recognize and adjust to restrictions in the timing of responses as well as the organization of turn-taking. The familiarity of the verbal mode of interaction – the conversation – suggested that the participants could behave as they were accustomed to in conversations with people. However, the technology did not efficiently communicate neither how to correctly time their responses nor how to conduct turn-taking conversations in interaction with a CA. Thus, the CAs inability to adhere to the children's mental models of the rules that usually govern conversations created profound challenges. To provide an example, only one of the three allocation mechanisms that govern the distribution of turns between speakers is applicable in this context, namely *self-selection*. The transition of turns to one specific participant would require the CA to select the next speaker, which it currently cannot do. We observed that the lack of specific distribution of turns between speakers often resulted in multiple responses or no response at all. Thus, breakdowns occurred because the CA demanded a response in order to continue the interaction, but only from one participant. It is demanding to unlearn or change ingrained behavior and our attempts at remedying some of these challenges by teaching the children about the "new" rules for timing and turn-taking proved unsuccessful. Consequently, the breakdowns following these trouble sources persevered and necessitated an excessive amount of adult interference during the sessions in the field trial.

We believe the abovementioned observations indicate that robustness in relation to design of CAs for children should entail clarity and visibility of system status regarding when the children can respond, especially because the technological restrictions diverge from what is considered normal conduct in conversations. Equally, the CA should communicate the "rules" for turn-taking more efficiently, or handle this in some other way. We provide suggestions in Section 6.3.

How can you help me?

Luger and Sellen (2016) assert that supporting correct evaluation of the systems capabilities and intelligence is critical in order to develop the users' mental models in a way that will promote system usage. During our attempts to support design for robustness in child-CA communication we identified unknown functionality as a significant issue. To provide an example, the possibility of asking the CA to repeat what it said was not evident to the children. During interaction, the children were required to remember the information and options provided by the agent. This created a cognitive load, which we saw that they had some difficulties in managing. Thus, it emphasized how important it was that the CA was able to repeat itself. However, during the evaluations, we observed an imbalance in the distribution of initiation and implementation of repeats as a repair strategy. We only observed *other-initiated self-repair*; the CA (receiver) initiated repair by prompting the children (speaker) to repeat their response. Our prototype had functionality that would allow the children to prompt repeats of the CA, but the children never attempted to trigger this intent. Rather, when in need of reminders they turned to the educator or to us. Thus, the possibility of asking for repeats would remain unknown if the children refrained from signaling these problems in interaction with the CA. Consequently, we suggest that observability also entails supporting the user in assessing the possibilities for cognitive support correctly. As it is with people, the CA has to speak in order to mediate its internal state. In this case, the agent's internal state is the current position in the dialogue-flow and the desired information concerns what the user must do to progress the conversation. For instance, if the CA perceives a delay in response, it might assume that the user did not detect the information and repeat the options for action: "do you want Elisa to save Freddy or go for a swim?" Alternatively, it might offer to repeat, or at some point mention that it has the ability to repeat: "I'm sorry, I didn't understand. Do you want me to repeat the options?".

Tell it like it is!

A prerequisite for providing satisfactory responses is that we understand what type of reply the sender expects to receive. Consequently, we believe that observability means creating a robust dialogue that supports the user in interpreting the message correctly. We observed that the children made very literal interpretations of the CA's utterances, which made the slightest ambiguity a precursor to breakdown. The following excerpt illustrates how one child interpreted the introductory message in an unexpected way:

> ***CA**: Boo! I love to make stories, give quizzes and give missions to children. Would you like to make a story, play a quiz or get a mission?*
>
> ***Louise**: Yes!*

An adult might infer that the CA is prompting them to make a choice by phrasing the questions this way. However, this phrasing does leave room for interpretation, and the answer provided by the child is by no means wrong. Hence, we suggest that in designing for observability in dialogue for children one should avoid ambiguity and adhere to the maxim of *manner* (Grice, 1982); be specific, brief and unequivocal in one's statements. Hall (2018) also describes that in the *action-phase* related to the four key moments which is essential for a user's relationship with a system, the users ought to perceive the exchange as non-ambiguous. As a designer, you should ask yourself "can this be interpreted any other way than what I intend it to?".

6.2.2 *Recoverability*: specificity in cooperative repair work

The principle of *recoverability* applies in situations where errors occur and demands that the system provides the user with sufficient resources to take corrective actions that are in furtherance of their objective of use (Dix et al., 2004, p. 271). Dix et al. distinguish between *forward error recovery* and *backward error recovery* where the former means recovery by progressing with the course of action while the latter means undoing an erroneous action. However, only forward error recovery is relevant to our context of use, as the conversation as interaction mode do not allow us to undo our actions or utterances – rather we correct them. Therefore, the user has to "accept the current state and negotiate from that state towards the desired state" (Dix et al., 2004, p. 272). In human-CA interaction, we suggest that this will entail either the user or the CA calling the others attention to an unsuccessful part of the interaction and provide the necessary recourses to

remedy the breakdown and progress with the initial line of pursuit. When using the prototype, the children often encountered troublesome repeat sequences, which did not provide them with sufficient information for identifying the correct strategy for repair and prevented the conversation from moving forward. Thus, the attempts at corrective action were ultimately unsuccessful or so consuming that the children became visibly frustrated and exasperated. We elaborate on what types of resources the CA might provide to remedy this in the following section.

Be specific, please!

We suggest that the lack of successful repair is essentially a result of the prototypes inability to offer specific information at the right time. Suchman (2007) argues that coherence in communication, as well as collaborative meaning-making, is dependent upon the availability of communicative resources. The first problem we identified in relation to jointly conducting repair work was consistent with her remarks on the asymmetries inherent in interactions between humans and machines; machines lack the capacity for receiving sensory input and are thus unable to make use of all situational cues that signal the need for corrective action (Suchman, 2007, p. 11). The need for immediate repair upon breakdown precipitated a need for the dialogue to support the user in making swift assessments of the trouble source as well as deciding upon an appropriate strategy for repair. The CA was lacking in its ability to detect and differentiate between specific precursors or trouble sources and was thus unable to respond with the level of specificity required in order to support the users in their efforts to identify appropriate repair strategies. For example, when the children talked either too loudly or too quietly, it could be beneficial if the CA told them this so that they could adjust the volume of their voice correctly. We saw that the prototype was unsuccessful in adequately supporting the users in this regard. Thus, the unspecific or generic prompts that initiated repair sometimes resulted in miscomprehension of the trouble source and implementation of inappropriate repair strategies, for example, when the CA was unable to distinguish between its inability to perform or respond to a certain request ("I cannot help you with that") and the need for repeat or paraphrasing ("I did not understand. Try saying it one more time"). Furthermore, we noted that in instances where the children initiated repair, they seldom required repetition of the entire response, but asked for partial repeats such as "what was the first thing it said, again?". We consider this another argument for the importance of specificity when supporting efficient cooperative repair work. Both the

user and the CA ought to be specific in their repair initiations, as this will benefit repair work. This argument is in line with *explicit indications of incomprehension* related to graceful interaction (Hayes & Reddy, 1983), where the speaker is specific when indicating what they did not understand as a way of helping their interlocutor in repairing the conversation. Additionally, it adheres to the maxim of *quantity* (Grice, 1982), which states that the actants in the conversation have to assess and contribute the right amount of information the situation requires – no more and no less.

6.2.3 *Responsiveness*: indications of data processing and the challenge of silence

The principle of *responsiveness* pertains to the rate of communication between the user and the system (Dix et al., 2004, pp. 172–173). It is important that the user perceives the system to communicate changes in its internal state efficiently. On occasions where immediate response time is unobtainable, the system should indicate that it has received a request for action and that it is working to create a response in return. For example, if the user has clicked the purchase-button the system should indicate that it has received the call to buy the items and that it is processing payment. If it fails to do so, the user will become uncertain as to whether they were able to attain their objective.

Responsiveness is also an essential element in conversations as the lack of a response leaves the communicative space open to interpretation and thus creates uncertainty in interaction. In this regard, we reiterate the convention of *implicit confirmation and acknowledgment* that Hayes and Reddy (1983) presents. The former refers to how the speaker assumes that the listener has received their message correctly if they do not indicate otherwise (initiate repair), and the latter refers to how the speaker assumes that the message has not been received if the listener fails to produce a reply. Hence, communication is largely founded on implicit assumptions, which is challenging considering that we never observed the children explicitly prompting the CA if they did not understand something or missed parts of the message.

In face-to-face conversations, response-time is usually immediate, and many cues besides explicit utterances indicate the receiving of a message. It might be the furrow of a brow, a pensive look, or a simple “hmm”. These are all communicative resources, which express that the actant is creating a reply by processing the message. Hence, responsiveness

largely involves paralinguistic signals. Our research indicates that responsiveness is an issue we should take seriously because of the expectations conversations as interaction mode creates regarding immediacy.

Silence – I need time to think!

Silence is an underrated communicative resource which current technology fails to account for. Our evaluations of the prototype suggested that silence is a property of communication that challenged the robustness of the interaction, mainly due to the technology's inability to interpret silence as anything other than the lack of response. Silence is an inherent property of situated talk and people apply it differently depending on the context. We can use silence for dramatic effect, to signal hesitation or uncertainty, to indicate the transference of turns at talk, or simply suggest that we are thinking of a response. Additionally, silence might indicate that paralinguistic signals are currently occupying the communicative space. There is always a chance that the participants in a conversation might interpret silences differently. We observed that silence often manifested when the children were uncertain of what to answer. Conversely, the children wrongly interpreted the silence that occupied the time between the end of the CAs response and activation of the microphone as an indication that they could reply.

The fact that the children are not necessarily familiar with the educator's agenda for use constitutes a different premise for their interaction with the CA. Unlike most pragmatically motivated conversations, the children did not have a predetermined aim such as ordering a flight ticket or writing a message. When performing such tasks, the inquiries from the CA would most likely demand little time to decide upon an answer because we expect questions like "when would you like to travel?" or "for whom is the message for?" In our context of use, however, the user is unaware of what to expect in the interaction. Hence, they sometimes needed time to decide which alternatives to choose.

Hayes and Reddy (1983) state that for a system to be graceful in its interaction, it has to indicate to the user if it needs time to process data so that the user does not implicitly assume that the system failed to receive their call for action. Linguistically, phrases such as "hold on just one moment" would be sufficient in this regard. Equally, we suggest that during communication between children and a CA, the signaling of "information processing" should go both ways. We observed that the children often needed time to think in order to decide upon a response for the CA. Hence, the user should also be able

to indicate the need for longer response time. The question is whether children would explicitly signal the need for more time in a conversation, or if this would be another case of "hidden functionality" similar to the possibility of repetition. One way of addressing this could be that the CA indicates this option in the following way: "do you need more time to decide?". This could be an alternative to the output our prototype provided ("hello? I didn't hear what you said"). If the user confirms the need for more time, the agent can follow up with a friendly reminder of the alternatives for action: "ok! You can choose if Elisa should go to into the woods or to the city".

6.2.4 *Task conformance*: enabling an effective conversation

Related to the context for the pedagogic CA, we propose to supplement the principle of *task conformance* (Dix et al., 2004); the degree to which the system services support all of the tasks the user wishes to perform and in a way that the user understand them, with what Dubberly and Pangaro (2009) describe as an *effective conversation*. As explained in Chapter 2, the definition of an effective conversation is one in which the participants are changed in some way that has lasting value to them, e.g. having gained an understanding of something, started to reflect upon a topic, or learned something new. They describe a dialogue as entailing stages, wherein the last stage the participants might agree to perform an act beyond the conversation. We compare this to the underlying objective of use for the pedagogic CA, which is to promote acts of sustainable behavior by supporting the development of an environmental identity.

This had us wondering, does it constitute a breakdown if an effective conversation fails to manifest? We argue that every child in the group is a participant of the conversation because we intend them to inhabit the role of active listeners. The message conveyed by the CA is meant for the group as a whole and not just the children who speaks directly to the CA. As in a classroom, the overarching goal is to have everyone learn about the current topic. This is a difficult job, but nevertheless the objective. We compare the classroom context to the group setting in which the CA is situated. Can we say that a teacher has failed in their job if only some children learn what they are supposed to? We think not, and ask the same question regarding the CA; has the CA failed in the interaction if only some of the children in the group pay attention? We choose to look at the group context as entailing multiple conversations and we understand each child as

having their own conversation with the CA, regardless of the role they choose to inhabit (speaker, listener, or both). Some of these conversations may break down if the CA fails to keep the child's attention. However, this does not mean that all the conversations have failed to achieve their purpose.

We consider the succeeding pedagogic conversation an extension of the interaction with the CA. Inattentiveness might exclude some children from benefiting from this discussion due to insufficient grounds for participation, thus we should strive to support attention and understanding of the content for every child in the group. Hence, we consider robustness in a group context to entail supporting all the children throughout the interaction. We find it important to engage and contain every child's attention as those who passively observe should actively listen. During the interviews preceding the field trial, the educators told us that the children needed support in staying focused and that concentration was an issue during use. We observed how the educators use concretes in their practice in order to ground the conversation, in addition to applying paralinguistic resources such as eye contact, prosody, and body language to engage the attention of the children. The CA does not currently possess these skills. We consider it important to explore how CAs can utilize paralinguistic strategies as a means for keeping the users attentive, and whether there are other ways to support focus and concentration in this context of use. This is something we comment on further when we put forth our recommendations for design and future work.

The group as a social resource for the effective conversation

During the design process, we identified the group as a trouble source in the interaction. In this section, we change our perspective and discuss the possibility of designing a CA so that the group becomes a *resource* in achieving an effective conversation. In her thesis, Geirbo (2017, pp. 242–243) writes about how social resources can make technology function. She exemplifies this using Claude Fisher's (1994; Geirbo, 2017) story about the history of telephony in the USA; when telephones were not a privilege of all classes in the society, people used public pay phones to reach one another. Children had the job of taking the phone calls and run to fetch the person it was meant for. The connection between caller and recipient was ensured through social, ad hoc connections rather than permanent poles and wires. Related to infrastructures, Geirbo contends that designers can utilize existing practices and habits in communities. Transferred to conversational

technology used in a pedagogic context, we ask if one can utilize already existing pedagogic practices in the design of CAs.

We observed how the educator became an essential part of the child-CA interaction by using well-known techniques to motivate the children, as well as to support repair and prevention of breakdowns. Ninio and Snow (1996) commented on several skills that are important to learn in order to function as an efficient conversational partner. It takes time to acquire these skills, and meanwhile children need help from adults to manage the complex cooperative requirements in conversations. When supported by adults, children are often able to function well in conversations. Adults are normally willing to enter complex repair sequences in conversations with children to enable them to both understand and to be understood, but also to progress the conversation if it has stalled or revert to the original topic if associations pivoted the direction of the communication. In the observed child-CA interaction, we find that the adults conducted repair work of this character. As the technology is not equipped to take on the task of such complex repair, we want to consider the possibility of including the educator as an integral part of the CA, functioning as a scaffold in the interaction by compensating for the CAs insufficient capabilities. We envision this as an approach to designing *for* the limitations within the technology, as the limitations are taken into account and remedied by another actant in the communication – the educator.

Exploration of how the CA and educator can cooperate is something we consider an interesting area of research. One approach could be for the educator to take on the role as an "assistant" or a friend of the CA, having them interact as a part of the child-CA interaction, e.g. enabling the educator to distribute turns as a part of the interaction (CA: "choose a child who can answer this question" – educator: "Lucy!"). We also find that exploring how we can integrate the facilitation and repair work currently done by the educator into the design of the CA itself is an interesting area of investigation. For instance, we could have the CA distribute turns by selecting respondents by name. We suggest that it is important to include the whole group in the conversation in order to initiate teamwork and engage all the participants' attention.

6.2.5 Accuracy versus progress: how the conversational objective affects repair

Because we use our CA as a pedagogic tool with an objective of enabling an effective conversation, the premise of the conversation is different from task-based CAs because the children do not have an overarching pragmatic motivation in use. Rather, the educator sets the agenda. We question in what ways this influences how we approach repair work during interaction. Hence, we discuss whether control can function as an efficient repair strategy in situations where progress is more important than accuracy.

People usually organize natural conversations ad hoc through continuous collaboration between the participants. However, Suchman (2007, p. 105) contends that certain settings comprise prescriptions about the practical direction and purposes of the interaction as well as the organization of turns at talk. To provide an example, when physicians conduct conversations with their patients, the organization of conversation usually comprise of question-answer adjacency pairs; the physician asks questions of the patient and thus asserts control over the agenda. We argue the communication between the children and the CA resemble these specialized forms of interaction because the CA pose strict rules on the turn-taking in the conversation, and furthermore sets the agenda as to what topics it is capable of engaging in. Hence, who talks and what is talked about is not decided through collaboratively constructing the course of the interaction, but through the children providing the correct responses in accordance with the output from the CA.

Designers should take *who* decides the objective of the conversation and *what* that objective is into consideration when deciding upon appropriate communicative repair strategies. If robustness indeed pertains to the level of support provided to the user to reach their goals, we find it important to ask what the children's goals are in interaction with a pedagogic CA. The context, in this case, is a group setting comprising the CA, the educators, and two or more children. The educator's intention of use is to engage children in a pedagogic conversation having them learn or reflect upon something through debating a topic. The children neither have a predefined understanding of the educator's purpose of use, nor a well-defined intention of their own regarding the communication with the CA. We suggest that the abovementioned argument affects what repair strategies are appropriate within the child-CA interaction because the situation does not require a high level of precision or accuracy regarding the interpretation of the input from the

participants. Conversely, in most task-based CAs, specific and accurate information is essential in order to accomplish the task correctly. However, the primary goal in this situation is to reach completion so that the CA can instigate reflection and conversation on the educators chosen topic thus making progress more important than the CA interpreting the children's responses exactly right. Concerning the children, we observed them wanting to get to the end of the story, which might be preferable to have complete control during the entirety of the interaction. Instead of incessantly prompting repeats during breakdowns in the story-activity, the CA might simply assume it has heard the children make a choice and progress to the next section; "I didn't quite hear you, but let's tell Elisa to go to the city!". Similarly, during the quiz, providing the correct answer does not require the CA to acknowledge the children's responses explicitly. This is not to say that the CA should ignore the children at all times, as this would probably become frustrating in the end. However, as a sparingly applied repair strategy, it might provide a more graceful approach to repair than repeatedly prompting for repetition of the same trigger word.

This, however, poses the question of whether we can call "ignoring" the user's response repair because in these situations, mutual understanding has failed to manifest. The CA does not understand the user, but still proceeds with the task. We find this to be repair in the sense that it progresses interaction to reach the objective of the task, which is to learn about a topic through both child-CA interaction and the succeeding pedagogic conversation.

6.3 Recommendations for design: towards a concept of robustness in design of CAs for children

We suggest that in the endeavor to design a robust CA for children, designers need to consider the following key implications: (1) support the user in understanding communicative constraints, (2) support the collaborative nature of conversation, and (3) understand what constitutes an effective conversation.

6.3.1 Support the user in understanding communicative constraints

Hayes and Reddy (1983) assert that in order for interaction to be graceful, users must be able to mend and repair communication with CAs in the same way that they would if their conversational partner was human. However, even though it has been many years in the making, the technology is still not able to partake in conversations with the same ease or effortlessness as people. Moreover, Luger and Sellen (2016) point out that people tend to overestimate the conversational competencies and capabilities of machines that are able to talk, which is a challenge in respect to design of CAs. Currently, we cannot completely circumvent or remove the asymmetries in communicative resources available to people and machines respectively. Consequently, we suggest that the concept of robustness in design of CAs for children is largely a question of accentuating the tacit and subtle variances that make conversations with CAs different from human-to-human conversations.

We find that one way to accomplish this is to provide the user with sufficient communicative resources during the interaction. We suggest that in our prototype, visual support might aid the children in timing their responses correctly, as well as distribute turns more efficiently. We envision an embodied agent that provide paralinguistic cues, e.g. through opening or closing their eyes to indicate data-processing and providing a response (closed: I'm thinking!, open: I'm talking!) or through ears that alternate between red and green light depending on whether the microphone is active or not (red = I can't hear you, green: you can talk now). Furthermore, eyes that fixate on one child at a time might indicate from whom the CA expects a response. Other types of CA might have constrictions that manifest in other ways during the interaction, but the point remains that designers should endeavor to identify and accentuate these differences so that users are not confounded as to what is correct behavior in use. Hence, we should strive to make the CA an asset in the collaborative act of communication, which brings us to our next recommendation.

6.3.2 Support the collaborative nature of conversation

We suggest that the concept of robustness entails designing to supporting the cooperative effort of sustaining the mutual intelligibility of dialogue. Paul Grice (1982) contended that conversations are a cooperative endeavor. Consequently, we choose to consider

communicative repair work in human-CA interaction a collaboration between the user and the CA. However, our focus on breakdowns implied that the complexity of collaborative repair proved challenging to the participants in many respects, mainly due to limitations in the technology and inadequate communicative competencies regarding both human and non-human actants. Supporting the collaborative nature will primarily entail designing for *specificity* as discussed in relation to *recoverability*. Designers can achieve this by providing sufficient resources such as unprovoked capacity indicators ("do you need more time?"), partial repeats ("what was the alternatives?", "Elisa can go to the left or to the right"), specific error-messages ("I am currently not able to help you with this"), and non-ambiguous output. Moreover, we argue that supporting collaboration entails an understanding of how the user's *competencies* affect communication.

