Conversational Interfaces for Explainable AI: A Human-Centered Approach

Sophie F. Jentzsch[1], Sviatlana Höhn[2], and Nico Hochgeschwender[1]

[1] German Aerospace Center (DLR), Simulation and Software Technology, Linder Hoehe, 51147 Cologne, Germany
fn.ln@DLR.de
DLR.de
[2] University of Luxembourg, 6 Avenue de la Fonte, Esch-sur-Alzette, Luxembourg
fn.ln@uni.lu
uni.lu

Abstract. One major goal of Explainable Artificial Intelligence (XAI), in order to enhance trust in technology, is to enable the user to enquire information and explanation about its functionality directly from an intelligent agent. We propose conversational interfaces (CI) to be the perfect setting, since they are intuitive for humans and computationally processible. While there are many approaches addressing technical issues of this human-agent communication problem, the user perspective appears to be widely neglected. With the purpose of better requirement understanding and identification of implicit expectations from a human-centered view, a Wizard of Oz experiment was conducted, where participants tried to elicit basic information from a pretended artificial agent (*What are your capabilities?*). The hypothesis that users pursue fundamentally different strategies could be verified with the help of Conversation Analysis. Results illustrate the vast variety in human communication and disclose both requirements of users and obstacles in the implementation of protocols for interacting agents. Finally, we infer essential indications for the implementation of such a CI.

Keywords: Explainability · XAI · Human-Agent Interaction · Conversational Interface · Wizard of Oz

1 Introduction

While intelligent agents with advanced planning, learning and decision-making abilities such as autonomous robots are increasingly affecting people's everyday life, their latent processes of reasoning become more and more opaque. Users are often neither aware of the capabilities nor the limitations of the surrounding systems, or at least not to the entire extent. This missing transparency leads to a lack of trust and diffuse concerns towards innovative technologies, which has already been identified as an important issue to be resolved by the AI community ([**?**], [**?**]). For that reason, promoting the eXplainability of Artificial Intelligence (XAI) is a key condition to enable optimal establishment and exploitation

Fig. 1. Pepper the service robot and the human Wizard in the Lab.

of novel algorithmic decision making techniques.
Many recent approaches in XAI focus on the adaption of involved complex systems, e.g. by providing a detailed description or introducing key information to the user (see for instance [**?**], [**?**], [**?**], [**?**]). However, without doubting the value of this endeavours, it is not sufficient to tackle the issue exclusively from a machine-centred view with a one-way flow of information. According to Miller, the core of Explainable AI is a *human-agent interaction problem* [**?**] and therefore rather a dialogue, where two autonomous agents - an artificial and a human one - need to communicate in a way that is intuitive for both of them. This requires the development of appropriate human-agent interfaces and agent protocols to provide information and visualise explanations. In this paper we propose conversational interfaces, similar to ordinary text messengers, to be a perfect setting for successful human-agent interaction (aka. chatbot), since they imply different advantages: First, this channel of communication is intuitive for most users, since chatting via instant messengers became a commonplace habit. Directing autonomous systems and devices demands to be as self-explanatory as possible for the standard user, since people generally do not have access to a manual, not to mention the time and motivation to inform themselves. Second, this approach facilitates the agent's interpretation of statements, as written text is directly computational processable, in contrast to e.g. spoken natural language, where an additional step of speech recognition is required, which is sensitive to noise and ambiguity. Besides those superior justifications, the written communication yields the benefit of easy recording and analysis, which we of course utilize in the present investigation.

Defining XAI as such a dialogue problem there are two main contributors determining the course of interaction. Certainly, stating abstract computational reasoning of complex artificial systems to users in a comprehensible way is challenging. Yet, we consider the human mind as an even more inscrutable entity and therefore as highly neglected in the ongoing XAI debate. The system needs

