Heuristic Rules for Human-Robot Interaction Based on Principles from Linguistics - Asking for Directions

Andrea Bauer and **Barbara Gonsior** and **Dirk Wollherr** and **Martin Buss** [1]

Abstract. Robots that are to assist humans in flexible and versatile ways, will not always possess all the information required to fulfill their task. Therefore robotic systems have to be able to retrieve information through natural language communication with humans. Natural language is often vague or ambiguous and thus hard to interpret by technical systems. To enable robotic systems to interpret natural language expressions correctly it is necessary to include findings from human-human communication into the dialog systems of robots. In this work the field of communication topics between human and robot is confined to the communication about space, more specifically to a robot asking a human for directions. This paper gives insight on theories from linguistics research focussing on asking for and giving directions. From these theories 10 heuristic rules for human-robot interaction are deduced, where 4 of the rules apply even to systems that are not able to communicate through natural language. Additionally a first experiment where a robot used the 4 basic heuristic rules to successfully ask passers-by for directions and find the way to an unknown goal location is presented.

1 Introduction

It is a goal of researchers around the world to design service robots that can be used by humans without technical training. This makes natural language communication robots a necessary ability for robotic systems. User-friendly robots should not only able to recognize speech and synthesize it, but also to understand the more complicated semantics. Therefore it is reasonable to include findings from linguistics in the form of heuristic rules for future dialog systems in robots.

Many examples of human-robot interaction (HRI) research exist where robots communicate with humans who have no experience with robotic systems, such as the museum guide robot Rhino [7] and its successor Minerva [23] which informed museum visitors about exhibitions. There are robots in shopping malls that relay useful precompiled information to humans, e.g. Toomas [11]. A communication robot for minor service tasks in hospitals, presented in [22], is intended to be used by people with little or no technical experience.

It can be assumed that systems which are to assist humans in a flexible and cognitive way will not always possess all the information required to fulfil their task. A central aspect of intelligent autonomous behavior is thus the ability to interact in order to retrieve information. As situations may often arise in which a robot will have to reach an unknown goal position our work focuses on HRI for asking for directions.

[1] Institute of Automatic Control Engineering (LSR), Technische Universität München, D-80290 Munich, Germany,
email: ab@tum.de, babsigonsior@gmx.de, dw@tum.de, mb@tum.de

Already there are robots communicating with humans about spatial information, such as [20, 12, 21]. A robot that retrieves spatial information from a grid map and translates it into linguistic spatial descriptions is presented in [20]. Human-augmented mapping [12] enables a robot to create a map of its environment by exploring it and asking a human to label the areas of interest for it, thus aligning a robotic map with human concepts. The robot Biron [21] integrates spoken dialog and visual localization to learn and label new places.

There are already some robots [18, 2, 17, 15] that ask humans for directions in simple settings and structured indoor environments. A wheelchair robot [18] can be given coarse qualitative route descriptions. The office robot Jijo-2[2] can learn the locations of offices and staff by moving around and asking humans for information. A robot asking for the way at a robotics conference is presented in [17]. A miniature robot that can find its way in a model town by asking for directions is described in [15].

As the ability to ask for directions seems to be a requirement in future robotic systems, we see the need in linguistic rules for dialog systems that will enable robots to ask humans for directions, reason with them about the given description and interpret the information correctly. Dialog systems in HRI are still mostly based on keywords, however as ”vagueness is one of the most salient, but also one of the most effective features of natural language” [13] there is more to understanding verbal information than to look up the lexical meaning of single words. Therefore it is necessary to include more complex rules in dialog systems. We propose 10 heuristic rules for HRI in an asking for directions situation.

The remainder of this paper is structured as follows. In Sec. 2 we give an overview of relevant theories from linguistics dealing with the problem of asking for directions. From these theories heuristic rules are derived and introduced in Sec. 3. An example of a robot following 4 of the proposed rules while interacting to retrieve direction information is presented in Sec. 4. Finally Sec. 5 provides concluding remarks.

2 Theoretical Background from Linguistics

Linguistics theories deal with the structures of complex verbal actions within human-human communication and the occurring psychological processes. Principles for reasoning about space can be found among them including the analysis of asking-for-directions dialogs and the different semantic meanings of verbal expressions or gestures. Theories relevant to the problem of asking for directions are the analysis of dialog structures and the complex deixis theory founded by Bühler [6] in 1934.