We have learned that in pursuit of an effective conversation the participants must follow four conversational maxims, the maxims of *quantity*, *quality*, *relevance* and *manner* (Grice, 1982). Hall (2018) suggests that these maxims have implications for dialogue design and conversational proficiency in the agent. Our observations propose that in interactions with CAs, these requirements also apply to the users. They have to assess the right amount of information to provide in each instance, they have to be specific and relevant to the context of the conversation, and they are required to be clear, unambiguous and efficient. However, young children are not yet fully educated in the arts of conversation and language. Moore (2018) contends that agents should inhabit basic conversational competencies, which he defines as the agent having the ability to respond appropriately to common actions in a conversation – "can the agent *do* conversation?" (Moore, 2018, p. 187). Seeing as young children do not necessarily coordinate or navigate a conversation with the proficiency required by the technology, we think it is sensible to turn this question around and ask whether *the user* can "do conversation", and if not; what implications might this have concerning spoken interaction?

Our research showed that the participants seldom provided idealized utterances, indicating that situated language is often messy. Some of the participants had speech impediments that caused trouble in the interactions with the CA (pronunciation of "r" and "s"). Other times, we saw that it was difficult to discern or remember the correct trigger words, indicating the need for cognitive support through repetition or other reminders. Hence, when designing to support the capabilities of the users, we argue that we should carefully consider the characteristics of members of this group. Must we use trigger

words containing awkward letters, or are there alternatives that are less problematic? How do we reveal required functionality efficiently so the user knows how the CA can help them when breakdowns occur? Is our dialogue design ambiguous? Supporting the collaborative effort in communication and in recovering from breakdown is essential in order to design a robust CA and to ensure that we achieve the objective of the conversation – creating an effective conversation.

6.3.3 Understand what constitutes an effective conversation

People have different motivations for engaging in conversation. Dubberly and Pangaro (2009) argue that the *effective* conversation changes ”something” in a way that has lasting value to the participants. Thus, an effective conversation with a CA might be one in which you are able to book a plane ticket efficiently and with certainty that everything went as planned, or it might be one where you have engaged in strange interactions with chatbots that made you laugh. In our case, we assess the effectiveness to comprise the degree to which the educators have achieved their pedagogic objective – did the conversation influence the children’s understanding of themselves in connection to the nonhuman natural environment. Consequently, we suggest that robustness in design of CAs entails a clear understanding of what constitutes an effective conversation, meaning that we should carefully consider who sets the agenda, what that agenda is, and how this influences aspects such as control, accuracy and how breakdowns should be handled. We believe that this, in turn, affects what strategies for repair that are applicable in the specific context of use. If you want to book a plane ticket, the CA cannot simply decide that it think it interpreted the information correctly and proceed with the task. In our instance, however, the situation might allow for more assertiveness and control.

7 Critical reflections

In this chapter, we begin by reflecting upon the tools and methods used in this project. We will then address how we worked towards including the children as informants with an overarching objective of enabling them to participate and learn. We conclude the chapter with a reflection regarding introducing conversational technology in a kindergarten setting.

7.1 Evaluation of methods and tools

7.1.1 Use of the technology in prototyping

In hindsight, we find that in addition to our explicit focus on preventing and mediating breakdowns, the design outcomes manifested in the prototype was largely predisposed by the performativity in the tool. Orlikowski (2005) find it important to distinguish between *human agency* and *material performativity* in technology studies. She highlights their importance and that they are related, but also not equal. Related to our study, we find that our agency in the design process manifested in the result, but the material performativity also played its part. Orlikowski talks to those who believe in the social construction of reality, stating that one should recognize the role of technological performativity, as this may help understand how our reality is materially constructed (Orlikowski, 2005). Our findings from the case study indicated that the children preferred an explorative approach to interaction based on curiosity and playfulness, but the final prototype did not account for this behavior. In hindsight, we suggest that the material performativity influenced the design result – a prototype not accounting for the openness and flexibility that interaction with children often necessitates.

In addition to this, we question to what degree we were able to evaluate the actual dialogue and user experience. During the evaluations, we continuously experienced that the children had trouble completing the activities presented by the CA. This was largely due to limitations and constraints pertaining to the technology, such as the Natural Language Processing and strict rules for receptivity of input. Our goal was to make the prototype as robust as possible, and the continuous breakdowns caused us to remove interactional elements rather than explore different structures and types of dialog. In

retrospect, we find that the troubles caused by the technology made the breakdowns that originated from flawed design less visible to us. This may have caused inattentiveness to potential improvements as well as alternative ways to approach the construction of conversations between the CA and the children and ways to improve user experience.

One solution to this problem could be to apply a *Wizard-of-Oz* approach; a simulation of functionality which allows for testing of applications that does not yet exist (Lazar et al., 2010). A person acts out the systems intended functions, and the participants are usually not aware that they are not interacting with the actual system. After the *role-playing* activity, we could have continued to make low-fidelity prototypes of flow-based scripts and acted these out with the children. By placing the actor in a separate room, the children might be convinced that they were talking to an actual CA. In the first iteration of the design process, we used enacting as a technique to create example dialogue. Despite of our efforts to explain the purpose and plan for the session, many of the children did not immediately understand that we were the ones playing the part of the robot. We never tried to conceal this fact and placed ourselves behind the curtain without any attempt at deception. It is difficult to say exactly why this happened, but we did observe the children being very preoccupied with the robot and not necessarily noticing us slipping under the curtain. Moreover, we argue that this can attest to the value of role-playing as an approach to envisioning conversations. Children have a unique ability to immerse themselves in a situation and imagine something as being real even when it is not. Thus, we believe that enacting scenarios with or without props can provide good indicators of how children would actually talk to a CA (if it could understand them as well as humans can, that is).

However, if the children become aware of the fact that the "CA" is not real, we find it probable that they would change their behavior. Hence, the data would not be truly representative of child-CA interaction. In our case, when the children exposed us as the ones playing the CA, the focus came to be on us as actors, not the task at hand – although only for a little while. Furthermore, the technique is precarious because of the probability of errors made by the actor. People do make mistakes, which could lead to incidents that would not be representative of child-CA interaction. Additionally, we wanted to understand how to design within the limitations and constraints that the technology sets. Not using the technology would diminish this and would have made the findings for our

purpose of research of little value. Hence, we found it problematic to refrain from using the actual technology.

However, we could apply a more balanced approach; test the design through an evaluation of the prototype and additionally implement a Wizard-of-Oz evaluation to see if the same issues occurred in those interactions. We suggest that simulating the technology and eliminating the technological causes of breakdown could be favorable in certain instances. If the objective was to evaluate elements that either the technology, the tool, or the abilities of the designer would prevent from being implemented in a prototype (e.g. playfulness and explorative interaction), or if the aim was to simply evaluate the wording of the output, then a Wizard-of-Oz approach might be favorable. By not using the technology, we could have examined solely whether the output was understandable and helpful during the interaction by observing the children's reactions. Although, in such cases we advise that it is important to be aware that children might behave differently if they understand that the "CA" they are talking to is an actual person.

7.1.2 Use of the prototype in research

We acknowledge that developing a prototype whilst doing research might have affected the research itself. We explored weaknesses in interaction with a prototype that was unfinished and far from perfect. Therefore, breakdowns in interaction can be a result of an incomplete and imperfect prototype and not derived from inherent characteristics of child-CA interaction. However, we did observe some of the same breakdowns in the children's interaction with Siri, which is arguably one of the most widely used digital assistants in the world. Hence, it has matured and gone through extensive training, although not with young children as users.

An argument that strengthens our belief that the prototype was a valuable tool in our research is that we conducted several pilot-tests with fellow students in which we observed few or no breakdowns. This suggests that our findings are connected to the context and children as users, and not the result of a deficient prototype. However, we do admit that it would be ideal to use a well-trained and thoroughly tested prototype for this purpose because it would remedy breakdowns that occurs due to lack of training or weaknesses in dialogue design. This was not feasible in this project, due to time constraints and limited knowledge concerning the prototyping tool and programming

language. However, there are additional examples of studies in which faulty technology leads to interesting findings, e.g. an error leading to inability to detect sound from young participants led to a focus on repair strategies in Cheng et al.'s (2018) study. We argue that developing the prototype whilst conducting the research did not negate the value of the findings in this project.

7.2 Children as participants in the design process

In this section, we reflect upon the involvement of the children in the process, our strategies for enabling participation, and debate what role the children had during the process.

7.2.1 Mindfulness of participation and process

Throughout the process, we wanted the children to be mindful of the interconnected activities of the design process as well as to learn about technology development and about the technology itself. We used several strategies to visualize and make clear the process as a whole. We also attempted to emphasize that we considered their input to be of value during the development of the prototype.

One of these strategies was *the Inventor's Book*, which we used as a tool for facilitating engagement and awareness of the design process. As previously stated, this tool did seem to engage and motivate the children, and it appeared to create a sense of understanding of the continuation in the design process. It enabled us to talk about the design process as constituting of several activities, and that the purpose of these activities differed. We do, however question to what extent they learned about technology development by participating in this project. We often failed to engage them in conversations about the technology and about how we actually developed the prototype. We considered these conversations rather complex and we found them to demand thorough preparation. Conducting the evaluations was in itself draining for the children. Thus, we were already pushing the limits as of what their endurance and concentration span could handle, and we had to keep the sessions as short as possible. Therefore, we prioritized the actual evaluation and the interaction with the prototype over methodical and technological discussions.

As a way of accentuating the children's contribution to the process, we made sure to thank them after each session, stressing that they were helping us to make the "robot" better. We concluded the design process by giving all the children their own "inventor-badge" (the same badge that was on the physical representation of the CA), as a way of symbolizing their status as "robot-inventors". We told them that they would get this badge after the first evaluation, which they were excited about. During the following activities, they continued to mention, "when we have invented a robot, we will get an inventor-badge". However, we are uncertain of how they perceived their actual contribution as it was obviously not entirely clear to them exactly how their participation aided the development of the prototype.

7.2.2 Supporting the children in participation

As stated in Chapter 3, we wanted to involve the children in the design process as informants and enable and support participation in a valuable way by giving them a fun and hopefully educational experience. In addition to this, we had to make sure that the evaluations provided data informing our research. This required thorough planning and was both time consuming and demanding. Furthermore, the design material we worked with was intangible, abstract, and unfamiliar to us. In all activities and techniques that we implemented during the process, we tried to understand how to give specific and definite form to the material and create methods for feedback that would support the children as competent actants. We attempted to make them understand the purpose of the activity, encourage them in expressing their opinions, as well as make clear what they were expressing opinions about. In this section, we will elaborate on these techniques and discuss their value in the process. We also comment on other interesting insights.

What am I supposed to do?

During the focus groups in the preliminary case study, we had our first experience of how things never pan out as planned when working with children. How the focus groups unfolded, highlighted our incorrect assumptions about children as participants, and especially the assumption that children are all creative and imaginative beings that would have little trouble improvising interaction with a CA. We failed to provide tasks before we asked them to interact with Siri. In retrospect, we acknowledge the irrationality of expecting it to be effortless to come up with questions and talking points without any

support. Additionally, we did not account for Siri being a passive participant in the conversation, as “she” does not initiate interaction with the user. Thus, the children became the only party driving the conversation forward, and the absence of an explicit intent or objective in the dialogue was obviously debilitating. Thus, we learned not to assign the children tasks without providing context and support during the activity.

In the succeeding robot-workshop, a happy accident occurred. After concluding one session, we forgot to disassemble the 2-dimentional robots the participants had made before bringing in the next group. The children saw the robots and became curious and eager to make their own. This was in complete contrast to the former group who did not know what to expect from the activity and thus approached it with uncertainty and hesitation. In the preceding sessions, we intentionally hinted to the nature of the proceeding activity by leaving some robots that the participants could look at when they entered the room. This had us reflecting upon the importance of predictability when working with children in design processes. When conducting activities in the kindergarten, the adults often provide an example beforehand because it contributes to awareness of what is to come, as well as inspiration. We found it important to strike a good balance between preparing the children and to avoid overtly influencing their choices during an activity. We should give them an idea of what is to come, but still leave enough room for them to be creative.

The difficulties in providing reasons for their choices

Furthermore, we quickly asserted that the children found it difficult to provide reasons for their choices. In the robot-workshop, we initially wanted to get feedback from the children by having them choose their favorite robot and provide a reason for their choice. However, when we asked questions like “why did you choose that one?” we were continuously met with “I don’t know” or “just because”. Consequently, we decided to rely mainly on our observations of the activities and products to try to understand their likes and dislikes. In retrospect, we reflected upon the fact that participants may not always be mindful of the reasons behind their choices. Sometimes, what you like or dislike is based on an emotional response and not a conscious thought process. Thus, being unable to rationalize such a reaction should be acceptable. Hence, we reconsidered the importance of having the children provide an explanation for their choices.

Copying as consensus?

As stated, during the robot-workshop, we had troubles getting the children to express their opinions regarding choices of elements when making robots. However, we did observe that the robots made by children within the same group were alike in many ways. The same elements where incorporated by several children, and during the workshop it was evident that they were copying each other (e.g. using big eyes, making a cape, make a tall robot, give it clothes, using the big geometric shapes). This made us consider whether, in the process of creation, the act of copying each other's work is a form of consensus, as it reveals what elements the participants all agree that they like. This can also signify the value of observing actions made over asking them to reason.

Support in expressing opinions

A large part of planning the evaluations pertained to understand what methods would be suitable when using an intangible design material in the activities with the children. How should we support the children in expressing their opinions about dialogue – which is highly intangible. As stated, the children found it difficult to express reasons for their decisions, hence we relied on our observations to understand the interaction. However, we still sought to gain access to the children's thoughts regarding their experiences, but it was not an aim to have the children provide reasons for their answer. We decided to try different approaches to support the children in conveying their opinions. In evaluation 2, we used a technique called "the monster game", which we had come across during our readings; two monsters have contrasting opinions and the children are to agree with one of them. We developed our tool based on this description, which consisted of illustrations[5] of 16 monsters, where two monsters were paired up, giving opposing statements regarding the prototype (e.g. "the robot always understood me" and "the robot never understood me") (Appendix O). We instructed the children to put a sticker on the monster they agreed with in each pairing. The intention was to give them concrete opinions to reflect upon as we considered it easier to consider whether they agreed with something rather that come up with feedback on their own.

[5] Illustrations collected from: https://wikiclipart.com/monster-clipart_10297/, https://wikiclipart.com/monster-clipart_10310/ and https://wikiclipart.com/monster-clipart_10273/

Figure 22 One of the statement pairs in The Monster Game

We observed that the children seemed engaged, and they payed attention to the statements given by each monster. However, when analyzing the data, we saw problems regarding the way we phrased the statements. We considered them too polarizing (very positive or very negative), thus the children found themselves disagreeing with both monsters. Some of the children wanted to put their sticker in the middle of the two monsters, which may indicate that they did not fully agree with either of them. We also consider the possibility of the children being polite, not wanting to put their sticker on a negative statement, and at the same time not lie and put it on the positive one. Additionally, some statements were phrased in a way that put the blame on the user ("*I* did not understand what the robot said" or "*I* thought it was difficult to talk to the robot") and the children might have perceived that putting a sticker on the negative statement reflected badly on them. We observed that the children clearly struggled in their interactions with the CA, but in the succeeding feedback session, they put their sticker next to statements like "the robot always understood me". Consequently, we question whether the data was of value as we were unsure about the intention and meaning behind their answers. We discarded the data from this activity, and solely relied on observations and transcripts when analyzing data in the succeeding iterations for this project.

In the field trial, we were not present to observe the children directly. However, we were still curious about their thoughts on using the CA. We also wanted to test different techniques when working with children. Therefore, we made a tool that we intended to elicit information about the participant's user experiences. We called it "the story of opinions" (Appendix P). We got inspiration from a game called "mad libs", in which the adjectives in a story is replaced with a blank space, and the participants fill in the blanks with random adjectives before someone reads the story aloud. We wrote the story as a

first-person description of use of the CA and left a blank space where the description of the children's thoughts should be noted. We provided some alternatives (positive and negative words describing the experience) but they were also able to provide their own suggestions. We told the educators to read the sentences aloud and discuss with the children what words they wanted to write down. In the concluding interview, the educators told us that the children were exhausted after the interaction with the CA and that they had difficulties concentrating during this activity. The educators were under the impression that the children provided the answers at random. Therefore, we considered the data from this activity to be of little value and leaned on the audio recordings to understand the experiences the children had during use.

The importance of being flexible

Our experiences suggest that working with children necessitates a high level of flexibility and an ability to improvise, as unexpected situations will often occur. We contend that a willingness to deviate from the plan is necessary in order to adapt to an unpredictable and heterogeneous user group. Children are not only different from adults, they are also different from each other – their personalities, strengths and challenges make each child unique. As the researchers at OsloMet told us, children often move in and out of situations because they need space and time to process information. We were mindful of this in the activities we conducted with them.

A rigid and strict plan was neither practical nor feasible during execution. Rather, we found that creating an outline and establishing a clear objective, allowed for more flexibility during implementation. We also found it important to have contingency plans prepared by discussing what could potentially go wrong, or what might be challenging for the children, in advance. Eventually, we learned who amongst the children were cautious and shy, who was easily distracted and playful, and who was focused and able to participate in lengthy discussions. Hence, we tried to adapt our approach in each activity to suit the particular participants. A continuous evaluation of the participant's level of attention, understanding, and motivation was essential in order to make quick assessments of the situation and take appropriate actions. Establishing a clear objective so that potentially improvised approaches or activities was more likely to produce data that could inform the development of the design was equally important.

In hindsight, we reflected upon whether we in some instances were *too flexible*, and that we perhaps deviated from our plan too quickly. After assessing the data from the field trial, we observed that the educators were firm with the children, encouraging them to finish the task even though they were clearly unmotivated. We chose not to push the children to this extent due to our newly established and somewhat fragile relationship. Moreover, we had to keep in mind that participation was voluntary. It was a continuous struggle to strike the correct balance between continuing with the planned activity in order to gather the data we needed and allowing the children to follow their impulses in instances where they seemed unengaged.

7.2.3 The evolution of the children's role in the process

We started this project with an idea that the children should be involved by partaking in the design process as informants (Druin, 2002). In hindsight, we saw that their role alternated between informant, user and tester. In the focus groups, we find their role to be twofold. Druins (2002) definition of *user* fits their role in this activity: they used an already released technology while being observed, and the researcher's goal was to understand the impact the technology has on the user to be able to create better technology on the basis on this understanding. In the focus groups, we also find that the children inhabited the role of informants. In this activity, they informed the design prior to development, as we observed and gained understanding of them as a user group, as well as their way of being in interaction with the technology. We intended to take their playfulness and curios nature into account during the development of the prototype. However, this did not happen, partly due to the performativity in the technology restricting the design, and partly because of our exclusive focus on creating a robust prototype, obscuring other valuable design choices.

Including the children as informants throughout the process proved difficult. Working with children as participants posed some challenges, and as mentioned, some of the techniques we used to support the children in voicing their opinion during the design process failed. We ended up founding our data collection on observations and transcripts of the interaction, rather than the children's opinions or thoughts. This led us to conclude that they participated primarily as testers in the evaluations in the design process.

To conclude, we argue that the children were a part of the whole process, informing the design before prototyping started, and they have been aware of the process and the activities leading to the last prototype. However, in the focus groups, they also inhabited the role as users, while in the last three evaluations their role resembled that of a tester, as we observed them using the technology and asked about their experiences afterward.

7.3 Introducing conversational technology in a kindergarten setting

In this section, we will reflect on potential issues regarding the introduction of conversational technology in kindergartens as well as why it may be in accordance with the national curriculum.

7.3.1 Designing for inclusion

We repeatedly observed that the children who mispronounced words withdrew from interacting with the CA because it was unable to understand them. These children realized that they caused the breakdowns by saying the word incorrectly and often blamed themselves ("I'm not able to do this. Can someone else try?"). The fact that the technology accentuates the children's difficulties with pronunciation is unfortunate. For some, it might lead to a devaluation of their abilities as well as bring attention to a sensitive issue. In hindsight, we reflected upon the possible consequences of introducing a technology that is incapable of interpreting some children's speech patterns in a group setting, and what the potential alternatives or workarounds are. One solution could be to avoid the use of trigger words that rely on the letters "r" and "s" as these are the most difficult sounds to pronounce. Nevertheless, we find that practitioners in the field of CA-design should design for the "weakest link" in the group, namely the children whose linguistic competencies are still in the early stages in the development. Related to this, we can also debate inclusion of children who are behind in the development regarding communicative competencies. We compare this to universal design and the objective of designing for inclusion by considering cognitive and physical limitations in the user group. This means that designers need an understanding regarding children's cognitive and linguistic development. We find that more research ought to be done pertaining to

how to support less competent conversationalists in CA interaction. We suggest some focus areas to meet this challenge in Section 8.2 regarding future work.