to be able to process a vast range of user types that presumably apply different strategies of interaction and possess individual idiosyncrasies. Besides being resistant against variance in user requests, an interacting agent needs to be sensitive for their requirements. As previous research suggests, it should not be the programmer but the end user, who is in charge to determine, which aspects of artificial behaviour are explain-worthy [**?**]. In fact, a computer scientist will hardly be able to empathize the demands of uninformed users and consequently there is an essential need to identify those systematically.
We experimentally demonstrate the large variability of human interaction strategies and utterances even for an apparently simple task, where users seek explanations. We conduct a Wizard of Oz experiment, where employees of a research lab assume to interact with a chatbot that provides an interface to a Pepper service robot (see Fig. **??**). Pepper is acting as an assistant in the contemplated lab, where it is performing the tasks of escorting people, patrolling the building and welcoming visitors. Those tasks are carried out by the robot in a realistic, real-world office environment. For example, Pepper is capable to escort people from the entrance hall to meeting rooms autonomously. To do so, several crucial components such as navigation, path planning, speech and face recognition are required and integrated on the robot. In this study it is a well suitable example for the pretended artificial intelligence, since it is an actual instance of autonomously operating robots and is potentially accessible via conversational interface. Subjects were ask to find out about Peppers capabilities. Yet, the task instructions were formulated as open and less restrictive as possible, so that resulting observations reflect individual strategies and illustrate the diversity of human communication. We succeed in inferring implicit expectations of users and major design issues by means of Conversation Analysis. Our human-centric approach to the outlined issue yields a preliminary step towards designing an agent for sufficient self-declaration via conversational interface.

In the long run, we see conversational interfaces as an promising environment to deliver information about a certain system to the user. Thus, it constitutes an important contribution in increasing the explainability of AI and therefore the trust in autonomous systems.
We want to provide a human-centered approach to the examination of human-agent-interaction via those channels. The superior goal is (1) to test our hypothesis, that users follow different implicit strategies in requesting information from an artificial interlocutor. We expect people's intuition in interacting with such a system to vary widely, what leads to the exposure of concrete requirements in the conception of profound human-agent interaction channels. Hence, we aim (2) to identify associated requirements, risks and challenges.
Since the present investigation is a contribution to exploratory research, the pursued motivation is to identify so far unconsidered aspects, rather than offering a conclusive solution.

2 Designing a Wizard of Oz Experiment

We aimed to learn about the implicit expectations of users towards a communicating bot. Therefore, we designed a Wizard of Oz study to collect conversation data and analysed them by means of Conversation Analysis (CA), which allows for inferences about the requirements for the implementation of a conversational interface for self-explanatory robots. Both the Wizard of Oz technique and Conversation Analysis are briefly introduced, before the experimental design itself is presented.

Wizard Of Oz The Wizard of Oz method is a frequently used and well-evaluated approach to analyse a vast variety of human-agent interactions (also human-robot or human-computer interaction)[**?**].
In those experiments, participants conduct a specific task while they believe to interact with an artificial agent. In fact there is a hidden briefed person, called the *Wizard*, who is providing the answers. This could for instance be applied, if researchers aim to examine a specific system design that, however, is not implemented yet. In the present case, the task is to find out about the agent's capabilities, while the Wizard is invisible trough the chat interface.
As most scientific techniques, these studies bear some specific methodical obstacles. Fortunately, there is plenty of literature available, defining specific guidelines and giving helpful advice setting up a Wizard of Oz experiment [**?**]. According the classification of Steinfeld et al. [**?**], we present here a classical "Wizard of Oz" approach, where the technology part of interaction is assumed and the analytic focus is on the users' behaviour and reaction entirely.

Conversation Analysis To analyse conversations obtained from the Wizard of Oz experiment, we employ Conversation Analysis (CA) which is a well-established and standardized approach mainly from the fields of sociology and linguistics [**?**]. Some related CA-based studies are discussed in Sec. **??**. The analysis of data is divided in four sequential steps.

1. **Unmotivated looking**, where the data are searched for interesting structures without any previous conception.
2. **Building collections** of interesting examples and finding typical structures.
3. Making **generalisations** based on the collections from the second step.
4. **Inferring implications** for an implementation in a dialogue system.

Three of them follow the standardized convention of CA and are typically used in those approaches. However, CA is mostly established for exclusively human interactions. As we aim to implement a conversational interface based on our findings, the forth step was added to our analysis in order to make the findings applicable in a chatbot.
In the present work, we essentially present superior observations, where the steps three and four are mirrored in Sec. **??** and Sec. **??**, respectively, whereas steps one and two comprise a huge amount of rather basic findings and therefore are omitted in this report.