2.1 Dialog Structure

Wunderlich [24] analyzed asking-for-directions dialogs and found a common structure of four consecutive phases:

I *Introduction*: The asker addresses a respondent and defines the task, i.e. giving directions to a specified goal location, possibly defining the mode of transportation or other individual requirements.

II *Giving directions*: The respondent provides the necessary information by means of natural language and gestures, sometimes additionally with the help of a sketch.

III *Confirmation*: Either of the two partners confirms the information. In this phase further inquiries can be made.

IV *Conclusion*: The asker thanks the respondent and they part.

This schematic structure is very flexible, i.e. some phases may be interchanged or recur. Nevertheless it is a well-proven guideline for human-human interaction reflecting the intrinsic cognitive processes involved. One of these cognitive processes for the respondent is planning the description by building a cognitive map which is based on individual experiences. This cognitive map includes ”objects which are salient landmarks for nearly everybody” [14]. The respondent has to complete the task of separating these salient landmarks from individual experiences and present them to the asker. If this is achieved by using appropriate means of communication the asker will be able to build a corresponding cognitive map that represents the route to the goal location, structured by landmarks.

Spatial information in general and route information in specific is communicated using deictics, or deictic words, which are analyzed by deixis theory.

2.2 Deixis Theory

Deixis theory [6] is based on the assumption that communication acts can be assigned to two different fields of language, namely the symbolic field, comprising nouns which are symbols independent of the context, and the deictic field, that includes deictics, which vary depending on the context. In natural language communication referring is generally managed by deictics which point to subspaces of the deictic field and are the verbal equivalents to pointing gestures. In order to point to subspaces within the deictic field both speaker and listener have to share a common ”deictic space”. At the beginning of a face-to-face route description this deictic space is given through the range of the visual perception. In the course of the conversation the route description usually leaves the shared visual perception range and a new deictic space is built by the geographical knowledge of both partners. Another deictic space can be introduced by involving a map or a sketch that represents the real geographic space.

Referring to subspaces in a certain deictic space by using deictics depends on particular contextual factors such as the position of the speaker and the direction of gaze [14], can be summarized as the speaker’s origo.

2.3 The Origo

Bühler introduces the term ”origo” [6] which is conceptually conceived as the origin of a ”coordinate system of subjective orientation”. It is derived from the need of a ”basic reference point” [14] in a given deictic space, which includes three deictic dimensions [1]: the *personal dimension* including personal pronouns of the first and second person, the *spatial dimension* including demonstratives, adverbs, movement verbs, and prepositions, and the *temporal dimension* including temporal expressions. Thus the origo is defined by the personal mark ’I’, the spatial mark ’here’ and the temporal mark ’now’. Accordingly the origo is the reference point for personal, spatial and temporal deixis, which relates to other elements within these three deictic dimensions. Thus deictics can only be interpreted considering the origo.

Deictics have certain functions within an asking-for-directions dialog depending on their deictic dimension.

I *Personal deictics* define the actors within communication.

II *Spatial deictics* indicate the directions from landmark to landmark. They can be accompanied by gestures.

III *Temporal deictics* refer to the time domains of actions which have to be carried out to reach from one landmark to another. They are indicated by movement verbs.

The functions of deictics can be refined further on the basis of contextual factors and relations, as described by Fillmore [9] and Lyons [16]. As these refinements are beyond the scope of this paper, they are not further explained here.

In some cases deictics can only be located by using an additional ”relatum”, an object which is related to the origo [10]. For instance if it is referred to a subspace relative to a landmark, the landmark functions as a relatum relative to the perspective given by the origo. However if it is referred to a subspace relative to the speaker’s body, then the origo, or the speaker, functions as a relatum. If no new landmark is introduced after a deictic in a route description, the relatum is the origo of the previous landmark.

Problems and difficulties in interpreting deictics can occur, as deictics can be interpreted differently according to the deictic dimension, the context or whether or not they are used in combination with a relatum.

2.4 Problems in Interpreting Deictics

Several problems and difficulties that may occur while interpreting deictics and identifying the subspaces they refer to, have been identified by Klein [14].