7.3.2 Digital practices and technology in the kindergarten

Digital practices are a central theme in the National curriculum for kindergartens in Norway, which states that the educators should actively involve the children when using digital tools and help them develop an ethical understanding of digital media early on. Hence, the educators are obligated to explore creative and inventive use of digital tools together with the children (Norwegian Directorate for Education and Training, 2017, pp. 44–45).

Conversational agents are a fairly new addition to mainstream technology and are quickly finding their way into people's homes. After Christmas break, several of the children expressed familiarity with the Google Home device, indicating that it was a popular Christmas gift that year. The educators who partook in the last evaluation stated that it is important for them to stay updated on the technology the children are using at home, and that the kindergarten has a responsibility for educating the children on media use and support critical reflections. This is in line with what the researchers at OsloMet told us regarding the importance of the adults' awareness of how the omnipresence of technology in children's lives today influences their childhood in significant ways. Adults in kindergartens can simply not ignore the fact that children today are users of technology and a part of their job is to support the children and enable them to use this technology wisely. We believe that introducing a CA in the kindergarten could enable conversation about the technology and how it ought to be used.

However, some argue that the introduction of technology undermines the values of interpersonal competence and human-to-human-relations in the kindergartens. How children use technology in the context of both school, kindergarten, and their spare time, is up for debate. Adult control and moderation are highlighted as important factors in developing healthy relations to technology. We argue that avoiding solitary use, imposing adult supervision and guidance, and using the technology with care and specific purpose, is important when introducing technology in kindergartens – including CAs.

8 Conclusion

In this thesis, we have presented a study with the objective of understanding *what the concept robustness entails in the design of conversational agents for preschool children*. The research process comprised of two distinct segments, where the first was a preliminary case study in which we sought to explore children's behavior in interactions with CAs as well as gain a deeper understanding of the kindergarten context. The second was an empirical study incorporating a user-centered design process in which we built and evaluated a CA-prototype with participants from two kindergartens. During the case study, we sought to immerse ourselves in the everyday workings of the kindergartens through participant observation. We also conducted an interview with an employee and an interview with two experts within the field of pedagogy. Thus, we were able to form an idea of how a CA might fit into the sociomaterial context of the kindergarten as a place of knowledge construction. Moreover, through a collection of focus groups conducted with 17 participants, we began to explore how children behave in interaction with conversational technology. The continuous breakdowns we observed during these sessions, made us consider the severe fragility in spoken communication between young children and machines. In the succeeding empirical study, we applied *broken world thinking* as a perspective in a design process to investigate whether breakdown and disintegration of communication could aid our understanding of CA-design for young children. The design process consisted of three phases:

1. *Engaging the children,* where we applied three strategies with the aim of supporting the participants as competent actants and informants in the duration of the process.

2. *Dialogue design,* in which we developed and evaluated a prototype in three iterations.

3. *A field trial,* where we deployed the prototype in Raspberry kindergarten so that the children and educators could use it and provide us with research data in the form of recordings as well as provide feedback in concluding interviews.

The continual involvement of children in this project has allowed us to point at specific challenges experienced by them during the interaction with the CA prototype.

Consequently, the findings and design recommendations we have derived in this thesis are rooted in comprehensive empirical data and are both relevant and favorable to those who wish to develop conversational technologies for similar user groups and contexts in the future.

8.1 Contributions

In this study, we have entered the field of Human-Computer Interaction (HCI) by exploring the intricacies and complications of designing dialogue between young children and conversational agents. Consequently, the contributions we make in this thesis can be useful for both researchers and practitioners who venture into similar undertakings. In this section, we attempt to clarify what we believe to be our main contributions and how they might be of value to others. We group our contributions under the following captions: robustness in child-CA interaction, broken world thinking as a design perspective, and methodological insights related to the design of dialog.

8.1.1 Robustness in child-CA interaction

The findings from this thesis have culminated in insights pertaining to challenges in child-CA interaction that can be beneficial to those who aim to understand what robustness entails in this context, and furthermore seek suggestions on how to design for robust conversations with children. Firstly, our research points to several trouble sources that are useful as they offer insights into what is specifically difficult when the technology relies on spoken interaction, most importantly precursors stemming from the children's playful and associative behavior, the group situation, and a lacking in conversational competencies in both actants. These findings along with central theoretical concepts borrowed from linguistics and conversation analysis informed our understanding of the concepts of *observability*, *recoverability*, *responsiveness* and *task conformance* (Dix et al., 2004) in CA-design. We argue that designing for observability in CAs entails both sending perceivable and understandable messages and supporting the children in understanding how to provide perceivable and intelligible responses in return. Recoverability concerns recovery from communicative breakdown and in conversations this entails designing for efficient communicative repair. When assessing how to design for responsiveness, we find that designers should consider how to support the children if

they need time to think before responding, and task conformance entails supporting an effective conversation by identifying the agenda and motivation behind the interaction. Other practitioners can learn from and apply this understanding and the suggestions we provide when designing CAs for children in the future.

Additionally, we contribute three design recommendations. Firstly, designers should *support the user in understanding communicative constraints*. We cannot completely avoid or remove the differences in communicative resources available to people and machines respectively. We suggest that designers should explore ways to mediate these asymmetries by accentuating the tacit and subtle discrepancies, which make conversations with CAs different from human-to-human conversations. Secondly, designers should endeavor to *support the cooperative nature of conversations*. We found that the cooperative nature of conversations demands that designers make sure the CAs provides adequate communicative resources, especially in situations that require repair because specificity and explicitness are vital when initiating repair sequences with young children. Furthermore, children rarely probe to reveal capabilities. Hence, the CA should reveal these at appropriate moments during the interaction. Moreover, human-CA interaction requires efficiency in communication not only from the CA, but from the users as well. We have pointed to the fact that children do not always adhere to established conversational conventions, thus when developing CAs designers should ask if the intended users can "do conversation" and strive to understand their challenges and consider how these might affect the interaction. Lastly, designers should attempt to *understand what constitutes an effective conversation*. The goals of a conversation are not necessarily pragmatic in nature. With regards to the pedagogic objective of the prototype that we developed, an effective conversation was one in which the children had learned something or reflected upon a topic during or after use. Thus, we found that we might support this objective more efficiently by designing explicitly for the group situation or by designing for the educator to be an integral actant in use. Furthermore, the motivations and objective for interacting with CAs affect what repair strategies are applicable. We suggest that designers should reflect upon the importance of accuracy in interaction, and thus consider whether asserting control can be an appropriate repair strategy in certain situations.

8.1.2 Broken world thinking as a design perspective

At the beginning of this thesis, we asked what we could learn by making breakdown the starting point for understanding the challenges in the interaction between children and CAs.

This study points to how *broken world thinking* as a perspective in a design process instigated a repair-centric focus and aided our understanding of the challenges inherent in child-CA interaction. Hence, this design process shows that by positioning breakdown as an inevitability of use and of communication, our understanding of how to conceptualize robustness in relation to the technology, context and target group gradually increased. However, we also contribute critical comments as we learned that an exclusive focus on how to remedy and prevent breakdown obscured other important design elements and influenced some decisions unfavorably. Consequently, we suggest a more balanced approach to those who aim to use this as a perspective in design processes themselves.

8.1.3 Methodological insights – designing with dialog

Our research approach provides suggestions and insights into how we can combine designing dialogue for CAs and working with children. We have offered our experiences with various methodical approaches such as enacting dialogue and techniques for supporting engagement and participation in a continuous but fragmentary design process. Others might use these approaches as we present them here or adapt them to similar situations in which they combine an intangible design material with a rather challenging target group. Additionally, we created an analytical framework that helped accentuate both trouble sources and repair strategies in interactions between people and CAs, and we suggest it is very much applicable by others when analyzing human-CA interaction. Moreover, the way we used descriptive concepts from the fields of conversation analysis and linguistics to understand phenomena such as communicative repair in this design and research process, can be a resource for others to learn more of how they can approach working with dialogue as a design material.

Lastly, the project illustrates how the performativity in the tool that was available to us (as well as our technological skills) influence how well we were able to explore different ways to design for the actual behaviors that the users displayed. Here we refer to how we were unable to explore more playful and open conversations with the children, even

though we had identified this as an essential motivating factor in their way of being. These experiences point to challenges that others might encounter, and that they should consider carefully in relation to the goals of their project. Moreover, we have offered suggestions to circumvent these tools by exploring other forms of conversations by applying methods such as Wizard-of-Oz.

8.2 Future work

More knowledge regarding child-CA interaction is needed in order to improve conversational technology for a user group with linguistic skills and technological understandings different from adults. The insights gained throughout this project has highlighted several opportunities for further exploration within the field of Human-Computer Interaction. Firstly, we suggest that future work should involve exploration of how to provide visual support for timing and turn-taking in child-CA interaction, as this study points to the importance of enabling an understanding of the subtle differences between human-to-human conversation and conversations with a machine. We have proposed some ways of doing this: opening and closing of eyes, ears that indicate the agent's ability to receive input or the possibility of looking at the child that should speak next. We believe that an evaluation of these or other strategies would provide interesting contributions to the field.

Secondly, future work should involve exploration of how to design CAs for use in a group setting. We have proposed that one possible approach is to design for the educator as an integral part of the interaction. As we learned that the educators preferred collective use of technology and our observations indicated that inactivity leads to inattentiveness during use, we find that more knowledge regarding use of CA in a group situation should be sought after.

Thirdly, others should investigate the possibility of the children being the ones setting the agenda for the conversation by designing CAs that are more playful and that supports exploration. We believe that designing for the inherent characteristics of the user group is essential to maintaining long-term use of the artifact when employed in a kindergarten setting. In addition to being robust, it ought to be fun and engaging.

Lastly, we find that exploring different methods for evaluating dialogue design and user experiences, with or without a functional prototype, would be a valuable contribution to the field of HCI. We used a functional prototype instead of the Wizard-of-Oz approach but contend that the Wizard-of-Oz approach could be an efficient method for the abovementioned purposes of evaluation.

Epilogue

We began this thesis by asking whether you have ever had a frustrating experience with chatbots or digital assistants, fully expecting that if you had used one you probably had. We believe that one day, spoken interaction with different types of technology will be just as normal as interacting with phones and tablets through touch screens is now. Although our imagining of the future is one in which interaction with conversational technologies feels effortless, the user-experiences these technologies presently offer suggest that the prospects of achieving this lie some time ahead. However, with each small step we take, and each new thing we learn, we believe that we are slowly approaching this now utopian vision.

Whilst remaining optimistic about the eventual benefits of well-functioning conversational technology, we also acknowledge that conversations are often deceivingly effortless – like everything else, it is easy if you know how to do it. Picture the often frustrating experience of speaking a language of which you are not proficient. It feels neither simple nor unproblematic to carry out a conversation if you are struggling to remember and pronounce words, if you feel uncertain of how to phrase questions, or if you find it difficult to interpret the unfamiliar conventions your conversational partner seems to follow. Now, imagine having this conversation over the phone, stripped of the gesturing, facial expressions and other paralinguistic signals that usually aid the communication. Not an easy task! At the closing of this thesis, we express a hope that the conversational technologies of the future incorporate the support mechanisms that adults apply in conversation with children. That way, we might avoid the children thinking that the only thing the CA knows how to do is to *not* understand them.

Bibliography

Baron. (2015). Shall We Talk? Conversing With Humans and Robots AU. *The Information Society*, *31*(3), 257–264. https://doi.org/10.1080/01972243.2015.1020211

Bratteteig, T., Bødker, K., Dittrich, Y., Mogensen, P. H., & Simonsen, J. (2013). Methods. Organising principles and general guidelines for Participatory Design projects. In J. Simonsen & T. Robertson (Eds.), *Routledge International Handbook of Participatory Design*. New York: Routledge International Handbooks.

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, *3*(2), 77–101.

Cheng, Y., Yen, K., Chen, Y., Chen, S., & Hiniker, A. (2018). Why doesn't it work?: voice-driven interfaces and young children's communication repair strategies. *Proceedings of the 17th ACM Conference on Interaction Design and Children - IDC '18*, 337–348. https://doi.org/10.1145/3202185.3202749

Clayton, S. D., & Opotow, S. (2003). *Identity and the Natural Environment: The Psychological Significance of Nature*. MIT Press.

Corsaro, W. A. (2000). Early childhood education, children's peer cultures, and the future of childhood. *European Early Childhood Education Research Journal*, *8*(2), 89–102. https://doi.org/10.1080/13502930085208591

Definition of workshop. (2019). Retrieved February 15, 2019, from Merriam-Webster website: https://www.merriam-webster.com/dictionary/workshop

Ditlefsen, H. (2018, January 10). Chatbot skal hjelpe ungdom som sliter. Retrieved February 19, 2019, from NRK website: https://www.nrk.no/sorlandet/chatbot-skal-hjelpe-ungdom-som-sliter-1.13856295

Dix, A., Finlay, J. E., Abowd, G. D., & Beale, R. (2004). *Human-Computer Interaction* (3rd ed.). Harlow, England: Pearson Education Limited.

Druga, S., Williams, R., Breazeal, C., & Resnick, M. (2017). "Hey Google is It OK if I Eat You?": Initial Explorations in Child-Agent Interaction. *Proceedings of the 2017 Conference on Interaction Design and Children*, 595–600. https://doi.org/10.1145/3078072.3084330

Druin, A. (2002). The role of children in the design of new technology. *Behaviour and Information Technology*, *21*(1), 1–25. https://doi.org/10.1080/01449290110108659

Dubberly, H., & Pangaro, P. (2009). What is conversation? How can we design for effective conversation? *Interactions Magazine*, 1–9.

Fischer, C. S. (1994). *America calling: A social history of the telephone to 1940*. University of California Press.

Fischer, G. (1994). Turning breakdowns into opportunities for creativity. *Knowledge-Based Systems*, *7*(4), 221–232. https://doi.org/10.1016/0950-7051(94)90033-7

Fogg, B. J. (2002). *Persuasive Technology: Using Computers to Change What We Think and Do* (1 edition). Amsterdam ; Boston: Morgan Kaufmann.

Gaver, W., Beaver, J., & Benford, S. (2003). Ambiguity as a resource for design. *Proceedings of the SIGCHI*. Retrieved from https://www.academia.edu/868863/Ambiguity_as_a_resource_for_design

Geirbo, H. C. (2017). *Crafting connections - practices of insfrastructuring. An ethnographic study of developing a village electricity grid in Bangladesh* (Ph.D. Thesis). University of Oslo, Oslo.

Gjems, L. (2009). *Å samtale seg til kunnskap. Sosiokulturelle teorier om barns læring om språk og gjennom språk.* (1st ed.). Bergen: Fagbokforlaget.

Golinkoff, R. M. (1986). 'I beg your pardon?': the preverbal negotiation of failed messages. *Journal of Child Language*, *13*(3), 455–476. https://doi.org/10.1017/S0305000900006826

Google. (2019a). Greetings - Conversational components - Conversation design. Retrieved January 11, 2019, from Google Design website: https://designguidelines.withgoogle.com/conversation/conversational-components/greetings.html#greetings-prompting

Google. (2019b). How does fulfillment work? | Dialogflow. Retrieved February 8, 2019, from https://dialogflow.com/docs/intro/fulfillment

Google. (2019c). Overview | Dialogflow. Retrieved February 8, 2019, from Dialogflow website: https://dialogflow.com/docs/intro

Google Home. (2019). Retrieved April 12, 2019, from Google Store website: https://store.google.com/product/google_home

Grice, H. P. (1982). Logic and Conversations. In *Studies in the Way of Words*. Retrieved from https://courses.media.mit.edu/2004spring/mas966/Grice%20Logic%20and%20Conversation.pdf

Gulz, A., Haake, M., Silvervarg, A., Sjödén, B., & Veletsianos, G. (2011). Building a Social Conversational Pedagogical Agent: Design Chalenges and Methodological approaches. In D. Perez-Marin & I. Pescual-Nieto (Eds.), *Conversational Agents and Natural Language Interaction: Techniques and Effective Practices*. IGI Global.

Hall, E. (2018). *Conversational Design* (1st ed.). A Book Apart.

Hayes, P. J., & Reddy, D. R. (1983). Steps toward graceful interaction in spoken and written man-machine communication. *International Journal of Man-Machine Studies*, *19*(3), 231–284. https://doi.org/10.1016/S0020-7373(83)80049-2

Hoey, E., & Kendrick, K. (2017). Conversation Analysis. In A. M. B. de Groot & P. Hagoort (Eds.), *Research methods in psycholinguistics: A practical guide*. Wiley Blackwell.

Hutchby, I., & Wooffitt, R. (1998). *Conversation Analysis. Principles, Practices and Applications* (5th ed.). Oxford: Blackwell Publishing Ltd.

Jackson, S. J. (2014). Rethinking Repair. *Media Technologies. Essays on Communication, Materiality and Society*, 221–239.

Jakobson, R. (1960). *Closing statement: Linguistics and poetics*. Cambridge, MA: MIT Press.

Johannessen, L. E. F., Rafoss, T. W., & Rasmussen, E. B. (2018). *Hvordan bruke teori? Nyttige verktøy i kvalitativ analyse*. Universitetsforlaget.

Jurafsky, D., & Martin, J. H. (2018). Dialog Systems and Chatbots. In *Speech and Language Processing* (Draft). Retrieved from https://web.stanford.edu/~jurafsky/slp3/24.pdf

Kanda, T., Hirano, T., Eaton, D., & Ishiguro, H. (2004). Interactive Robots As Social Partners and Peer Tutors for Children: A Field Trial. *Human Computer Interaction*, *19*(1), 61–84. https://doi.org/10.1207/s15327051hci1901&2_4

Kim, Y., Baylor, A. L., & Shen, E. (2007). Pedagogical agents as learning companions: the impact of agent emotion and gender. *Journal of Computer Assisted Learning*, *23*(3), 220–234. https://doi.org/10.1111/j.1365-2729.2006.00210.x

Lazar, J., Feng, J. H., & Hochheiser, H. (2010). *Research Methods in Human-Computer Interaction*. John Wiley & Sons Ltd.

Leigh Star, S., & Ruhleder, K. (1996). Steps Toward an Ecology of Infrastructure: Design and Access for Large Information Spaces. *Information Systems Research*, *7*, 111–134. https://doi.org/10.1287/isre.7.1.111

Lowdermilk, T. (2013). *User-Centered Design* (1st ed.). California, United States of America: O'Reilly Media Inc.

Löwgren, J., & Stolterman, E. (2004). *Thoughtful Interaction Design: A Design Perspective on Information Technology*. Cambridge, Massachusetts: The MIT Press.

Luger, E., & Sellen, A. (2016). "Like Having a Really Bad PA": The Gulf Between User Expectation and Experience of Conversational Agents. *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, 5286–5297. https://doi.org/10.1145/2858036.2858288

Moore, R. J. (2018). A Natural Conversation Framework for Conversational UX Design. In R. J. Moore, M. H. Szymanski, R. Arar, & G.-J. Ren (Eds.), *Studies in Conversational UX Design* (pp. 181–203). Springer Interantional Publishing.

Moore, R. J., Arar, R., Ren, G.-J., & Szymanski, M. H. (2017). Conversational UX Design. *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems*, 492–497. https://doi.org/10.1145/3027063.3027077

Myers, M. D. (1997). Qualitative Research in Information Systems. *Management Information Systems Quarterly*, *2*(21).

Newman, D. (2018). Top 10 Digital Transformation Trends For 2019. Retrieved March 29, 2019, from Forbes website: https://www.forbes.com/sites/danielnewman/2018/09/11/top-10-digital-transformation-trends-for-2019/

NILU. (2018). About NILU. Retrieved March 15, 2018, from https://www.nilu.no/OmNILU/tabid/67/language/en-GB/Default.aspx

Ninio, A., & Snow, C. E. (1996). *Pragmatic Development*. Westview Press Inc.

Norwegian Directorate for Education and Training. (2017). *Framework Plan for Kindergartens*. Retrieved from https://www.udir.no/globalassets/filer/barnehage/rammeplan/framework-plan-for-kindergartens2-2017.pdf

Orlikowski, W. J. (2005). Material Works: Exploring the Situated Entanglement of Technological Performativity and Human Agency. *Scandinavian Journal of Information Systems*, *17*(1).

Pagliari, C. (2007). Design and Evaluation in eHealth: Challenges and Implications for an Interdisciplinary Field. *Journal of Medical Internet Research*, *9*(2), e15. https://doi.org/10.2196/jmir.9.2.e15

Pierson, J., Jacobs, A., Dreessen, K., Van Den Broeck, I., Lievens, B., & Van Den Broeck, W. (2006). Walking the interface: uncovering practices through "proxy technology assessment." *EPIC*, 29–40. American Anthropological Association.

Preece, J., Rogers, Y., & Sharp, H. (2015). *Interaction Design: Beyond Human-Computer Interaction* (4th ed.). John Wiley & Sons Ltd.

Richards, Jack. C., & Schmidt, R. W. (1983). Conversational Analysis. In *Language and Communication*. New York: Longman Inc.