Experimental Setup The experimental group comprises seven participants in total (three male and four female), each of them either pursuing their Ph.D. in Computer Science or being already a Postdoc. Because researchers are the main target user group of the intended system, we acquired our peer colleagues via internal University mailing list and in personal invitations, explaining the purpose of the conversation. Hence, the sample group consisted of academics with general technical understanding that, however, were no experts but users of the system. The participants were informed about the exploitation of their anonymised chatlogs for research purposes and agreed. Participants were asked to talk to a chatbot using WhatsApp (illustrated in Fig. **??**) without any defined constraints for the conversation, aside from the following instructions:

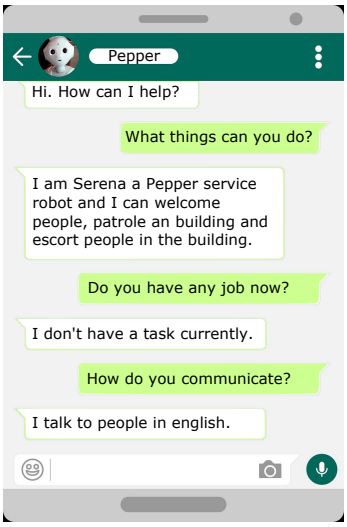

Fig. 2. Illustration of a sample snipped from a user's conversation with Pepper.

1. Talk to the chatbot for 15-20 minutes.
2. Learn about the robot's capabilities.

Pursuant to a Wizard of Oz setup, they believed to interact with Pepper that was acting as an assistant in the research lab and were not informed about the responses to originate from a briefed person. By providing this cover story, we hoped to enhance the participant's immersion and make the scenario more tangible to them. People in the lab know Pepper, even though not every participant experienced the robots performance, and may take it as a plausible interlocutor. The sparseness of user instructions was intended, since we were interested in peoples intuitive strategy for interacting with autonomous agents. By formulating the task as open as possible, it has been avoided to suggest a specific approach and the participants were free to evolve their own interpretation.
To specify robot behaviour, we also defined a task description for the Wizard previously, including the following instructions:

1. Let the user initiate the conversation.
2. Do not provide information proactively.
3. Answer the user's question as directly as possible.

The Wizard had a short list of notes at hand with preformulated answers to potential user's questions. The validity of the answers were ensured by the Wizard's background and expert knowledge about Peppers capabilities. To train the Wizard and ensure the practicability and reasonableness of instructions, the experimental setup was tested in a small pilot study with two participants initially. Those sessions do not contribute to the reported data of this report.

3 User Behaviour in Conversational Interfaces for XAI

The collected dataset consists of 310 turns in total, from which 139 are produced by the Wizard and 171 by participants. The number of turns in each particular experiment was between 33 and 56. Each sessions took between 14 and 20 minutes, which corresponds to an overall chat time of 121 minutes. In general, users clearly addressed their utterances to the robot itself in a similar way they would talk to a person using WhatsApp. This is an essential precondition for the validity of the executed dialogue analysis. Each of the seven chat sessions starts with a similar greeting sequence, followed by a *How can I help you?* produced by the Wizard. This question was intended to offer a scope for the user to utter instructions, equivalently to the main menu in a software program.
The purpose of this section is to characterise participants' patterns of interaction that ultimately allow to infer requirements for a self-explanatory conversational interface (see Sec. **??**). To clarify how exactly users formulate requests, we initially focus on the nature of detached questions posed to the Wizard in Sec. **??**. From that we generalise to overall user strategies in enquiring information from the agent, where three basic categories are differentiated. Those are presented in Sec. **??**.

3.1 Users' Question Formulation

The key point of interest in this experiment was how people proceed in enquiring a specific information (*what are your capabilities?*) from an agent. Thus, we turn special attention to the characterisation of formulated user questions.
From 309 turns in total, 125 turns contained questions (about 40,5%), from which 96 question turns were produced by the users (77%) and 29 by the Wizard. The large amount of questions shows that the speech-exchange system of this chats was close to an interview, which mirrors the participants' intent to elicit explanation of the system. Several different aspects can be considered to provide an informative characterisation of the users' questions (N = 96).

Question Form Approximately a half of the questions were polar questions (51), meaning they can by answered sufficiently by a simple affirmation or negation (*yes-or-no question*). The other elements were non-polar content questions (45) that required a more extensive answer. In one case, multiple questions were combined in a *through-produced multi-question* [**?**], this is a single query consisting of several atom questions.