ROBUST | meaning in the Cambridge English Dictionary. (2019). In *Cambridge English Dictionary*. Retrieved from https://dictionary.cambridge.org/dictionary/english/robust

Robustness (computer science). (2018). In *Wikipedia*. Retrieved from https://en.wikipedia.org/w/index.php?title=Robustness_(computer_science)&oldid=865723730

Ryokai, K., Vaucelle, C., & Cassell, J. (2003). Virtual peers as partners in storytelling and literacy learning. *Journal of Computer Assisted Learning*, *19*(2), 195–208. https://doi.org/10.1046/j.0266-4909.2003.00020.x

Sacks, H., Schegloff, E., & Jefferson, G. (1974). *A Simplest Systematics for the Organization of Turn-Taking for Conversation* (Vol. 14). https://doi.org/10.17323/1728-192X-2015-1-142-202

Sanders, E. B.-N., & Stappers, P. J. (2008). Co-creation and the new landscapes of design. *CoDesign*, *4*(1), 5–18. https://doi.org/10.1080/15710880701875068

Schegloff, E. (2007). *Sequence organization in interaction: a primer in conversation*

analysis (Vol. 1). Cambridge: Cambridge University Press.

Schegloff, E., Jefferson, G., & Sacks, H. (1977). *The Preference for Self-Correction in the Organization of Repair in Conversation* (Vol. 53). https://doi.org/10.2307/413107

Schön, D. A. (1983). *The Reflective Practitioner: How Professionals Think in Action*. USA: Basic Books.

Sciuto, A., Saini, A., Forlizzi, J., & Hong, J. I. (2018). "Hey Alexa, What's Up?": A Mixed-Methods Studies of In-Home Conversational Agent Usage. *Proceedings of the 2018 Designing Interactive Systems Conference*, 857–868. https://doi.org/10.1145/3196709.3196772

Shedroff, N., & Noessel, C. (2012). *Make It So: Interaction Design Lessons from Science Fiction*. New York: Rosenfeld Media.

Sorin, R. (2014). Sharing Postcards about Where We Live: Early Childhood Environmental Understanding. *The International Journal of Early Childhood Learning*, *20*(3), 35–49. https://doi.org/10.18848/2327-7939/CGP/v20i03/48423

Sorin, R., & Galloway, G. (2006). *Construct of childhood: Constructs of self* (Vol. 31). https://doi.org/10.1017/S1035077200011081

Stake, R. E. (2005). Qualitative Case Studies. In N. Denzin & Y. Lincoln (Eds.), *The Sage Handbook of Qualitative Research* (pp. 443–466). Retrieved from https://www.researchgate.net/publication/232438903_Qualitative_Case_Studies

Stoltenberg, H. (2018, October 12). Jeg Begynner å Utvikle Følelser. Retrieved February 26, 2019, from www.dn.no website: https://www.dn.no/staticprojects/special/2018/10/11/0600/teknologi/eq-ai/

Suchman, L. (2007). *Human-Machine Reconfigurations: Plans and situated actions.* (2nd ed.). Cambridge University Press.

Sun, M., Leite, I., Lehman, J. F., & Li, B. (2017). *Collaborative storytelling between robot and child: A feasibility study*. 205–214. https://doi.org/10.1145/3078072.3079714

Susi, T., & Ziemke, T. (2005). On the Subject of Objects: Four Views on Object Perception and Tool Use. *TripleC: Communication, Capitalism & Critique. Open Access Journal for a Global Sustainable Information Society*, *3*(2), 6–19. https://doi.org/10.31269/triplec.v3i2.19

The Design Process: What is the Double Diamond? (2015, March 17). Retrieved March 27, 2019, from Design Council website: https://www.designcouncil.org.uk/news-opinion/design-process-what-double-diamond

Veletsianos, G., & Russell, G. S. (2014). Pedagogical Agents. In J. M. Spector, M. D. Merrill, J. Elen, & M. J. Bishop (Eds.), *Handbook of Research on Educational Communications and Technology* (pp. 759–769). https://doi.org/10.1007/978-1-4614-3185-5_61

Walsham, G. (1995). Interpretive case studies in IS research: nature and method.

European Journal of Information Systems, *4*(2), 74–81.

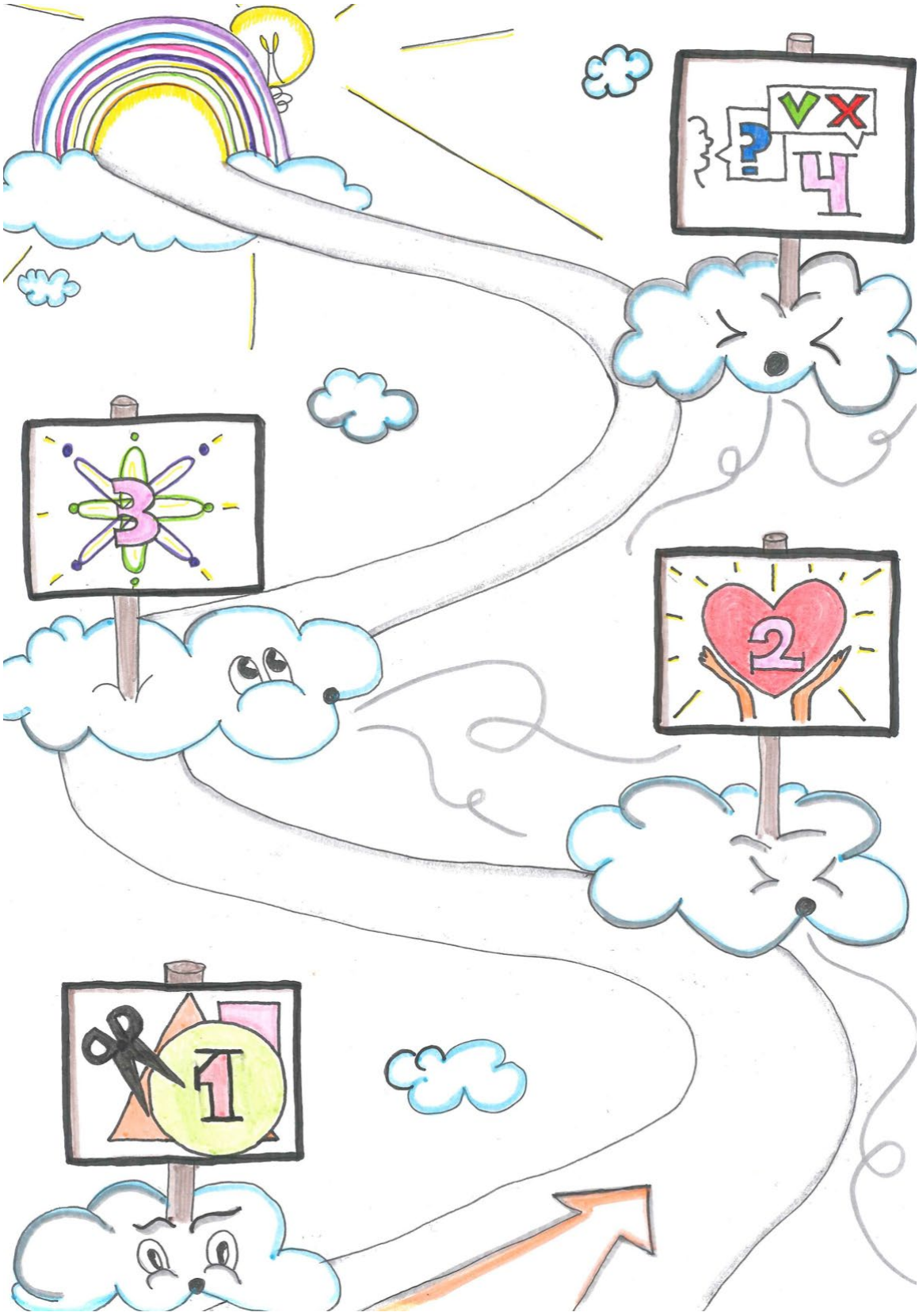
Watt, W. C. (1968). Habitability. *American Documentation*, *19*(3), 338–351. https://doi.org/10.1002/asi.5090190324

Weizenbaum, J. (1966). ELIZA—a Computer Program for the Study of Natural Language Communication Between Man and Machine. *Commun. ACM*, *9*(1), 36–45. https://doi.org/10.1145/365153.365168