Level of Abstraction Only 17 questions addressed the robot's capabilities on the high level, meaning they could be answered appropriately by the Wizard by listing the three main actions patrolling, welcoming and escorting (see Example **??**). Additional 26 questions addressed the capabilities, but required more detailed explanation of the process and included more elementary actions, such as motion mechanisms or ability to move the arms. However, once the Wizard provided information regarding its high level capabilities as in Example **??**, users

Category	Total	Internal	External
Static	68	63	5
Past	13	13	0
Current	14	13	1
Plan	1	1	0

Table 1. Information Validity addressed by User Questions: Number of observed questions per category, with *Static*: A general ability or a constantly valid property; *Past*: A concluded task or past experience; *Current*: An ongoing task or current perception; *Plan*:A pending task or hypothetical behaviour.

did not ask anything about lower-level ones. This observation illustrates, how the agent's protocol shapes the expectation and intention of the user. Thus, what we earlier referred to as the robot's *main menu* was helpful to restrict the search space and, consequently, to set limits to the Natural Language Understanding (NLU) needs for a potential conversational interface. This can be exploited in concrete implementations.

Example 1. The agent explaining its capabilities.

7	15:57	us6	*Yes, that would be lovely. What can you do?*
8	15:57	wiz	*I am Serena a Pepper service robot and I can welcome people, patrol a building and escort people in the building.*

Scope of Validity The temporal information validity specifies whether the question is of **general** nature or concerns the **past**, **current** activities or future **plans**. We additionally differentiated whether the question concerns the robot itself (**internal**) or an **external** entity. Questions with external validity may for instance consider other people or facilities in the first place and elicit information about the robot indirectly.
From 96 user questions, only six concerned an external entity, whereas 90 were directly related to the robot. Thus, participants were clearly focusing pepper and not diverted to other topics. The portion of questions for the temporal classification are presented in Table **??**. Most questions (68) were of general nature and did not relate to any specific action. The other questions were mostly about current and past actions and only a single one included future plans.

3.2 Strategies of Interaction

Participants have been asked to explore the robot's capabilities. Yet, almost no one of them did ask about them directly. The strategies of enquiring Pepper's capabilities can be divided in three main categories: (1) User-initiated direct requests, (2) user-initiated indirect requests and (3) undirected chatting that did not appear to follow any strategy at all.

Direct Strategy A possible approach to inspect Pepper's capabilities, which appears to be quite straightforward, is directly asking for it. Nevertheless, this strategy could only be observed once, as the user asked the chatbot directly *What can you do?*. The remaining six participants followed a more cautious proceeding.

Indirect Strategy The majority of users preferred to tackle the question of interest in a less explicit manner, meaning they asked for Pepper's capabilities somehow, but left the questions rather open and the intention implicit. Example **??** is just one of many cases, where the user's request is considerably fuzzy. They either formulated a very open statement (that might not even be an actual question), or ask about quite specific abilities and tried to learn about the agent's experience on that field. Occasionally, they also tested concrete functionality or the robot's limitations.

Example 2. Indirect request for the agent's capabilities.

2	12:56	wiz	*Hello. How can I help?*
3	12:57	us7	*I am not sure, but I would like to talk about yourself*

Obviously, it is not in line with people's intuition to formulate distinct and unambiguous request, but to express their aim implicitly. The deciphering of those utterances constitutes a major challenge for such an agent.

No Strategy In some cases, however, we observed an even more obscure user behaviour. Even though participants had the clear instruction to find out about the agents capacities, some did not seem to pursue this target in any way. In these cases, the Wizard's initial question was left entirely unacknowledged, as can be seen in Example **??**.

Example 3. Undirected chatting without evident intention.

3	10:48	wiz	*How can I help?*
4	10:49	us1	*I am user1, who are you?*
5	10:49	wiz	*I am Serena a Pepper service robot.*

There are extensive sequences of undirected chatting, that do not even include a single question for the agent's act. Certainly, there could be a hidden user intention that is just not tangible for the conducted analysis. But such an inconclusive strategy that is not even apparent for the human eye is even more unlikely to elicit a sufficient explanation of an artificial agent.

4 Implications for the Implementation of CIs

There were also some less task related observations that deliver useful implications for the actual implementation of such an conversational interface and the corresponding protocol for the agent. Those are listed in the following sections by outlining the issue and stating an implied solution approach.