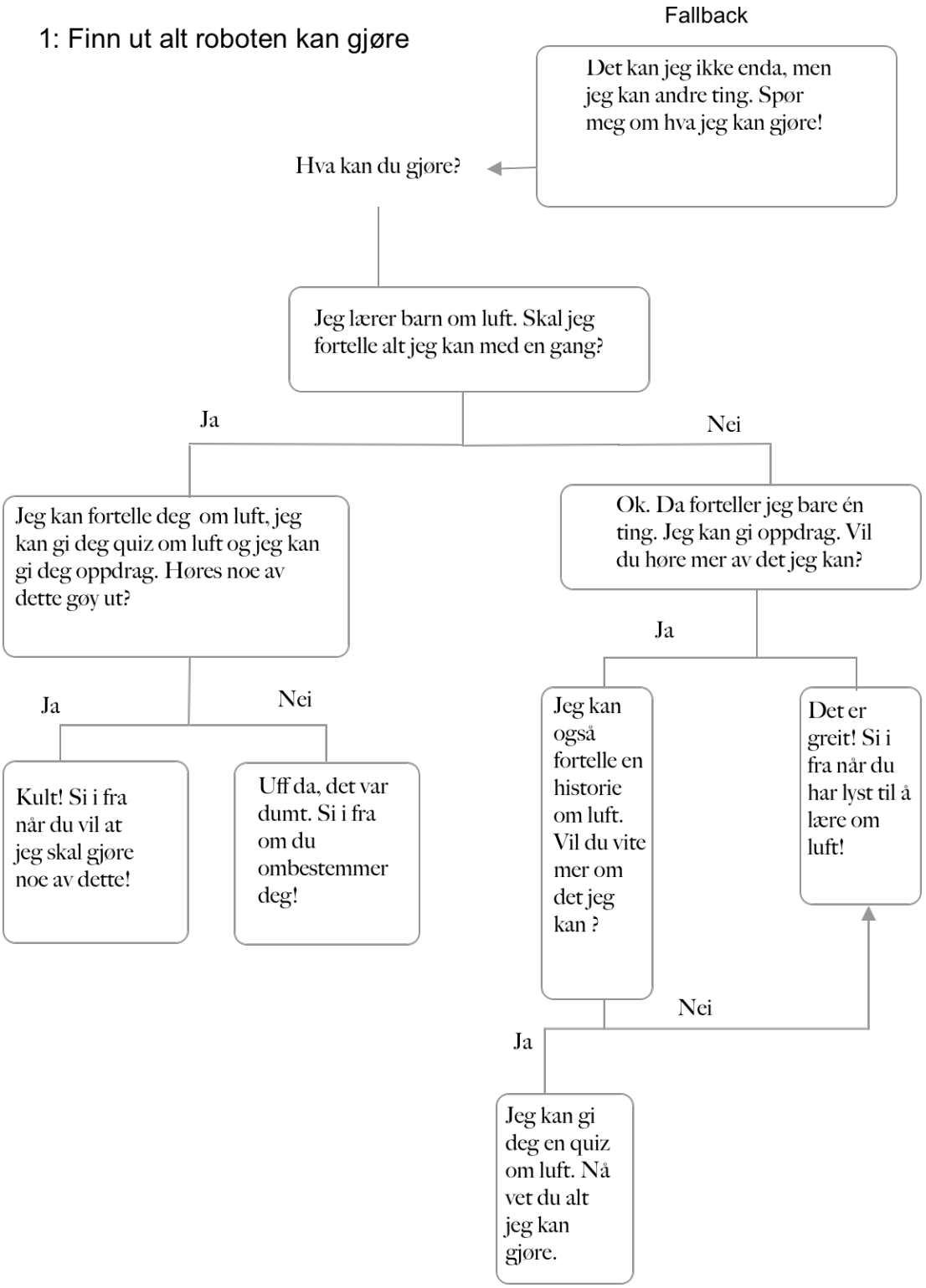
Appendices

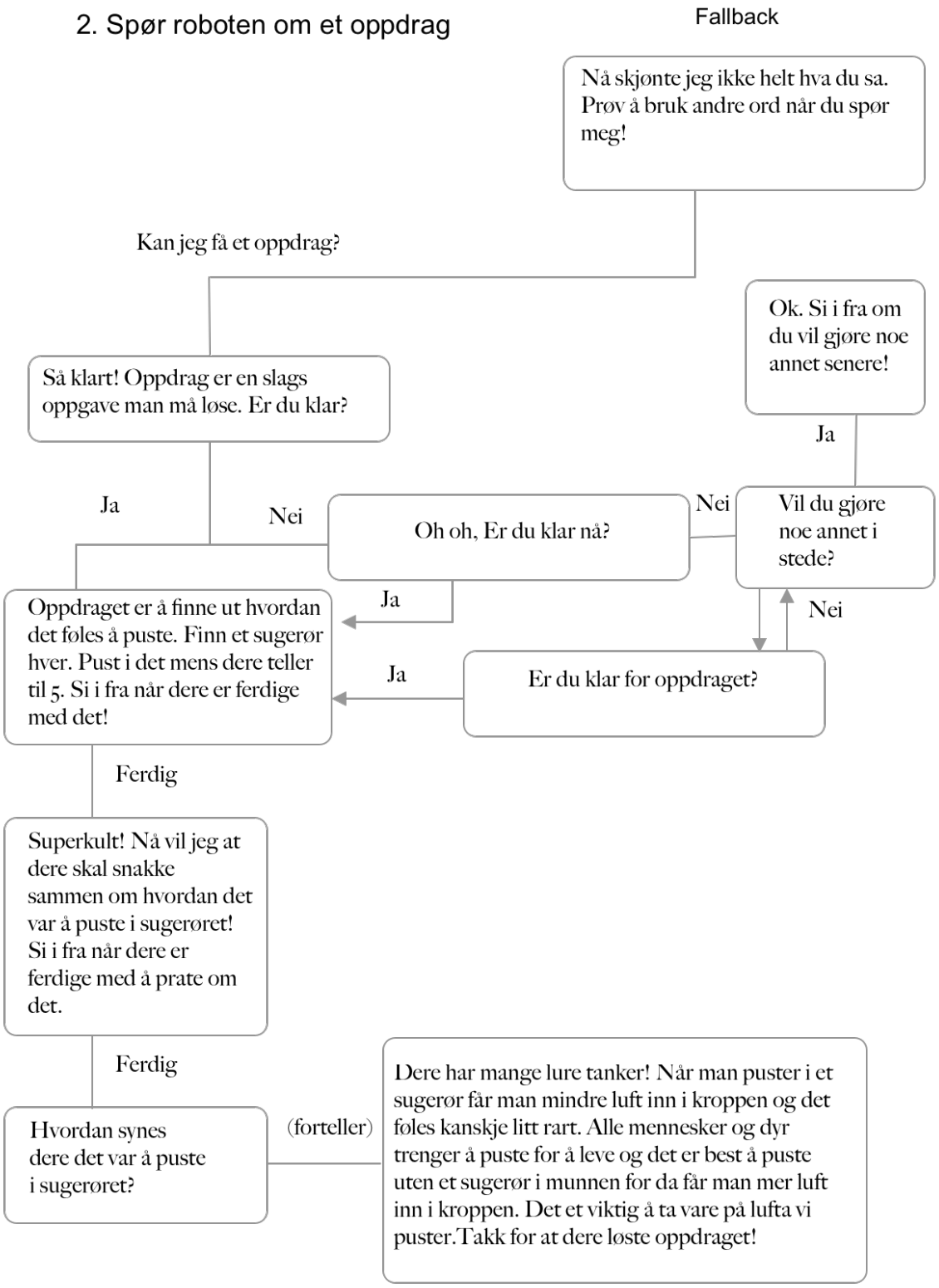
A The Inventor's Book front cover

B The Inventor's Book back cover

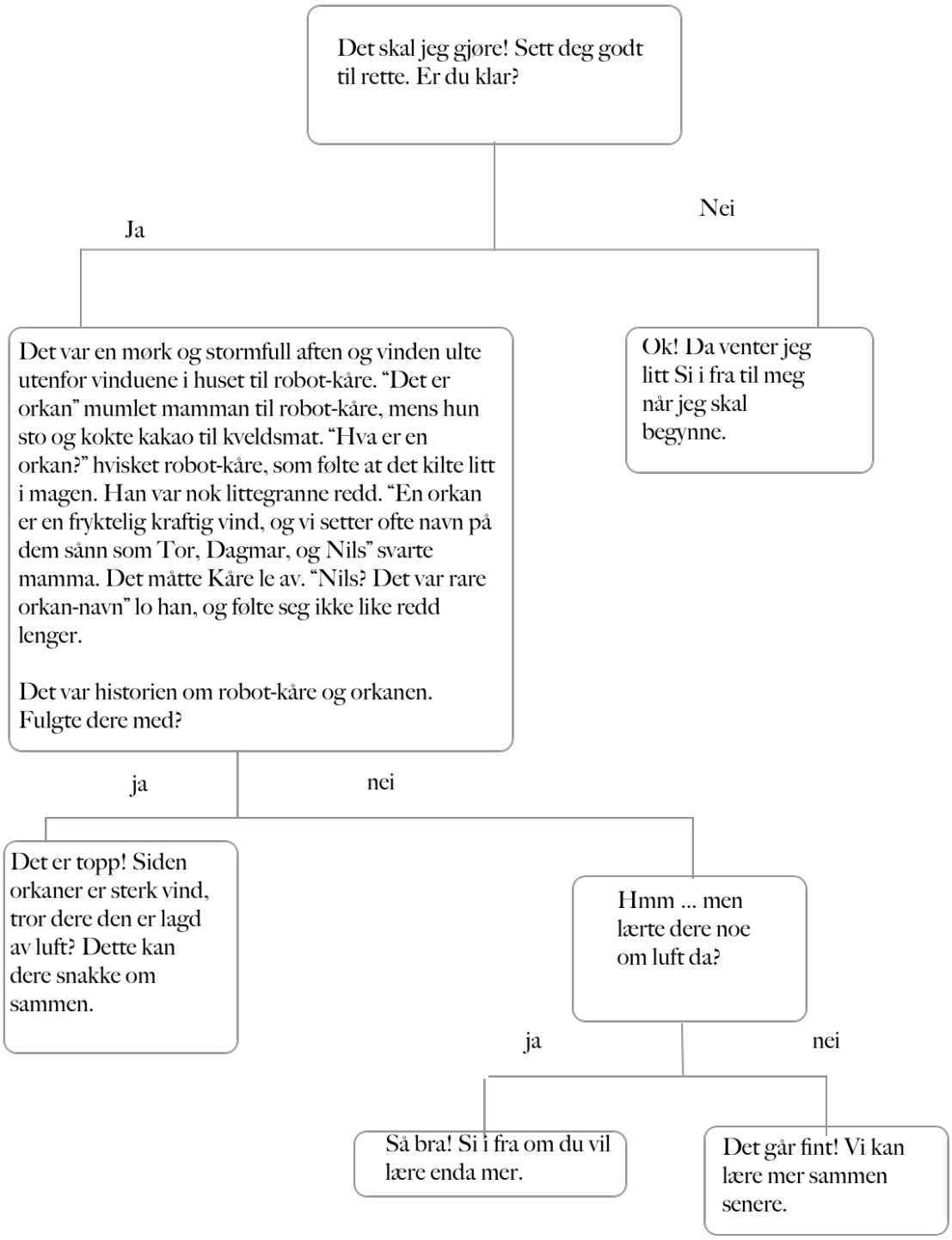


C Prototype iteration 1

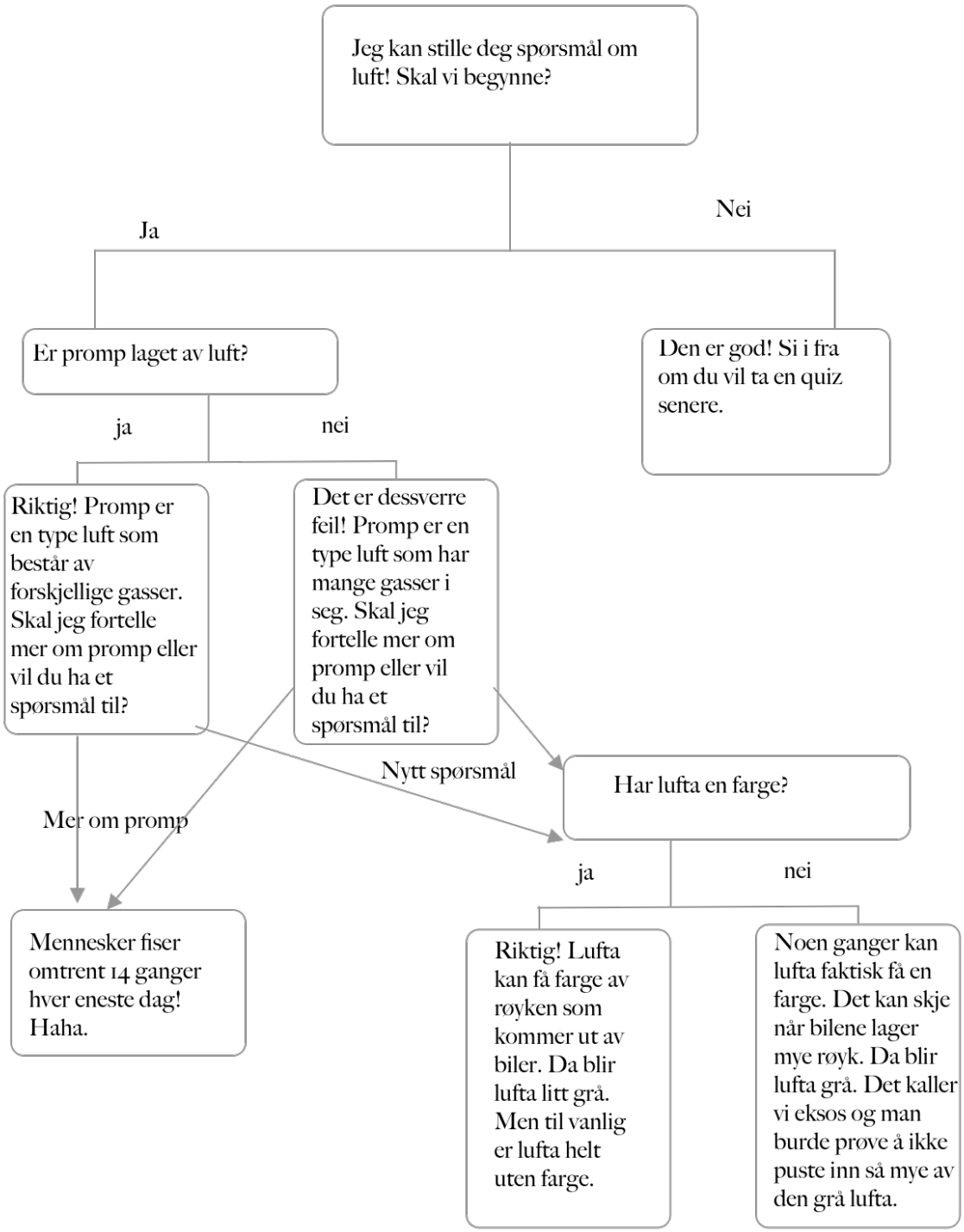



2. Spør roboten om et oppdrag
Fallback
Nå skjønte jeg ikke helt hva du sa. Prøv å bruk andre ord når du spør meg!
Kan jeg få et oppdrag?
Ok. Si i fra om du vil gjøre noe annet senere!
Så klart! Oppdrag er en slags oppgave man må løse. Er du klar?
Ja
Ja
Nei
Oh oh, Er du klar nå?
Nei
Vil du gjøre noe annet i stede?
Ja
Nei
Oppdraget er å finne ut hvordan det føles å puste. Finn et sugerør hver. Pust i det mens dere teller til 5. Si i fra når dere er ferdige med det!
Ja
Er du klar for oppdraget?
Ferdig
Superkult! Nå vil jeg at dere skal snakke sammen om hvordan det var å puste i sugerøret! Si i fra når dere er ferdige med å prate om det.
Ferdig
Hvordan synes dere det var å puste i sugerøret?
(forteller)
Dere har mange lure tanker! Når man puster i et sugerør får man mindre luft inn i kroppen og det føles kanskje litt rart. Alle mennesker og dyr trenger å puste for å leve og det er best å puste uten et sugerør i munnen for da får man mer luft inn i kroppen. Det et viktig å ta vare på lufta vi puster.Takk for at dere løste oppdraget!

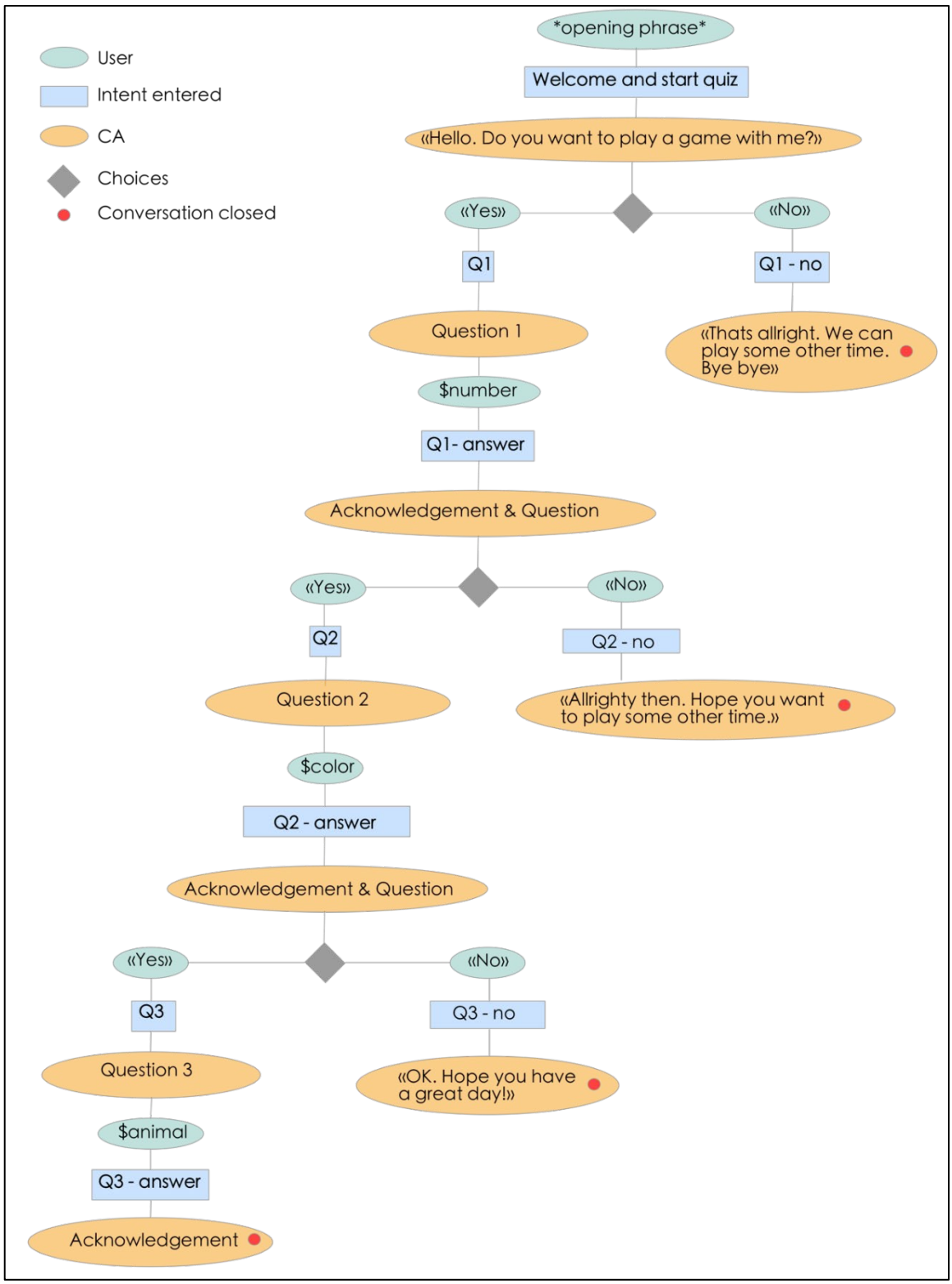
3. Få roboten til å fortelle en historie



4 . Få roboten til å spille et spørrespill



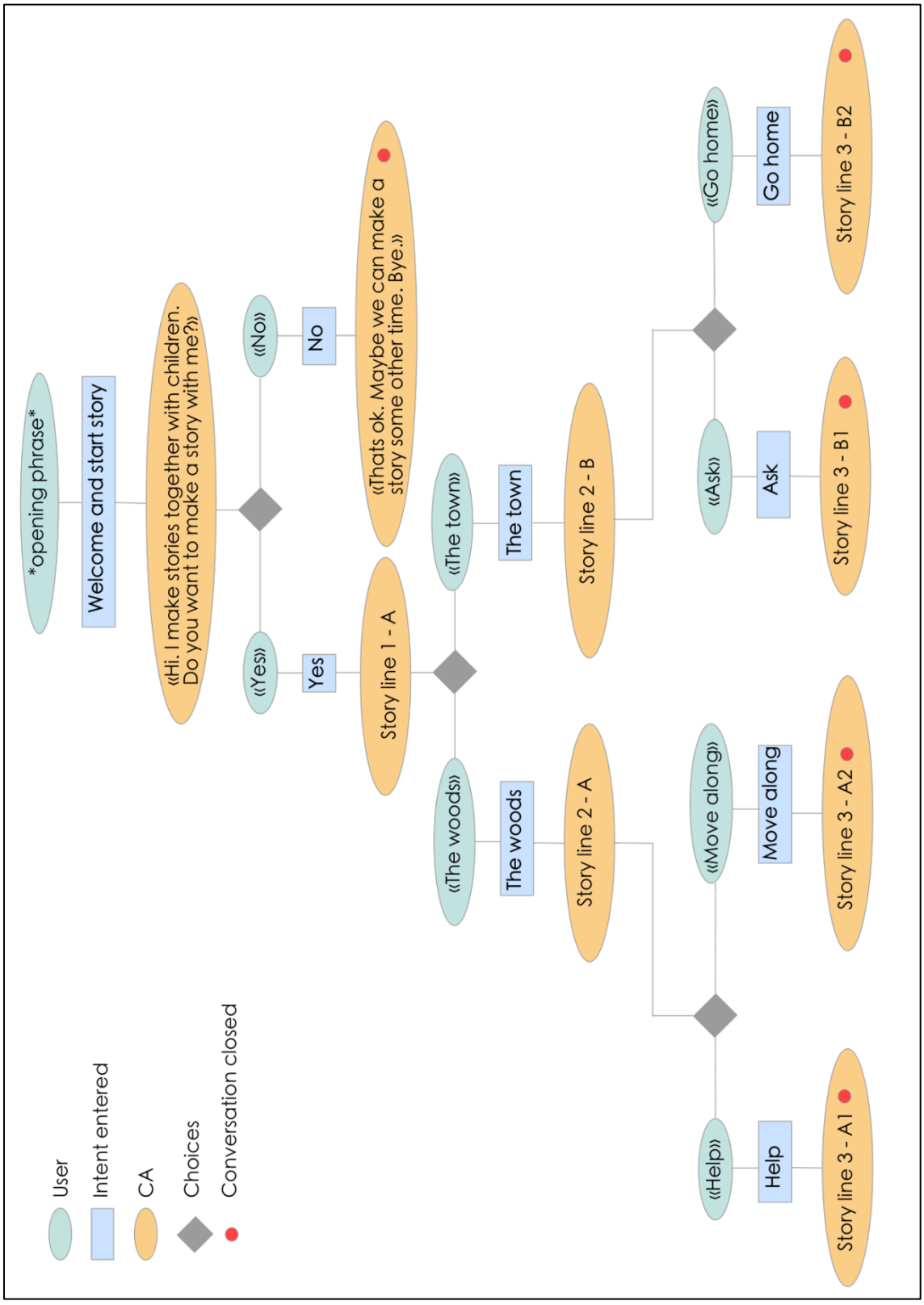
D Flow and content prototype v1: quiz



Who	Sequence	Alt. input A/ Output	Alt. input B/ Output	Intent
User	Greeting	Opening phrase		Welcome and start quiz
CA	Greeting	Hello.		
CA	Self-identification	I can give children games to play.		
CA	Question	Do you want to play a game with me?		
User	Answer	Yes	No	A: Q1 B: Q1 – no
CA	Question	I'm thinking of a number between one and 10. Guess which.	That's allright. We can play some other time. Bye.	
User	Answer	$number		Q1 – answer
CA	Acknowledgement	$number is correct. You must be a mind reader.		
CA	Question	One more question?		
User	Answer	Yes	No	A: Q2 B: Q2 – no
CA	Question/Acknowledgement	What is my favourite colour?	All righty then. Hope you want to play some other time.	
User	Answer	$colour		Q2 – answer
CA	Acknowledgement	Yes! $colour is correct. You are awesome.		
CA	Question	One more question?		
User	Answer	Yes	No	A: Q3 B: Q3 – no
CA	Question/Acknowledgement	Last Question. What pet do I have at home?	OK. Hope you have a great day.	

User	Answer	$animal	Q3 – answer
CA	Acknowledgement	You are just awesome. $animal is correct. That was the last question. Have a great day.	

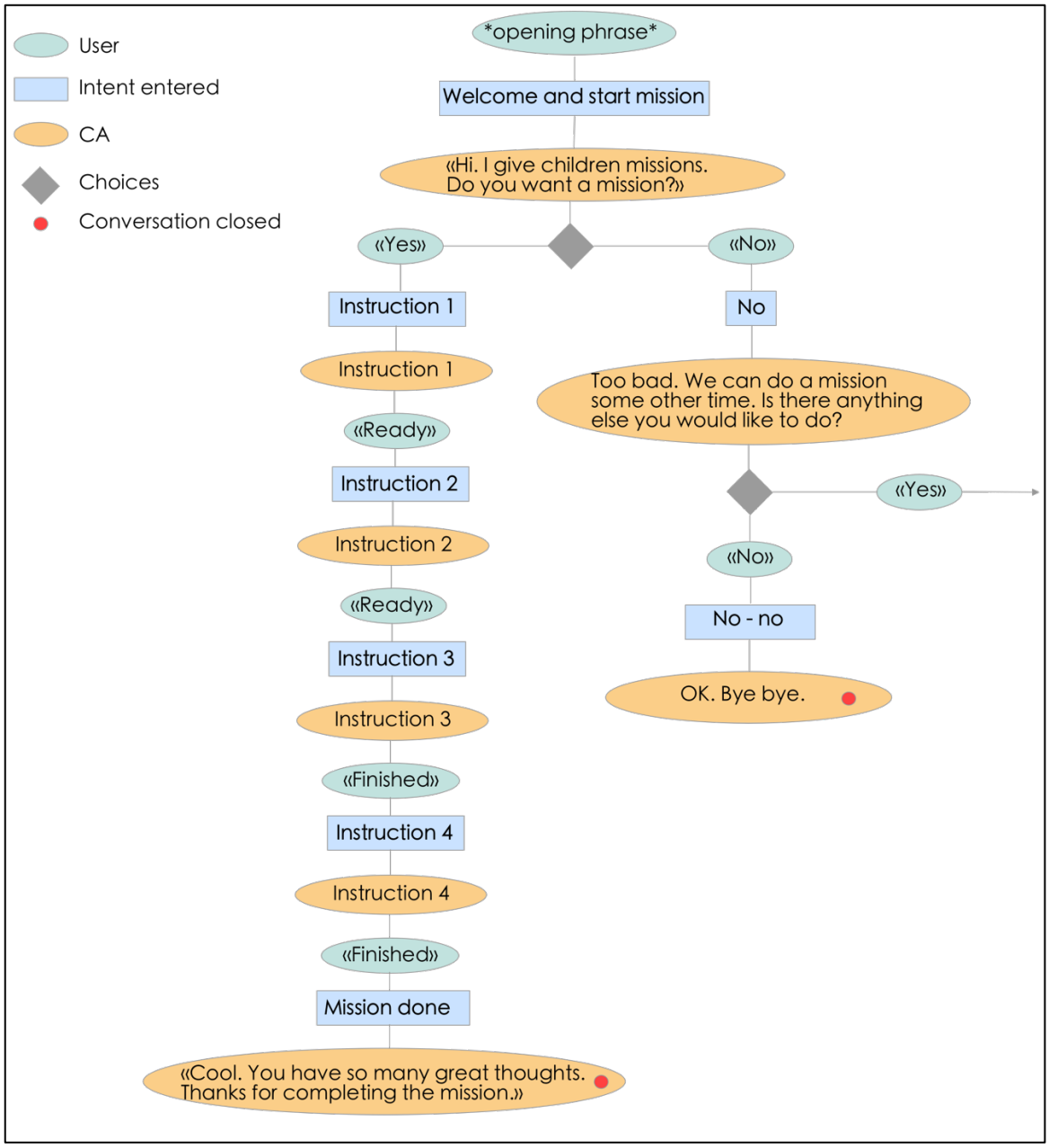
E Flow and content prototype v1: story



Story line#	Who	Sequence	Alt. input A/ Output	Alt. input B/ Output	Intent
	User	Greeting	opening phrase		Welcome and start story
	CA	Greeting	Hi		
	CA	Self-identification	I make stories together with children		
	CA	Question	Do you want to make a story with me?		
	User	Answer	Yes	No	A: Yes Line B: No
1	CA	Question	The story is about Josefine. She goes through a small forest. In front of her the path devides into two. One way goes further into the woods and the other goes into town. Which path should she follow?	That's ok. Maybe we can make a story some other time. Bye.	
	User	Answer	The woods	The town	A: The woods B: The town
2	CA	Question	Josefine starts walking into the woods. Josefine meets a small bunny. The bunny's name is Ola. He tells that there is a lot of garbage in the woods. He is angry at the people who don't throw their garbage in the garbage bins. Should Josefine help Ola clean up or should she move along?	Josefine starts to go towards the town. Josefine meets a boy. His name is Ola. He tells that the air in the city is contaminated. Should Josefine ask Ola what contaminated means or should she just go home?	
	User	Answer	Help/move along	Ask/go home	

3	CA	Question	Help: A1: Josefine says she can help clean up. Ola is very excited and say thanks. He tells Josefine that many animals eat and get stuck in the garbage and can get ill or injured. Josefine thinks this is very sad. She is going to tell everyone she knows that they have to throw the garbage in the garbage bins. This was the end of our story about Josefine Move along: A2: Josefine moves along because she do not want to help the bunny Ola. Josefine do not need to clean up after anyone else. This was the end of our story about Josefine.	Ask: B1: That the air is contaminated means that the air is dirty. You should try not to breath in too much dirty air. Ola says the air is dirty because the adults drive cars. It comes dirty air out of the cars. Josefine thinks the adults should drive a little less, because then the air will stay cleaner. This was the end of our story about Josefine. Go home: B2: Josefine goes home because she does not care about what contaminated means. This is the end of our story about Josefine.	A:Help/move along B: Ask/go home

F Flow and content prototype v1: mission



Instruction#	Who	Sequence	Alt. Input A/ Output	Alt. input B/ Output	Intent
	User	Greeting	opening phrase		Welcome and start mission
	CA	Greeting	Hi		
	CA	Self-identification	I give children missions		
	CA	Question	Do you want a mission?		
	User	Answer	Yes	No	A: Instruction 1 B: No
1	CA	Instruction	Now I want you to find out what it is like to breathe through a straw. Find a straw each. Let me know when you are ready.	Too bad. We can do a mission some other time.	
	User		Ready		Instruction 2
2	CA	Instruction	Great! Now I want you to take the straw in your mouth. Let me know when you are ready.		
	User		Ready		Instruction 3
3	CA	Instruction	Superb. Breathe in and out of the straw in five seconds. Let me know when you are finished.		
	User		Finished		Instruction 4
4	CA	Instruction	Great job! Now I want you to talk amongst yourself about how it was to breathe through the straw. Let me know when you are finished.		
	User		Finished		Mission Done
	CA	Achnowledgement	Cool. You have so many great thoughts. Thanks for completing the mission.		

G Content prototype v2: quiz

Quiz			
Category	Question	Wrong Answer	Correct Answer
Air	Intent name: Q1 Du skal få et spørsmål om luft. Her er spørsmålet: Hva er luft laget av? Er det laget av ingenting eller er det laget av mange små partikler?	Intent name: Q1AltA Trigger word: Ingenting oooh, lufta er faktisk laget av mange små partikler. Vet dere hva partikler er? Spør gjerne en voksen om dette, for det er ganske kult!	Intent name: Q1AltB Trigger word: Mange små partikler Helt riktig! Lufta har mange små partikler i seg. Vet dere hva partikler er? Spør gjerne en voksen om dette, for det er ganske kult!
Air	Intent name: Q2 Hold deg fast for her kommer et spørsmål om luft. Hva er skyer laget av? Er de laget av vanndåper eller av bomull?	Intent name: Q2AltA Trigger word: Bomull Nå tror jeg dere tuller. Skyer er jo ikke laget av bomull, men av små vanndråper. Er ikke det rart? Snakk med hverandre om dere tror man kan drikke skyene	Intent name: Q2AltB Trigger word: Vanndråper Dere er smarte altså. Vanndråper er helt riktig. Er ikke det rart? Snakk med hverandre om dere tror man kan drikke skyene.
Air	Intent name: Q3 Gjør deg klar for et spørsmål om luft. Noen ganger kan lufta være skitten. Hva tror dere gjør lufta skitten? Er det sykler eller biler?	Intent name: Q3AltB Trigger word: Sykler Ups, det er nok biler som gjør lufta skitten. Har du sett at det kommer grå røyk ut av bilene? Det kalles eksos og gjør lufta skitten. Snakk med hverandre om hva vi gjøre for at lufta skal holde seg ren	Intent name: Q3AltB Trigger word: Biler Helt riktig! Har du sett at det kommer grå røyk ut av bilene? Det kalles eksos og gjør lufta skitten. Snakk med hverandre om hva vi gjøre for at lufta skal holde seg ren
Nature	Intent name: Q1 Da skal du få et spørsmål om natur. Hvilken av disse tingene hører ikke hjemme i skogen: Fugler eller søppel	Intent name: Q1AltA Trigger word: Fugler Nå tuller du litt med meg tror jeg. Det er søppel vi ikke vil ha i skogen. Snakk med hverandre om hvordan vi kan unngå at det kommer søppel i skogen.	Intent name: Q1AltB Trigger word: Søppel Du er jammen flink! Søppel hører ikke hjemme i skogen, men noen ganger finner man det der allikevel. Snakk med hverandre om hvordan vi kan unngå at

			det kommer søppel i skogen.
Nature	Intent name: Q2 Hold på hatten, her kommer et spørsmål om natur. Hvis man kaster en brusflaske i naturen og ingen plukker den opp. Hvor mange år tar det før den er helt borte av seg selv? Tar det 10 år eller 450 år?	Intent name: Q2AltA Trigger word: 10 Det var ikke helt riktig. Det tar faktisk fire hundre og femti år før den blir helt borte. Det er kjempelenge! Snakk med hverandre om hva man burde gjøre hvis man finner søppel på bakken.	Intent name: Q2AltB Trigger word: 450 Wow, dere er kjempeflinke. Det tar fire hundre og femti år før brusflasken er borte. Det er kjempelenge! Snakk med hverandre om hva man burde gjøre hvis man finner søppel på bakken.

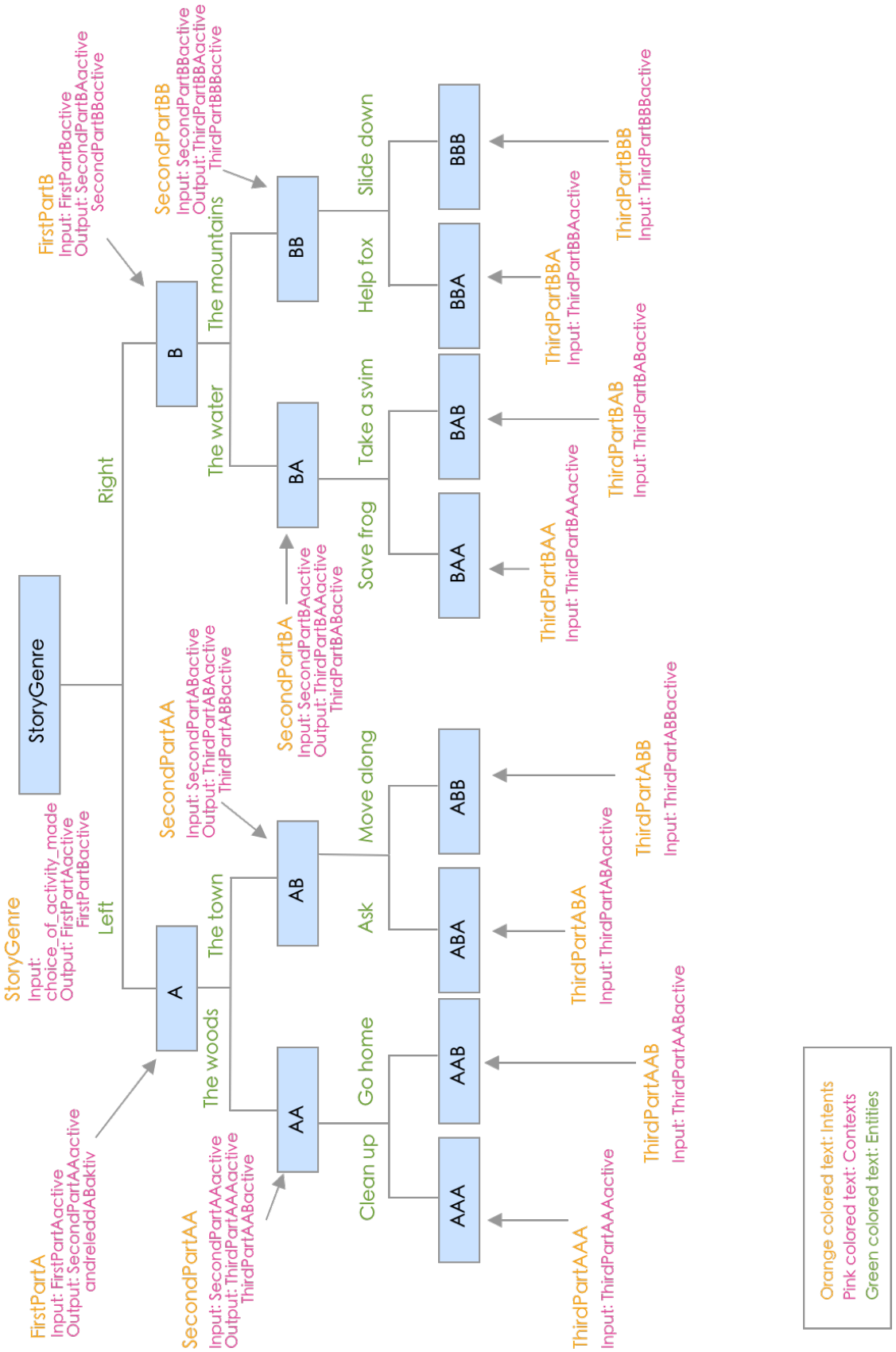
H Content prototype v2: story

Story			
Intent	Content	Trigger Word to Enter	Trigger Word to Exit
StoryGenre	Det var en gang en jente som het Elisa. Hun gikk en tur i skogen. Foran henne deler veien seg i to. Elisa må bestemme seg for hvilken vei hun skal gå. Skal hun gå til venstre eller til høyre?	Kul	Venstre/høyre
FirstPartA	Elisa går til venstre innover i skogen. Hun har kommet ganske langt inn i skogen nå. Stien deler seg igjen. Hun kan velge om hun vil gå videre innover i skogen eller inn til byen. Skal Elisa gå til skogen eller byen?	Venstre	Skogen/byen
SecondPartAA	Elisa går videre innover i skogen. Hun hører noen som gråter. Det er et lite ekorn som sitter oppe i et tre. Ekornet heter Erik og er kjempe lei seg. Han sier at mennesker sånn som Elisa har forsøplet skogen hans. Han peker ned og viser Elisa at det ligger masse søppel på bakken. Skal Elisa rydde opp søppelet eller skal hun dra hjem?	Skogen	Rydde/dra hjem
SecondPartAB	Elisa går inn i byen. Der møter hun en fugl som heter Ulrik. Han hoster og sier at det er så skitten luft. Elisa har aldri hørt om at luft kan være skitten. Skal hun spørre hva som gjør at lufta er skitten eller skal hun gå videre?	Byen	Spørre/gå videre
ThirdPartAAA	Elisa begynner å rydde søppelet hun finner på bakken oppi en plastpose. Hun synes det er veldig trist at mennesker kaster søppel i skogen for da blir dyrene lei seg. De kan til og med skade seg på søppelet eller spise det og få vondt i magen. Elisa synes alle skal kaste søppelet sitt i søppelkasser. Dette var slutten på historien vår om Elisa. Hva synes du om at mennesker kaster søppel i skogen? Snakk med hverandre om hva man kan gjøre for at det skal bli mindre søppel i skogen.	Rydde	Close

ThirdPartAAB	Elisa snur og drar hjem. Hun gidder ikke rydde søppelet noen andre har kastet. Erik får rydde det opp selv. Dette var slutten på historien vår om Elisa. Hva synes du om at mennesker kaster søppel i skogen? Snakk med hverandre om hva man kan gjøre for at det skal bli mindre søppel i skogen.	Dra hjem	Close
ThirdPartABA	Elisa spør Ulrik hva som gjør at lufta er skitten. Ulrik sier at lufta er skitten fordi det kjører så mange biler i byen. Ut av bilene kommer det grå luft som er skitten. Man burde prøve å ikke puste inn så mye grå luft. Den grå lufta heter forurensning. Det var slutten på vår historie om Elisa. Hva synes du om at mennesker som kjører bil gjør lufta skitten? Snakk med hverandre om hvordan vi kan prøve å holde lufta ren.	Spørre	Close
ThirdPartABB	Elisa går videre. Hun bryr seg ikke noe om at lufta er skitten. Det var slutten på vår historie om Elisa. Hun bryr seg ikke om at lufta er skitten, men bryr du deg om det? Spør en voksen om hva det betyr at lufta er skitten.	Gå videre	Close
FirstPartB	Elisa går til høyre innover i skogen. Litt lenger inn deler stien seg igjen. Hun kan velge om hun vil gå ned til et vann eller opp i fjellet. Skal Elisa gå til vannet eller fjellet?	Høyre	Vannet/fjellet
SecondPartBA	Elisa går bort til vannet. Der hører hun rop om hjelp. Det er en liten frosk som roper. Frosken heter Freddy. Han har satt beinet sitt fast i søppel og kommer seg ikke løs. Han forteller at han har sittet fast i mange dager. Skal Elisa hjelpe Freddy eller skal hun heller ta seg et bad?	Vannet	Redde/bade
SecondPartBB	Elisa går oppover fjellet. Plutselig hører hun noen som piper. Det er en liten rev. Reven heter Mikkel og har satt fast poten sin i en metallboks. Han får kjempevondt når han går. Han spør om Elisa kan hjelpe han med å få metallboksen av poten. Skal Elisa hjelpe Mikkel eller ake ned fra fjellet?	Fjellet	Hjelpe/ake

ThirdPartBAA	Elisa vil redde Freddy. Hun går bort og vikler Freddys ben ut av masse plastposer. Han satt ordentlig godt fast. Stakkars Freddy. Det var bra Elisa ville redde han. Han takker Elisa og hopper glad og fornøyd videre. Dette var slutten på vår historie om Elisa. Det er veldig dumt at noen kaster søppel i naturen. Da kan dyr slik som Freddy sette seg fast. Snakk med hverandre om hva dere synes man skal gjøre for at dyr ikke skal sette seg fast i søppel.	redde	Close
ThirdPartBAB	Elisa vil heller bade enn å hjelpe Freddy. Det er jo ikke hennes feil at han sitter fast. Freddy blir sittende igjen mens Elisa bader og kommer seg ikke løs i dag heller. Dette var slutten på vår historie om Elisa. Det er veldig dumt at noen kaster søppel i naturen. Da kan dyr slik som Freddy sette seg fast. Snakk med hverandre om hva dere synes man skal gjøre for at dyr ikke skal sette seg fast i søppel.	Bade	Close
ThirdPartBBA	Elisa vil hjelpe Mikkel med å få av metallboksen. Hun tar tak i poten hans og hjelper med å dra den av. Mikkel blir kjempeglad og sier tusen takk til Elisa. Dette var slutten på vår historie om Elisa. Når mennesker kaster søppel i naturen kan dyr slik som Mikkel sette seg fast i det. Snakk med hverandre om hva dere synes man skal gjøre for at dyr ikke skal sette seg fast i søppel.	Hjelpe	Close
ThirdPartBBB	Elisa vil heller ake ned den kule akebakken ned fra fjellet. Det er jo ikke hennes feil at Mikkel har satt seg fast. Dette var slutten på vår historie om Elisa. Når mennesker kaster søppel i naturen kan dyr slik som Mikkel sette seg fast i det. Snakk med hverandre om hva dere synes man skal gjøre for at dyr ikke skal sette seg fast i søppel.	Ake	Close

I Flowchart prototype v2: story



J Content prototype v2: mission

Mission	
Level of Difficulty	Content
Lett (easy)	Du skal få et lett oppdrag av meg. Hør godt etter nå. Oppdraget i dag handler om lufta vi puster. Dere skal holde pusten i fem sekunder. Når dere har gjort det vil jeg at dere prater om hvordan det føltes å ikke få luft inn i kroppen. Lykke til! teller for dere. Klar, ferdig, GÅ. 1, 2, 3, 4, 5
Vanskelig (hard)	Oppdraget i dag handler om lufta vi puster. Vi trenger ekstra mye luft i kroppen når vi gjør aktiviteter som gjør oss slitne. Nå setter vi i gang! Når jeg har telt ferdig til fem må alle sitte på huk på gulvet. 1, 2, 3, 4, 5. Flott! Nå vil jeg at dere skal ta fem spensthopp. Jeg teller høyt for dere! 1, 2, 3, 4, 5. Dere er flinke! Nå vil jeg at dere prater om hvordan det føltes når kroppen trengte mer luft når dere gjorde en aktivitet som gjør oss slitne. Vi snakkes!