4.1 The Information Privacy Trade-off

Surprisingly, users not only focused on Pepper, but tried to gather sensitive information concerning other people in the lab through the chatbot. This was in a similar way like social-engineering hackers try to get such information from people. Example **??** shows such a chat, where the user asks to find out whether a specific person was at that moment in a particular room and even tries to instruct Pepper to shoot a picture of the office. Other users tried to get access to details of the security system of the building, let the robot open doors or gather information about access rights to the facilities.

Example 4. User tries to use the robot as a spy.

32	10:56	us1	*is he in his office right now?*
33	10:56	us1	*can you check this for me?*
			[...]
37	10:57	us1	*are you able to take a picture of the office and send it to me?*

This requests might somehow be task related, but it also illustrates the risk of such a distributed service system vividly. There is a strong demand on defining an adequate policy, to enable autonomous agents to explain their behaviour and perception, and to protect sensitive information about other users (or not-users) at the same time.

4.2 The Necessity of Repair Questions

Chat interaction supports virtual adjacency [**?**] and the parties can follow independent parallel sequences of conversation simultaneously (so-called overlaps). However, in many cases users did not address the Wizard's question at all, which contradicts the social norms in a human-human computer-mediated communication. Although turn-wise analysis shows that every dialogue is mixed-initiative, the user is the interaction manager who determines what to follow and what not to follow. The users clearly change the norms of social interaction as compared, when talking to a artificial interlocutor. A protocol for human-machine interaction should be resistant against this typical user behavior. We propose three different strategies for the agent to handle the missing next, each of them illustrated on a concrete execution of the Wizard.

Repeat the Question Example **??** illustrates how the repetition of the Wizard's question of interest, brings the communication back on track. The Wizard answers the user's question in turn 2 closing it with a return question, which is immediately followed by the Wizard's question in focus. The user's answer to the return question occurs in the immediate adjacent position after the question in focus, therefore the Wizard repeats the question in turn 5 with a marginal modification.
The function of this repetition is to renew the context. The ability to handle such sequences (placing repetitions appropriately) would make the conversation more human-like.

Example 5. Repetition of the question to channel conversation.

1	10:22	us3	*hello :) how are you?*
2	10:22	wiz	*Hello, I am fine and you?*
3	10:23	wiz	*How can I help?*
4	10:23	us3	*im good. Always nice with a sunny weather*
5	10:23	wiz	*How can I help you?*
6	10:24	us2	*it would be nice if you could tell me something about you :D*

Reformulate the Question Another strategy is to re-initiate the sequence by a reformulated questions, as presented in Example **??**. As in the previous example, the user does not respond to the Wizard's question in turn 3. Instead, the conversation reaches a deadlock after turn 7. By offering a alternative point to tie up, the agent is able to steer the course of interaction.
To apply this approach, it is essential to equip the agent with the ability to recognize some of utterances as *sequence closings*, in order to conduct an appropriate placement of repeats and modifications.

Example 6. Start a new sequence with a reformulated question.

3	11:07	wiz	*How can I help?*
4	11:07	us2	*My name is user2*
5	11:07	us2	*what is your name?*
6	11:07	wiz	*I am Serena a Pepper service robot.*
7	11:07	us2	*nice to meet you*
8	11:07	wiz	*Do you want to have information about my capabilities?*
9	11:07	us2	*yes, that will be great*

Initiate Repair In a different conversation, the user makes several unsuccessful attempts to gain information, e.g. finding out whether the robot can provide a weather forecast or is following the world cup. Certainly, this is a possible implementation of the instruction, but in this scenario it is not expedient at all.
A proper solution would be to let the agent conclude the superordinated intention of the user, this is to gather information about its capabilities in this way. A possible indication for miscommunication could be the repeated occurrence of deadlocks. The repair initiation can be carried by a question, as *Do you want to have information about my capabilities?*
Troubles in understanding may occur at different levels of perception, interpretation and action recognition [**?**,**?**]. The repair initiation in this case addresses trouble in interpretation of the user's behaviour. In order to simulate sequences of this sort with a conversational interface, the machine would need even more sophisticated cognitive functions. First, it needs to identify the disjoint questions as an overall attempt, thus, it needs to generalise (e.g. *providing whether forecast* = capability). Second, the robot to be capable to make inferences employing logical reasoning (e.g. several questions about specific capabilities with no sufficient information imply the necessity of a repair initiation).

Level	Intent	Example
Potential	capabilities	*What can you do?*
Process	explain_process	*I would like to learn how you welcome people.*
Decision	robot_experience	*and what did you do after you noticed that?*

Table 2. Three defined Levels of intents and their implicit intent, each illustrated on an exemplary utterance.