K Code prototype v3

```
'use strict';

// Import the Dialogflow module from the Actions on Google client
library.
const {dialogflow} = require('actions-on-google');

// Import the firebase-functions package for deployment.
const functions = require('firebase-functions');

// Instantiate the Dialogflow client.
const app = dialogflow({debug: true});

//intents
const WELCOME_INTENT = 'Default Welcome Intent'
const FALLBACK_INTENT = 'Default Fallback Intent'
const ACTIVITY_INTENT = 'velg_aktivitet'

const STORYFORSTELEDD_A_INTENT = 'StoryforsteleddA'
const STORYFORSTELEDD_B_INTENT = 'StoryforsteleddB'
const STORYANDRELEDD_AA_INTENT = 'StoryandreleddAA'
const STORYANDRELEDD_AB_INTENT = 'StoryandreleddAB'
const STORYANDRELEDD_BA_INTENT = 'StoryandreleddBA'
const STORYANDRELEDD_BB_INTENT = 'StoryandreleddBB'
const STORYTREDJELEDD_AAA_INTENT = 'StorytredjeleddAAA'
const STORYTREDJELEDD_AAB_INTENT = 'StorytredjeleddAAB'
const STORYTREDJELEDD_ABA_INTENT = 'StorytredjeleddABA'
const STORYTREDJELEDD_ABB_INTENT = 'StorytredjeleddABB'
const STORYTREDJELEDD_BAA_INTENT = 'StorytredjeleddBAA'
const STORYTREDJELEDD_BAB_INTENT = 'StorytredjeleddBAB'
const STORYTREDJELEDD_BBA_INTENT = 'StorytredjeleddBBA'
const STORYTREDJELEDD_BBB_INTENT = 'StorytredjeleddBBB'

const REPEAT_INTENT = 'repeat'
const PARAPHRASE_INTENT = 'paraphrase'
const PROMP_INTENT = 'promp'
const VITS_INTENT = 'siste svar vits'
const TUTORIAL_INTENT = 'tutorial'
const TUTORIAL2_INTENT = 'tutorial2'
const TUTORIAL3_INTENT = 'tutorial3'

//entities
const TYPE_AKTIVITET_ENTITY = 'aktivitet'
const RETNING_ENTITY = 'retning'
const STED_ENTITY = 'sted'
const ELISASVALG_ENTITY = 'Elisas_valg'
const ANSWERJN_ENTITY = 'answerJN'
```

```
//variables
var previousItem = null;
var paraphraseItem = null;
var repromptTeller = 0;

function saveAndSend(conv, respons) {
  console.log(respons)
  previousItem = respons
  conv.ask(respons)
}

//repeat last output
app.intent(REPEAT_INTENT, (conv) => {
  repromptTeller = 0;
    repeat(conv);
})

function repeat(conv) {
  repromptTeller = 0;
  var items = Array('Det jeg sa var',
      'Beklager, jeg sa',
      'La meg si det igjen');
  var item = items[Math.floor(Math.random()*items.length)];
  if(previousItem.startsWith('<speak>')) {
    previousItem = previousItem.replace('<speak>','');
    if(previousItem.endsWith('</speak>')) {
      previousItem = previousItem.replace('</speak>', '');
    } if(previousItem.startsWith('Det jeg sa var')) {
      previousItem = previousItem.replace('Det jeg sa var', '');
    } if(previousItem.startsWith('Beklager, jeg sa')) {
      previousItem = previousItem.replace('Beklager, jeg sa', '');
    } if(previousItem.startsWith('La meg si det igjen')) {
      previousItem = previousItem.replace('La meg si det igjen',
'');
    }
    previousItem = '<speak>' + item + '<break time = "0.5s"/>' +
previousItem + '</speak>';
  }
  conv.ask(previousItem);
}

app.intent(PARAPHRASE_INTENT, (conv) => {
  repromptTeller = 0;
  paraphrase(conv);
})
```

```
function paraphrase(conv) {
  repromptTeller = 0;
  conv.ask(paraphraseItem);
}

// Handeling Dialogflow NO_INPUT intent.
// Triggered by no input
app.intent('actions_intent_NO_INPUT', (conv) => {
  repromptTeller = 0;
  const repromptCount =
parseInt(conv.arguments.get('REPROMPT_COUNT'));
  if (repromptCount === 0) {
    saveAndSend(conv, '<speak> Hallo? <break time="0.5s"/> Jeg
hørte ikke hva dere sa. <break time="0.5s"/> Vent med å svare til
lysene mine står i ro. </speak>');
    paraphraseItem = 'Prøv å svare meg igjen når lysene mine står
i ro'
  } else if (repromptCount === 1) {
    saveAndSend(conv, '<speak> Jeg klarte ikke å høre dere <break
time="0.5s"/> Dere må svare meg for at jeg skal vite hva dere vil
gjøre</speak>');
    paraphraseItem = 'Dere må svare meg'
  } else if (conv.arguments.get('IS_FINAL_REPROMPT')) {
    previousItem = ('<speak>Beklager, nå har jeg problemer med å
høre. <break time="0.5s"/> La oss prøve igjen senere. <break
time="0.5s"/>Ha det på badet.</speak>');
    conv.close(previousItem);
  }
});

app.intent(FALLBACK_INTENT, (conv) => {
    if(repromptTeller === 0) {
      conv.ask('<speak> nå forstod jeg ikke. <break time="0.5s"/>
Prøv å si det en gang til.</speak>');
      repromptTeller++;
    } else if(repromptTeller === 1) {
      conv.ask('<speak>jeg skjønte ikke hva du sa. <break
time="0.5s"/> Prøv å si det igjen</speak>');
      repromptTeller++;
    } else if(repromptTeller === 2) {
      conv.ask('<speak> Nå har jeg problemer med å forstå
deg.<break time="0.5s"/> La oss begynne på nytt: Jeg kan gi deg
oppdrag, lage fortelling eller spille spørrespill. </speak>');
      repromptTeller++;
    } else if(repromptTeller === 3) {
```

```
      conv.ask('Fortell meg om du vil ha oppdrag, lage fortelling
eller spille spørrespill');
      repromptTeller++;
    } else if(repromptTeller === 4) {
      repromptTeller = 0;
      conv.close('<speak> Unnskyld, nå skjønner jeg ingenting.
<break time="0.5s"/> La oss prøve igjen senere. <break
time="0.5s"/>Ha det på badet.</speak>');
    }
})


app.intent(WELCOME_INTENT, (conv) => {
  repromptTeller = 0;
  var items = Array('<speak> <audio src =
"https://actions.google.com/sounds/v1/cartoon/clang_and_wobble.ogg
"> </audio> Hei! Jeg kan gi deg oppdrag, lage fortelling eller
spille spørrespill. <break time="0.5s"/> Velger du oppdrag,
fortelling eller spørrespill? </speak>',
      '<speak>Halloisen! <break time="0.5s"/> Jeg er en robot som
kan tre ting.<break time="0.5s"/> Lage fortelling, gi oppdrag
eller spille spørrespill.<break time="0.5s"/> Vil du lage
fortelling, få et oppdrag eller spille spørrespill?</speak>',
      '<speak>Bø!<break time="0.5s"/> Jeg liker å spille
spørrespill, gi barn oppdrag og lage fortellinger. <break
time="0.5s"/>Vil du spille spørrespill, få et oppdrag eller lage
en fortelling?</speak>');
  var item = items[Math.floor(Math.random()*items.length)];
    saveAndSend(conv, item);
    paraphraseItem = 'Velg om jeg skal gi deg et oppdrag, lage
fortelling eller spille spørrespill'
})

app.intent(ACTIVITY_INTENT, (conv) => {
  console.log(ACTIVITY_INTENT);
  repromptTeller = 0;
    const aktivitet =
conv.parameters[TYPE_AKTIVITET_ENTITY].toString().toLowerCase();
    if(aktivitet == "oppdrag") {
      conv.close('<speak> Du skal få et oppdrag av meg. <break
time="0.5s"/> Hør godt etter nå. <break time="0.5s"/> Oppdraget i
dag handler om lufta vi puster. <break time="0.5s"/> Vi trenger
ekstra mye luft i kroppen når vi gjør aktiviteter som gjør oss
slitne. <break time="0.5s"/> Nå setter vi i gang! <break
time="0.5s"/> Når jeg har telt ferdig til fem må alle sitte på huk
på gulvet <break time="0.5s"/>  1 <break time="1.0s"/> 2 <break
```

```
time="1.0s"/> 3 <break time="1s"/> 4 <break time="1.0s"/> 5 <break
time="1.0s"/> Flott! <break time="0.5s"/> Nå vil jeg at dere skal
ta fem spensthopp. <break time="0.5s"/> Jeg teller høyt for
dere!<break time="0.5s"/> 1 <break time="2.0s"/> 2 <break
time="2.0s"/> 3 <break time="2.0s"/> 4 <break time="2.0s"/> 5
<break time="2.0s"/> Dere er flinke! <break time="0.5s"/> Nå vil
jeg at dere prater om hvordan det føltes når kroppen trengte mer
luft når dere gjorde en aktivitet som gjør oss slitne. <break
time="0.5s"/> Vi snakkes!</speak>')
    } else if (aktivitet == "fortelling") {
          saveAndSend(conv, '<speak> Da lager vi en fortelling
sammen. <break time="0.5s"/>Hør godt etter nå.<break time="0.5s"/>
Det var en gang en jente som het Elisa. <break time="0.5s"/>Hun
gikk en tur i skogen. <break time="0.5s"/>Foran henne deler veien
seg i to. <break time="0.5s"/>Elisa må bestemme seg for hvilken
vei hun skal gå. <break time="0.5s"/> Skal hun gå til venstre
eller til høyre?</speak>');
          paraphraseItem = 'Vil du at Elisa skal gå til høyre
eller venstre?'
    } else if (aktivitet == "spørrespill") {
            var items = Array
            ('<speak>Du skal få et spørsmål om luft. <break
time="0.5s"/>Her er spørsmålet: <break time="0.5s"/>Hva er luft
laget av? Er det laget av ingenting eller er det laget av mange
små partikler?</speak>',
            '<speak>Hold deg fast for her kommer et spørsmål om
luft. <break time="0.5s"/>Hva er skyer laget av? Er de laget av
vanndråper eller av bomull?</speak>',
            '<speak>Gjør deg klar for et spørsmål om luft. <break
time="0.5s"/>Noen ganger kan lufta være skitten. <break
time="0.5s"/>Hva tror dere gjør lufta skitten? Er det sykler eller
biler?</speak>',
            '<speak>Du skal få et spørsmål om natur. <break
time="0.5s"/>Hvilken av disse tingene hører ikke hjemme i skogen:
Fugler eller søppel</speak>',
            '<speak>Hold på hatten, her kommer et spørsmål om
natur. <break time="0.5s"/>Hvis man kaster en brusflaske i naturen
og ingen plukker den opp. Hvor mange år tar det før den er helt
borte av seg selv? Tar det 10 år eller 450 år?</speak>');
            var item =
items[Math.floor(Math.random()*items.length)];
            saveAndSend(conv, item);
    } else {
        saveAndSend(conv, "Du må velge om vi skal gjøre oppdrag,
fortelling eller spørrespill");
```

```
        paraphraseItem = 'Velg oppdrag, fortelling eller
spørrespill'
    }
})

app.intent(STORYFORSTELEDD_A_INTENT, (conv) =>{
  repromptTeller = 0;
    const retning = conv.parameters[RETNING_ENTITY].toLowerCase();
    if(retning == "venstre") {
        saveAndSend(conv, '<speak> Elisa går til venstre innover i
skogen. <break time="0.5s"/>Hun har kommet ganske langt inn i
skogen nå. <break time="0.5s"/>Stien deler seg igjen. <break
time="0.5s"/>Hun kan velge om hun vil gå videre innover i skogen
eller inn til byen. <break time="0.5s"/>Skal Elisa gå til skogen
eller byen?</speak>');
        paraphraseItem = 'Skal Elisa gå til skogen eller byen?'
    } else {
        saveAndSend(conv, "du må velge om Elisa skal gå til høyre
eller venstre");
        paraphraseItem = 'du må velge høyre eller venstre'
    }
})

app.intent(STORYFORSTELEDD_B_INTENT, (conv) =>{
  repromptTeller = 0;
    const retning = conv.parameters[RETNING_ENTITY].toLowerCase();
    if (retning == "høyre") {
        saveAndSend(conv, '<speak>Elisa går til høyre innover i
skogen. <break time="0.5s"/>Litt lenger inn deler stien seg igjen.
<break time="0.5s"/>Hun kan velge om hun vil gå ned til et vann
eller opp i fjellet. <break time="0.5s"/>Skal Elisa gå til vannet
eller fjellet?</speak>');
        paraphraseItem = 'Skal Elisa gå til vannet eller fjellet?'
    } else {
        saveAndSend(conv, "du må velge om Elisa skal gå til høyre
eller venstre");
        paraphraseItem = 'du må velge høyre eller venstre'
    }
})

app.intent(STORYANDRELEDD_AA_INTENT, (conv) => {
  repromptTeller = 0;
    const sted = conv.parameters[STED_ENTITY].toLowerCase();
    if(sted == "skogen") {
        saveAndSend(conv, '<speak>Elisa går videre innover i
skogen. <break time="0.5s"/>Hun hører noen som gråter.<break
```

```
time="0.5s"/> Det er et lite ekorn som sitter oppe i et tre.
<break time="0.5s"/>Ekornet heter Erik og er kjempe lei seg.
<break time="0.5s"/>Han sier at mennesker sånn som Elisa har
forsøplet skogen hans. <break time="0.5s"/>Han peker ned og viser
Elisa at det ligger masse søppel på bakken. <break
time="0.5s"/>Skal Elisa rydde opp søppelet eller skal hun dra
hjem?</speak>');
        paraphraseItem = 'Skal Elisa rydde opp eller dra hjem?'
    } else {
        saveAndSend(conv, "Du må velge skogen eller byen");
        paraphraseItem = 'Velg hvor Elisa skal gå, til skogen
eller til byen'
    }
})

app.intent(STORYANDRELEDD_AB_INTENT, (conv) => {
  repromptTeller = 0;
    const sted = conv.parameters[STED_ENTITY].toLowerCase();
    if(sted == "byen") {
        saveAndSend(conv, '<speak>Elisa går inn i byen. <audio
src=
"https://actions.google.com/sounds/v1/ambiences/subway_station_nyc
.ogg" clipBegin= "27.0s" fadeOutDur="2.0s" clipEnd = "30.5s">
</audio> <break time="0.5s"/>Der møter hun en fugl som heter
Ulrik. <break time="0.5s"/> Han hoster og sier at det er så
skitten luft. <break time="0.5s"/>Elisa har aldri hørt om at luft
kan være skitten. <break time="0.5s"/>Skal hun spørre hva som gjør
at lufta er skitten eller skal hun gå videre?</speak>');
        paraphraseItem = 'Skal Elisa spørre eller gå videre?'
    } else {
        saveAndSend(conv, "Du må velge skogen eller byen");
        paraphraseItem = 'Velg hvor Elisa skal gå, skogen eller
byen'
    }
})

app.intent(STORYANDRELEDD_BA_INTENT, (conv) => {
  repromptTeller = 0;
    const sted = conv.parameters[STED_ENTITY].toLowerCase();
    if(sted == "vannet") {
        saveAndSend(conv, '<speak>Elisa går bort til vannet.
<break time="0.2s"/>Der hører hun rop om hjelp. <break
time="0.5s"/>Det er en liten frosk som roper. <break
time="0.5s"/>Frosken heter Freddy. <break time="0.5s"/>Han har
satt beinet sitt fast i søppel og kommer seg ikke løs. <break
time="0.5s"/>Han forteller at han har sittet fast i mange dager.
```

```
<break time="0.5s"/>Skal Elisa redde Freddy eller skal hun heller
ta seg et bad?</speak>');
        paraphraseItem = 'Skal Elisa redde Freddy eller ta seg et
bad?'
    } else {
        saveAndSend(conv, "Du må velge om Elisa skal gå til vannet
eller til fjellet");
        paraphraseItem = 'Velg hvor Elisa skal gå, til vannet
eller til fjellet'
    }
})

app.intent(STORYANDRELEDD_BB_INTENT, (conv) => {
  repromptTeller = 0;

    const sted = conv.parameters[STED_ENTITY].toLowerCase();
    if(sted == "fjellet") {
        saveAndSend(conv, '<speak>Elisa går oppover fjellet.<break
time="0.5s"/> Plutselig hører hun noen som piper.<break
time="0.5s"/> Det er en liten rev.<break time="0.5s"/> Reven heter
Mikkel og har satt fast poten sin i en metallboks. <break
time="0.5s"/>Han får kjempevondt når han går. <break
time="0.5s"/>Han spør om Elisa kan hjelpe han med å få
metallboksen av poten. <break time="0.5s"/>Skal Elisa hjelpe
Mikkel eller ake ned fra fjellet?</speak>');
        paraphraseItem = 'Skal Elisa hjelpe Mikkel eller ake ned?'
    } else {
        saveAndSend(conv, "Du må velge om elisa skal gå til vannet
eller til fjellet");
        paraphraseItem = 'Velg hvor Elisa skal gå, til vannet
eller fjellet'
    }
})

app.intent(STORYTREDJELEDD_AAA_INTENT, (conv) => {
  repromptTeller = 0;
    const Elisas_valg =
conv.parameters[ELISASVALG_ENTITY].toLowerCase();
    if(Elisas_valg == "rydde") {
        conv.close('<speak>Elisa begynner å rydde søppelet hun
finner på bakken oppi en plastpose. <break time="0.5s"/>Hun synes
det er veldig trist at mennesker kaster søppel i skogen for da
blir dyrene lei seg. <break time="0.5s"/>De kan til og med skade
seg på søppelet eller spise det og få vondt i magen. <break
time="0.5s"/>Elisa synes alle skal kaste søppelet sitt i
søppelkasser. <break time="0.5s"/>Dette var slutten på historien
```

```
vår om Elisa. <break time="0.5s"/>Hva synes du om at mennesker
kaster søppel i skogen? <break time="0.5s"/>Snakk med hverandre om
hva man kan gjøre for at det skal bli mindre søppel i skogen.
<break time="0.5s"/> Ha det bra!</speak>')
    } else {
        saveAndSend(conv, "Du må velge om Elisa skal rydde eller
dra hjem" );
        paraphraseItem = 'Velg hva Elisa skal gjøre, rydde eller
dra hjem'
    }
})

app.intent(STORYTREDJELEDD_AAB_INTENT, (conv) => {
  repromptTeller = 0;
    const Elisas_valg =
conv.parameters[ELISASVALG_ENTITY].toLowerCase();
    if(Elisas_valg == "dra") {
        conv.close('<speak>Elisa snur og drar hjem. <break
time="0.5s"/>Hun gidder ikke rydde søppelet noen andre har kastet.
<break time="0.5s"/>Erik får rydde det opp selv. <break
time="0.5s"/>Dette var slutten på historien vår om Elisa. <break
time="0.5s"/>Hva synes du om at mennesker kaster søppel i skogen?
<break time="0.5s"/>Snakk med hverandre om hva man kan gjøre for
at det skal bli mindre søppel i skogen.<break time="0.5s"/> Vi
snakkes!</speak>')
    } else {
        saveAndSend(conv, "Du må velge om Elisa skal rydde eller
dra hjem" );
        paraphraseItem = 'Velg hva Elisa skal gjøre, rydde eller
dra hjem'
    }
})

app.intent(STORYTREDJELEDD_ABA_INTENT, (conv) => {
  repromptTeller = 0;
    const Elisas_valg =
conv.parameters[ELISASVALG_ENTITY].toLowerCase();
    if(Elisas_valg == "spørre") {
        conv.close('<speak>Elisa spør Ulrik hva som gjør at lufta
er skitten.<break time="0.5s"/> Ulrik sier at lufta er skitten
fordi det kjører så mange biler i byen. <break time="0.5s"/>Ut av
bilene kommer det grå luft som er skitten.<break time="0.5s"/> Man
burde prøve å ikke puste inn så mye grå luft. <break
time="0.5s"/>Den grå lufta heter forurensning. <break
time="0.5s"/>Det var slutten på vår historie om Elisa.<break
time="0.5s"/> Hva synes du om at mennesker som kjører bil gjør
```

```
lufta skitten? <break time="0.5s"/>Snakk med hverandre om hvordan
vi kan prøve å holde lufta ren.<break time="0.5s"/> Ha det
bra!</speak>')
      } else {
        saveAndSend(conv, "Du må velge om elisa skal spørre eller
gå videre" );
        paraphraseItem = 'Velg hva Elisa skal gjøre, spørre eller
gå videre'
    }
})

app.intent(STORYTREDJELEDD_ABB_INTENT, (conv) => {
  repromptTeller = 0;
    const Elisas_valg =
conv.parameters[ELISASVALG_ENTITY].toLowerCase();
    if (Elisas_valg == "gå") {
        conv.close('<speak>Elisa går videre. <break
time="0.5s"/>Hun bryr seg ikke noe om at lufta er skitten. <break
time="0.5s"/>Det var slutten på vår historie om Elisa.<break
time="0.5s"/> Elisa bryr seg ikke om at lufta er skitten, men bryr
du deg om det? <break time="0.5s"/>Spør en voksen om hva det betyr
at lufta er skitten.<break time="0.5s"/> Vi snakkes! </speak>')
    } else {
        saveAndSend(conv, "Du må velge om elisa skal spørre eller
gå videre" );
        paraphraseItem = 'Velg hva Elisa skal gjøre, spørre eller
gå videre'
    }
})

app.intent(STORYTREDJELEDD_BAA_INTENT, (conv) => {
  repromptTeller = 0;
    const Elisas_valg =
conv.parameters[ELISASVALG_ENTITY].toLowerCase();
    if(Elisas_valg == "redde") {
        conv.close('<speak>Elisa vil redde Freddy.<break
time="0.5s"/> Hun går bort og vikler Freddys ben ut av masse
plastposer. <break time="0.5s"/>Han satt ordentlig godt fast.
<break time="0.5s"/>Stakkars Freddy.<break time="0.5s"/> Det var
bra Elisa ville redde han. <break time="0.5s"/>Han takker Elisa og
hopper glad og fornøyd videre.<break time="0.5s"/> Dette var
slutten på vår historie om Elisa. <break time="0.5s"/>Det er
veldig dumt at noen kaster søppel i naturen. <break
time="0.5s"/>Da kan dyr slik som Freddy sette seg fast. <break
time="0.5s"/>Snakk med hverandre om hva dere synes man skal gjøre
```

```
for at dyr ikke skal sette seg fast i søppel.<break time="0.5s"/>
Ha det bra!</speak>')
      } else {
        saveAndSend(conv, "Du må velge om elisa skal redde freddy
eller bade" );
      paraphraseItem = 'Velg hva Elisa skal gjøre, redde freddy
eller bade';
    }
})

app.intent(STORYTREDJELEDD_BAB_INTENT, (conv) => {
  repromptTeller = 0;
    const Elisas_valg =
conv.parameters[ELISASVALG_ENTITY].toLowerCase();
    if(Elisas_valg == "bade") {
        conv.close('<speak>Elisa vil heller bade enn å hjelpe
Freddy. <break time="0.5s"/>Det er jo ikke hennes feil at han
sitter fast. <break time="0.5s"/>Freddy blir sittende igjen mens
Elisa bader og kommer seg ikke løs i dag heller.<break
time="0.5s"/>Dette var slutten på vår historie om Elisa. <break
time="0.5s"/>Det er veldig dumt at noen kaster søppel i naturen.
<break time="0.5s"/>Da kan dyr slik som Freddy sette seg fast.
<break time="0.5s"/>Snakk med hverandre om hva dere synes man skal
gjøre for at dyr ikke skal sette seg fast i søppel.<break
time="0.5s"/> Ha det bra!</speak>')
    } else {
        saveAndSend(conv, "Du må velge om elisa skal redde freddy
eller bade" );
      paraphraseItem = 'Velg hva Elisa skal gjøre, redde freddy
eller bade';
    }
})

app.intent(STORYTREDJELEDD_BBA_INTENT, (conv) => {
  repromptTeller = 0;
    const Elisas_valg =
conv.parameters[ELISASVALG_ENTITY].toLowerCase();
    if(Elisas_valg == "hjelpe") {
        conv.close('<speak>Elisa vil hjelpe Mikkel ut av
metallboksen. <break time="0.5s"/>Hun tar tak i poten hans og
hjelper med å dra den av.<break time="0.5s"/> Mikkel blir
kjempeglad. <audio
src="https://actions.google.com/sounds/v1/animals/cat_purr_close.o
gg"></audio> Han sier tusen takk til Elisa. <break
time="0.5s"/>Dette var slutten på vår historie om Elisa. <break
time="0.5s"/>Når mennesker kaster søppel i naturen kan dyr slik
```

```
som Mikkel sette seg fast i det. <break time="0.5s"/>Snakk med
hverandre om hva dere synes man skal gjøre for at dyr ikke skal
sette seg fast i søppel. <break time="0.5s"/> Ha det
bra!</speak>')
      } else {
        saveAndSend(conv, "Du må velge om elisa skal hjelpe mikkel
eller ake" );
        paraphraseItem = 'Velg hva Elisa skal gjøre, hjelpe mikkel
eller ake'
    }
})

app.intent(STORYTREDJELEDD_BBB_INTENT, (conv) => {
  repromptTeller = 0;
    const Elisas_valg =
conv.parameters[ELISASVALG_ENTITY].toLowerCase();
    if(Elisas_valg == "ake") {
        conv.close('<speak>Elisa vil heller ake ned den kule
akebakken ned fra fjellet. <break time="0.5s"/>Det er jo ikke
hennes feil at Mikkel har satt seg fast. <break time="0.5s"/>Dette
var slutten på vår historie om Elisa. <break time="0.5s"/>Når
mennesker kaster søppel i naturen kan dyr slik som Mikkel sette
seg fast i det. <break time="0.5s"/>Snakk med hverandre om hva
dere synes man skal gjøre for at dyr ikke skal sette seg fast i
søppel<break time="0.5s"/> Ha det bra!</speak>')
    } else {
        saveAndSend(conv, "Du må velge om elisa skal hjelpe mikkel
eller ake" );
        paraphraseItem = 'Velg hva Elisa skal gjøre, hjelpe mikkel
eller ake'
    }
})

app.intent(PROMP_INTENT, (conv) => {
  repromptTeller = 0;
  conv.ask('<speak> <audio src =
"https://actions.google.com/sounds/v1/human_voices/human_fart.ogg"
> </audio> Ops, beklager! Det skal jeg ikke gjøre igjen. Vil du ha
oppdrag, lage fortelling eller spille spørrespill?</speak> ');
})

app.intent(VITS_INTENT, (conv) => {
  repromptTeller = 0;
  const answerJN = conv.parameters[ANSWERJN_ENTITY].toLowerCase();
  if(answerJN == "ja") {
```

```
    conv.close('<speak>Rompa til presten. <audio src =
"https://actions.google.com/sounds/v1/human_voices/man_laugh_and_k
nee_slap.ogg" clipEnd = "3.0s"> </audio> Jeg er ganske
morsom</speak>');
  } else if (answerJN == "nei") {
    conv.close('<speak> Nå var jeg jo nesten ferdig. <break
time="0.5s"/> Da får du ikke vite siste del av vitsen
min</speak>');
  }
})

app.intent(TUTORIAL_INTENT, (conv) => {
  repromptTeller = 0;
  saveAndSend(conv, '<speak>Jeg kan hjelpe dere med å forstå
hvordan man skal prate med roboter. <break time="0.5s"/> Roboter
er ikke så flinke til å forstå når mennesker prater. <break
time="0.5s"/> Men det er noen ting dere kan gjøre for at jeg skal
forstå bedre. <break time="0.5s"/> Hvis flere snakker til meg er
det ikke sikkert jeg forstår. <break time="0.5s"/> Bare en av dere
kan svare meg om gangen <break time="0.5s"/>Hvor mange kan svare
meg?</speak>');
})

app.intent(TUTORIAL2_INTENT, (conv) => {
  repromptTeller = 0;
  saveAndSend(conv, '<speak> Dere er flinke!. <break time="0.5s"/>
Én siste ting dere må huske på: <break time="0.5s"/> Etter at jeg
har sagt noe trenger jeg litt tid for å gjøre meg klar for å høre.
<break time="0.5s"/>  Vent med å svare til lysene mine står i ro!
<break time="0.5s"/> Gjenta til meg. <break time="0.5s"/> Hva må
dere vente på før dere svarer meg?</speak>');
})

app.intent(TUTORIAL3_INTENT, (conv) => {
  repromptTeller = 0;
  saveAndSend(conv, '<speak>Hurra! <break time="0.5s"/> Nå er dere
klare for å snakke med meg. <break time="0.5s"/> Si "Hei" til meg
for å finne ut hva vi kan gjøre sammen.</speak>');
})


// Set the DialogflowApp object to handle the HTTPS POST request.
exports.dialogflowFirebaseFulfillment =
functions.https.onRequest(app);
```