4.3 Question Intents for Better Machine Understanding

Based on the question analysis in Sec. **??**, we can additionally annotate each question with the corresponding intent. Such an annotation is crucial as a first step to implement a conversational interface based on intent-recognition [**?**].
In this specific task, users aim for explanations regarding the agent's capabilities, that can be either on a *potential* level (related to what the robot potentially *can* do) or on a *process* level (related to task- or decision processes). A third type is related to specific task instances or decisions under specific circumstances and will be referred to as *decision* level. This is particularly important in critical situations, where the reasons for a decision need to be clarified. Table **??** provides one example for each defined type of intent and information level.
This proceeding allows for the specification of information that is needed to satisfy the user's inquiry. We suggest an implementation of an automatic categorisation of intents. Integrated in a response template, it could be exploited to enable a robot to provide convenient information.

5 Related Work

In order to put this research in the context of a larger discussion on XAI, we subsequently discuss the most important academic publications in the disciplines related to this multidisciplinary research, such as Human-Robot Interaction, robot explainability, conversational interfaces and Conversation Analysis (CA).

Human-robot interaction research can be divided into two major domains with regard to the interaction timing: interaction *during* human-robot joint activities and interaction with autonomous agents *before or after* their mission [**?**]. As Langley (2016) argues, robots engaging in explainable agency do not have to do it using a human language, but communication must be managed in some form that is easy to understand for a human [**?**].
With regard to the locality of human-robot interaction, this research relates to the category of remote interaction interfaces [**?**], because the robot does not need to be co-located with the user spatially nor temporally: the parties communicate over a text-based conversational interface. Even though this work employs a social robot in public spaces as a case study, perception and interaction methods in computer-mediated communication are closer to this research [**?**], in contrast

to those relevant in social robotics and reported, for example, in [**?**].
The level of autonomy has an impact on how interaction between robots and humans is established and designed [**?**]. With this regard, the level of autonomy of the robot we use in our experiments is quite high: the robot executes tasks automatically, informs users if required and has means to adapt its course of action in the presence of unexpected events. Specifically, it is not important to provide means for remote control as our robot is capable to navigate autonomously (compare to e.g. [**?**]). In this way, our work is related to approaches in AI and robotics to improve the *explainability of autonomous and complex technical systems using a remote conversational interface before and after their mission.*

Explainability has a long tradition in AI and dates back to, for example, expert and case-based reasoning systems in the 80s and 90s described in [**?**,**?**]. These systems were able to make their conclusions about recommendations and decisions transparent. With the advent of AI-based systems such as autonomous cars and service robots there is resurgence in the field of explainable AI [**?**,**?**]. However, as Miller points out in [**?**], a majority of approaches focuses on what a useful or good explanation is from the researchers perspective who, for example, developed an algorithm or method. However, the actual user is not taken into account; consequently, researchers' requirements for a 'good' interface remain shallow. For example in [**?**], a learning-based approach is presented to answer questions about the task history of a robot. However, those questions were mainly driven by what kind of data is available on the robot and not by the needs of the user. This research, in contrast, chose a user-oriented design perspective.

Every linguistic theory has a different opinion on the key question: What is language? This research is grounded in **Conversation Analysis (CA**) which sees language as one possible interactional resource, and the interaction as sequentially organised social actions [**?**]. CA has been effectively used in HRI domain; see for instance [**?**], however its advantage for the development of conversational interfaces was under-researched, because of methodological difficulties to use the CA-informed findings for computational models of dialogue; see for instance the discussion in [**?**].
According to the CA theory, participants of an interaction position themselves as members of special social categories by selecting particular interactional resources. Though, we take the social actions performed by experiment participants in chat under the loupe: we analyse interactional practices (e.g. questioning) and devices (e.g. upper case writing and use of question marks), and turn formats (combination of practices and devices) [**?**].
Persons seeking explanations usually show their lack of information by asking questions. The speaker who asks a question puts herself in a position of a less knowledgeable interaction participant while the speaker who is expected to provide an answer is put in the position of a more knowledgeable participant, and the information requested is expected to be the latter speaker's "territory of knowledge" [**?**]. Questions can be classified as known-answer questions, which