L Complete list of identified trouble sources

Trouble Source	Precursor
Children's behaviour	• Associative: unpredictable behaviour and random acts. Violate the contextual limits of the conversation. • Playful: play and silliness confound communication. • Inefficient speech patterns: fractured speak, stops, hesitations, restarts. • Interrupting: Responds to what the CA says before it is done talking
Unrealistic expectations of capabilities	• Expectations of humanlike capacity in interactions. However, CA is incapable of engaging in complex exchanges
Technological inexperience	• Children not aware of when the technology can detect their spoken input
The group setting	• Multiple responders: simultaneous talk, succeeding responses or conflicting answers. • No response: uncertainty of who should talk results in no one talking. • Talking amongst themselves. Debating answers means the CA can detect talk not intended as responses.
Linguistic skills	• Mispronunciation
Technological weakness	• Flawed Natural Language Processing • Limited training

M Pamphlet last evaluation

Evaluering av prototype i felt

Vår kontaktinformasjon

Kristine: –telefonnummer–, –e-post–
Julie: –telefonnummer–, –e-post–

Barn med samtykke fra foreldre til deltagelse:

–Liste med navn her–

Huskeliste for gjennomføring

1. Sett på opptaker
2. Avtal hvilket barn som skal svare stemmeagenten
3. Start stemmeagent: "Hei Google. Snakke med roboten vår"
4. Gjennomfør tutorial: "Hvordan snakker man med roboter?"
5. La barnet velge hvilken aktivitet de skal gjøre (oppdrag, fortelling eller spørrespill)
6. Gjennomfør aktiviteten
7. Gjenta steg 5-6 til alle aktivitetene er forsøkt gjennomført minst én gang (Dere må starte stemmeagenten på nytt: "Hei Google. Snakke med roboten vår.")
8. Fyll ut arket til Oppfinnerboken
9. Skru av opptaker
10. Husk å noter kjapt i tilbakemeldingsskjemaet :-)

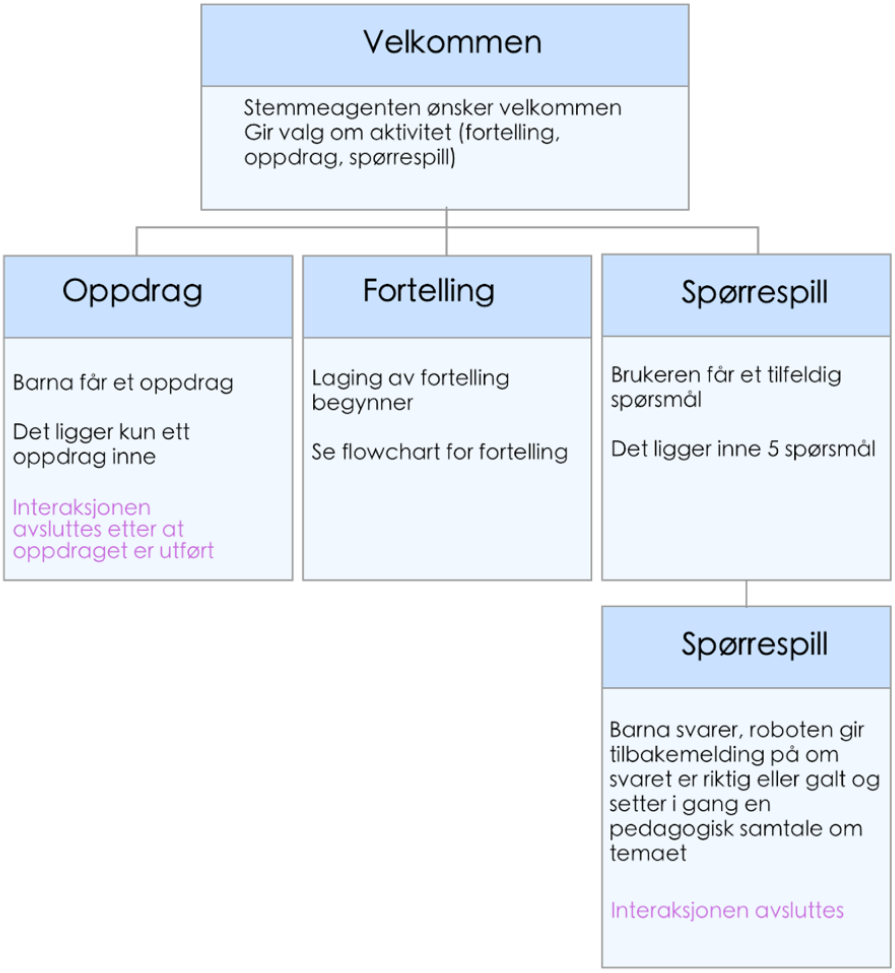
Vanlige årsaker til problemer i interaksjonen:

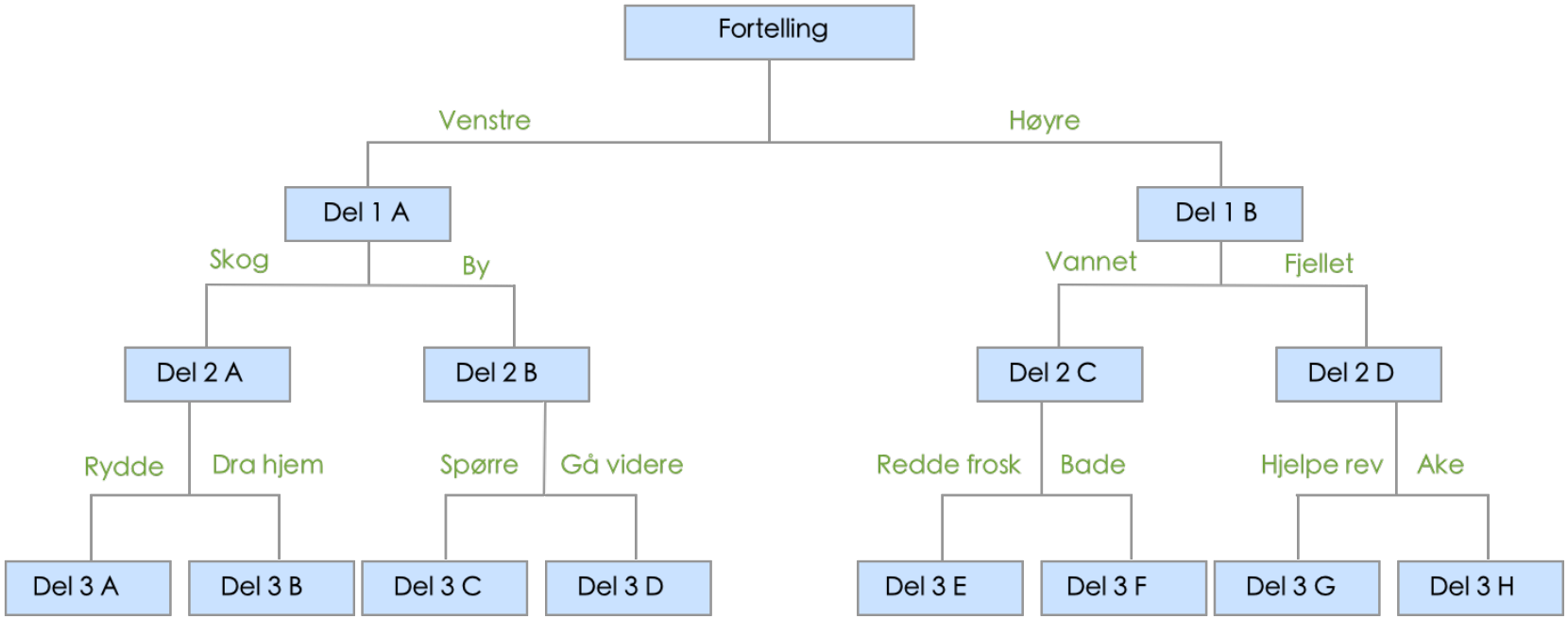
1. Barna begynner gjerne å snakke litt før mikrofonen har skrudd seg på - da skjønner ikke stemmeagenten hva de sier fordi den bare får med seg litt, eller ingenting av det som blir sagt.
2. Barna snakker ofte samtidig eller etter hverandre - gjør det vanskelig for stemmeagenten å tolke det som blir sagt

I en tutorial forsøker stemmeagenten å lære barna hvordan de skal prate med den, se side 5.

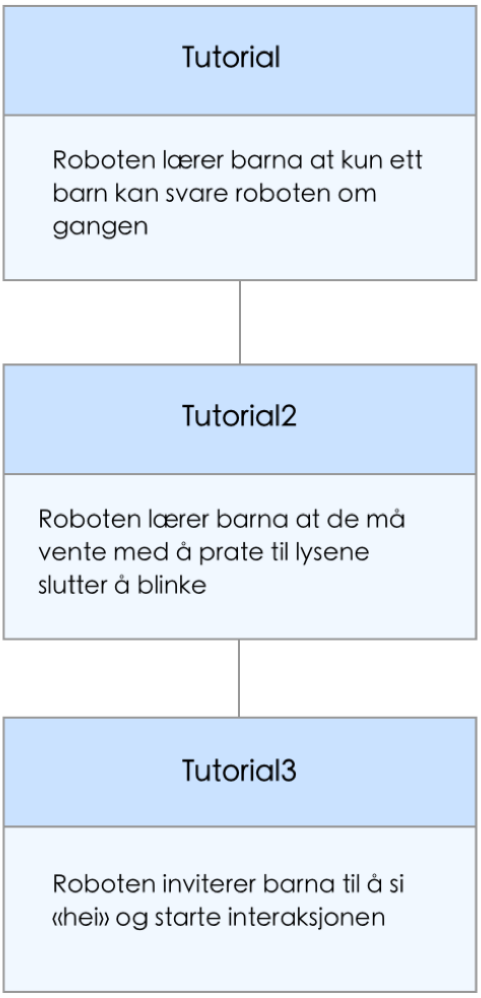
For å starte stemmeagenten, si:
«Hei, Google. Snakke med roboten vår»

PS! Legg gjerne trykk på «å» i «vår», da Google Home enda er litt dårlig på å tolke norsk uttale




Fortelling
Venstre
Høyre
Del 1 A
Del 1 B
Skog
By
Vannet
Fjellet
Del 2 A
Del 2 B
Del 2 C
Del 2 D
Rydde
Dra hjem
Spørre
Gå videre
Redde frosk
Bade
Hjelpe rev
Ake
Del 3 A
Del 3 B
Del 3 C
Del 3 D
Del 3 E
Del 3 F
Del 3 G
Del 3 H

Start ved å si:
«Hvordan snakker man med roboter?»



N Elicitation diary last evaluation

Tilbakemeldingsskjema

Gruppe nr:_____

Dato:_____ Tidspunkt:_____ Varighet:_____

Deltakere (barn):_____

Pedagog tilstede:_____

	Ja	Nei	Kommentar/utdyping
Opplevde du problemer med å komme inn i selve stemmeagenten? ("Hei Google, snakke med roboten vår").			

I hvilken grad vil du si du måtte hjelpe barna med å gjøre seg forstått?

Svært liten_____Svært stor

Kommenter svaret ditt med stikkord:

I hvilken grad vil du si barna forstod det stemmeagenten sa?

Svært liten_____Svært stor

Kommenter svaret ditt med stikkord:

I hvilken grad opplevde du at stemmeagenten forstod det barna sa?

Svært liten_____Svært stor

Kommenter svaret ditt med stikkord:

Hvordan opplevde du barnas holdning til stemmeagenten?

Negativ_____Positiv

Kommenter svaret ditt med stikkord:

Oppdrag

Spørsmål	Ja	Nei	Kommentar/utdyping
Fikk dere gjennomført oppdraget?			Hvis nei: Hva var årsaken?
Fulgte barna instruksjonen som ble gitt av stemmeagenten uten din oppfordring?			
Fikk barna med seg oppfordringen om å snakke sammen etter de var ferdige med å hoppe?			
Hvordan prøvde du å løse eventuelle problemer som oppstod i interaksjonen?			
Andre kommentarer/ observasjoner/ opplevelser			

Spørrespill

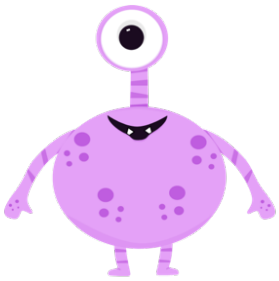
Spørsmål	Ja	Nei	Kommentar/utdyping
Fikk dere gjennomført spørrespillet?			Hvis nei, hva var årsaken?
Opplevde dere problemer med å gjøre dere forstått når dere skulle svare på spørsmålet?			Hvis ja: Hva oppfattet du som grunnen til at den feilet?
Fikk barna med seg oppfordringen om å snakke sammen etter at de fikk resultatet (riktig/feil)?			
Hvordan prøvde du å løse eventuelle problemer som oppstod i interaksjonen?			
Andre kommentarer/ observasjoner/ opplevelser			

Fortelling

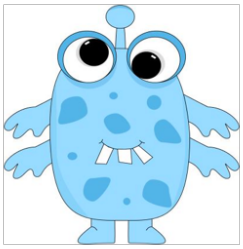
Spørsmål	Ja	Nei	Kommentar/utdyping
Fikk dere gjennomført hele fortellingen?			Hvis nei, Hva var årsaken? Hvor langt kom dere?
Opplevde barna problemer med å gjøre seg forstått når de skulle svare stemmeagenten?			Hvis ja: Hva oppfattet du som grunnen til at den feilet?
Fikk barna med seg oppfordringen om å snakke sammen på slutten av fortellingen?			
Hvordan prøvde du å løse eventuelle problemer som oppstod i interaksjonen?			
Andre kommentarer/ observasjoner/ opplevelser			

O The Monster Game

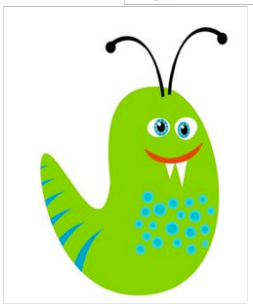
Jeg synes roboten er morsom



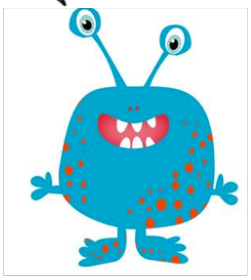
Jeg synes ikke roboten er morsom



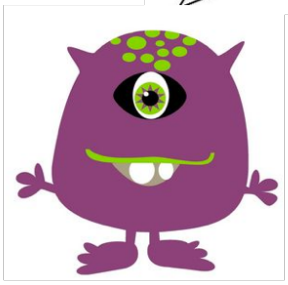
Jeg synes roboten bruker vanskelige ord



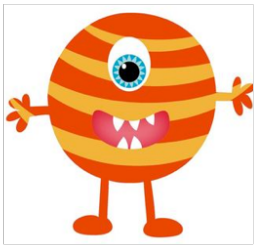
Jeg synes roboten brukte lette ord

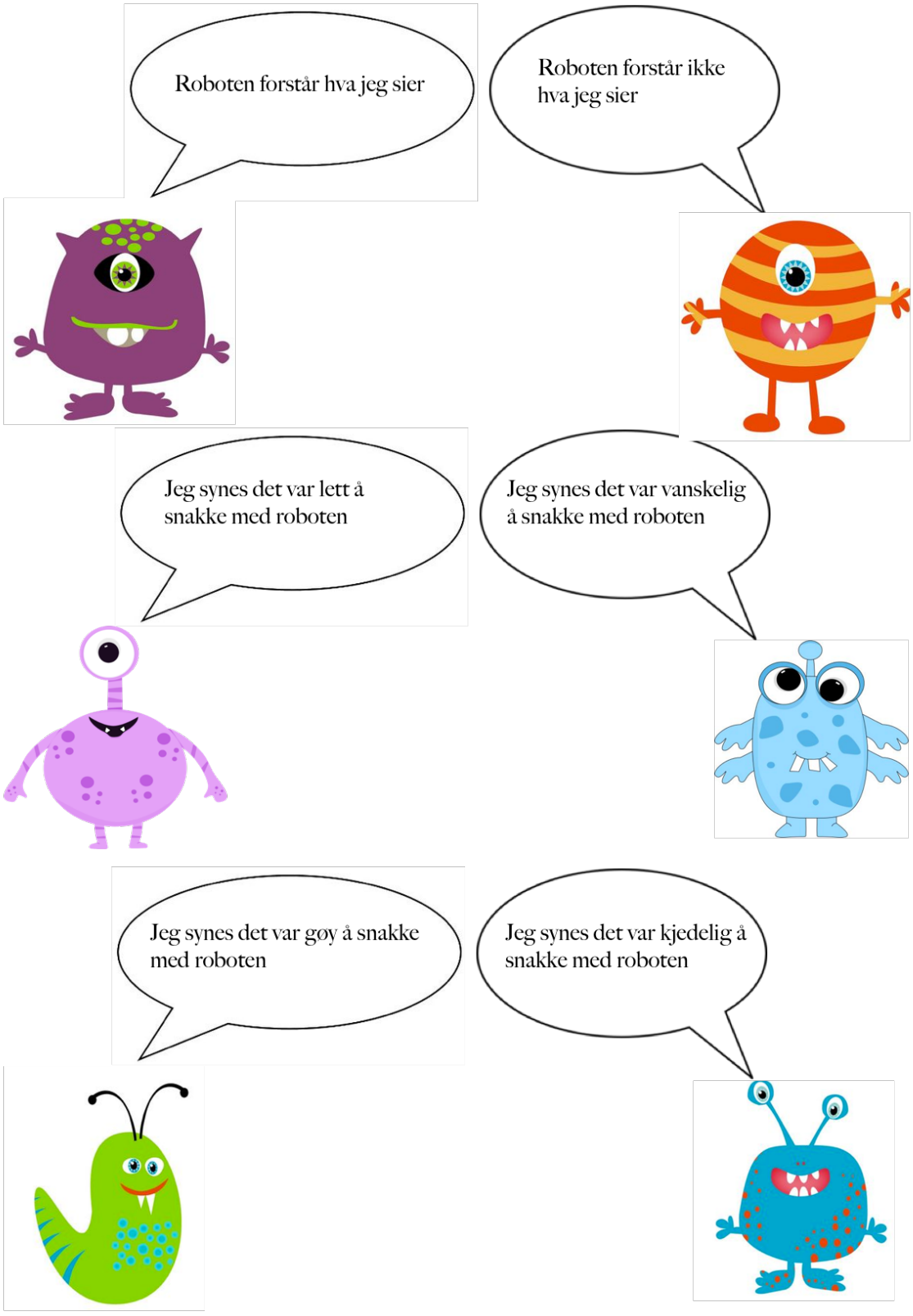


Jeg forstår hva roboten sier



Jeg forstår ikke hva roboten sier

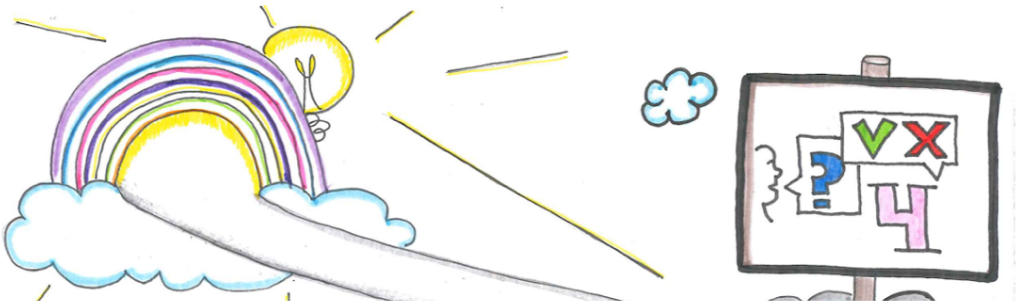




6

[6] Illustrations collected from: https://wikiclipart.com/monster-clipart_10310/, https://wikiclipart.com/monster-clipart_10297/ and https://wikiclipart.com/monster-clipart_10273/

P “The Story of Opinions”



Gruppe nr._____

Dette er en historie om da vi skulle snakke med en (kul/teit/morsom/rar/dum/smart/luftig) _____ snakke-robot. Snakke-roboten forstod det vi sa (ofte/noen ganger/aldri) _____. Da ble vi (irriterte/sure/sinte/brydde oss ikke) _____. Når den forstod oss ble vi (glade/overrasket/brydde oss ikke)_____). Vi synes selv at vi (ikke hørte etter/hørte etter) _____ når snakke-roboten pratet.

Vi gjorde et oppdrag og vi synes det var veldig (morsomt/kjedelig) _____ . Vi synes at vi lærte (masse/litt/ingenting) _____ om at vi trenger luft når vi er i aktivitet. Vil (ville/ville ikke) _____ gjøre et nytt oppdrag en annen gang!

Vi lagde en (kul/teit/kjedelig/morsom/spennende) _____ fortelling sammen med snakke-roboten. Vi synes dette var (kjedelig/morsomt/spennende) _____ . Vi synes selv at vi lærte (masse/litt/ingenting) _____ om naturen/lufta. Vi vil (vil/vil ikke) _____ lage en ny historie en annen gang!

Vi spilte også et _____ (kjedelig/rart/morsomt) spørrespill med snakke-roboten. Spørsmålene var ganske _____ (lette/vanskelige) å svare på. Vi var enige om at vi _____ (har lyst til/har ikke lyst til) spille spørrespill med roboten igjen!