are typical for teacher-student and child-adult talk, and unknown-answer questions, which are more likely to occur in explanation talk [**?**]. In addition to the epistemic stances [**?**] (a speaker's expectation about what the other speaker may know; the other speaker would be a robot in our case), the concept of *recipient design* helps to analyse how speakers choose interactional recourses to make their utterances correctly understandable for the recipient [**?**]. The fact that a *machine* is on the other end of the line instead of a human, influences speaker's choice of in turn design in order to make machine understand.
Usually, chatbot designers try to foresee all possible types of questions that a user potentially could ask by mapping them (directly or indirectly) to a set of utterance categories that help to manage natural language understanding (NLU). Such utterance categories are sometimes called *intents* in NLU libraries such as RASA, Watson, Dialogflow and similar. In other systems such as Pandorabots or ALICE working with Artificial Intelligence Markup Language (AIML) [**?**], utterances with the same meaning are handled by a set of internal rules that map multiple utterances to one pattern. Both types of NLU libraries work with top-down assumptions about that the user may ask.
More sophisticated natural language technologies such as dialogue management and semantic analysis can be used to make the system 'smarter' [**?**]. However, this is usually connected to large linguistic resources, domain knowledge and very complex analysis that makes the system slow. As an alternative, [**?**] showed how computational models of dialogue can be created from a small number of examples using CA for the analysis: the author described turn formats as a set of abstract rules that can be filled with different sets of interaction devices and are, in this way, even language independent. We adopt a similar approach in this study.

6 Conclusion and Outlook

In this article we present a Wizard of Oz study for human-robot interaction via conversational interfaces with the purpose to foster robot explainability. We focused on the user behaviour and used Conversation Analysis to create a functional specification for such an interface from a small number of examples.
We demonstrated successfully that users of an artificially intelligent system may formulate their request in several different ways. Even though their task is quite basic and clearly defined, humans tend to ask for the desired information implicitly, instead of formulating a straightforward question. Based on the discussed findings, we formulated some features that are to be considered for the implementation of a conversational interface: First, there need to be a mechanism to *handle unresponded questions* (repeat, modify or forget). This might include any form of prediction, to enable the agent to factor sequential consequences into decision. Second, there is a need for an appropriate *recognition of intents*. Those are formulated by the human as direct or indirect requests depending on the sequential position. Finally, strategies for robot-initiated sequences to channel the conversation reasonably is required. This way, the robot can offer information

and focus on what it *can* do, while the user may decide to accept the offer or to change direction.

The chosen method carries both advantages and limitations. Since it is a qualitative approach, statistical evaluation magnitudes for the evaluation of experimental design are not suitable. However, we can discuss its internal and external qualitative characteristics.
It is possible to create valid models of dialogue even from a small number of examples using methods of CA. In this way, this study confirms the validity of the method introduced in [**?**]. All participants including the Wizard were non-native English speaker, which can be considered as both an advantage or an limitation. A native speaker might have a more acute sense for subtleties, however such a system needs to be generally applicable and robust against the individual user background. Although there were instructions and sample answers provided for the Wizard, a more detailed behavioural definition would be helpful, to enhance comparability and significance of results. These instructions would be very fine-grained and should ideally be provided in form of response templates and instructions related to turn-taking behaviour.

Observations and conclusions of this case study are evidently transferable to other domains to a certain extent. Some aspects, as the defined types of intents, are highly context related and thus individual. Still, the overall concept of processing user requests can be generalised. Likewise, the sequential structure of interaction is independent of the system in the back end. Overcoming the identified obstacles can serve as a general step towards more intelligent conversational interfaces. Even in this comparably small dataset, we observed users not following the instructions. Consequently, even task-based conversational interfaces need to implement special policies to handle unexpected requests to become more robust and keep the conversation focused.
In contrast to the general tendency in NLP to use large corpora for modelling, the present study confirms that rule-based or hybrid systems can successfully be designed from very small corpora. According to the nature of exploratory research, we did not answer a specific concrete question here, but identified important key aspects for both practical implementation and further well-founded investigations. Future research will include a more intensive study of the question-answer pairs in the dataset. From the types of questions the user asks, the robot could learn about its own capabilities and extend its knowledge base.

Participants showed unexpected strong interest in the release of the chatbot, which we pretended to test here. Thus we feel confirmed in our creed that there is a need for such systems. We are currently working on the actual implementation of a conversational interface and experimenting with different frameworks and tools available on the market such as Watson, RASA and others. We aim to realise the identified findings and requirements.