I) Coordination problem: In a dialog situation all participants have their own origines. Since the origo is usually defined by the position and orientation of the current speaker, the listener must project that origo into her own system of orientation. As soon as the roles of speaker and listener are changed, the origo of the new speaker becomes essential and the other person has to adapt to it. Sometimes it is not clear whether the origo in a route description is the one of the speaker or the one of the listener. In the case of an asking-for-directions dialog the origines of the speaker and the listener are the same in terms of the position, as ’here’ encloses both speaker and listener. However, the orientations of the two communication partners differ and have to be coordinated, so as to avoid ambiguities of deictics.

II) Problem of the shifted origo: In the course of the asking-for-directions dialog both speaker and listener shift their origines into the perspective of an ”imaginary walker” [14] representing the addressee on her way projected into the future. A place that would be normally referred to as ’there’ may be called ’here’ within the route description, e.g. ’go straight until you see the park, here you need to turn left’. Thus the deictic ’here’ does not necessarily refer to the actual location but to the position of the shifted origo and must be interpreted depending on the context.

III) Problem with the use of an analogon: When humans use a sketch or a map to illustrate the described route they introduce a new deictic space. Consequently there are two deictic spaces involved, the map and the real geographic space represented by the map. The map functions as an analogon, where pointing to an element within it represents pointing to an element in the real space. Problems arise when the assignment of the two deictic spaces is not clear.

IV) Delimitation problem The subspace of the deictic space that a deictic points to can not be fully identified just by coordinating the origines of the dialog partners. The extend of a subspace is often vague and depends on context and environment, where the borders of the subspace must be established by gestures, verbal explanations or factual knowledge. For example, 'here' can be characterized as a subspace of the deictic space including the origo.

V) Problem of deictic oppositions: Deictic oppositions follow a certain system that, depending on the language can consist of two (e.g. in English: 'here' - 'there'), three (e.g. in German: 'hier' - 'da' - 'dort') or even more components. Examples of other deictic oppositions are the opposed meanings of 'left' and 'right' or 'this' and 'that'. Due to the fact that every communication partner has her own reference system there may emerge at least two different meanings for oppositional deictics. Their meaning depends on the context and additional information about the location in space and the delimitation or the interpretation of the shifted origo, where oppositional deictics may occur referring to subspaces in another deictic space, e.g. 'go straight on to the next intersection and turn right here'.

Even for humans it is sometimes hard to interpret deictics accurately, as any of these problems may occur in dialogs. For robots it is even harder to interpret deictics correctly, as they have to infer the right meaning from a series of single keywords retrieved by their speech recognition systems. In the following we propose guidelines in the form of heuristic rules that can be implemented in robotic systems to circumvent or solve the presented problems in HRI.

3 Heuristic Rules for Human-Robot Communication

Based on the theories from linguistics about human-human communication presented in the previous section we derive rules for the human-robot dialog of asking for directions. These rules should in hypothesis make the dialog more successful in terms of rendering it more natural and intuitive for the human partner, enabling the robot to interpret the possibly vague deictics correctly, and retrieve unambiguous route information that can be used by the robot to navigate.

Rules 1 to 4 are general rules for a robot asking for directions and can be applied by any interactive robot even if it is not equipped for natural language communication. Rules 5 to 10 are additional rules that apply to robots with natural-language-based dialog systems (text or speech). These rules provide guidelines that clarify and structure the vague deictics.

Heuristic Rule 1. *The dialog must be structured according to the four phases Introduction, Giving Directions, Confirmation, and Conclusion.*

This rule is a simple measure that adopts the structure from human-human communication to a human-robot dialog and therefore in hypothesis, should render the interaction more natural and intuitive for the human partners. Fig. 1 depicts the asking-for-directions dialog with the four phases, where the partner with the leading role is indicated in grey for each respective phase. It is the robot's task to

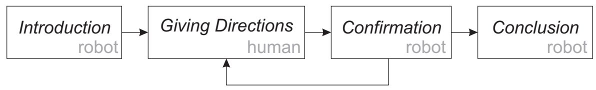

Figure 1. The structure of an asking-for-directions dialog. The partner with the leading role is indicated in grey in each phase.

initiate the *Introduction* phase, by addressing the human and introducing itself. A robot should introduce itself before asking for directions, as it is a strange novelty to humans in this situation. The *Introduction* phase is concluded by the robot when it asks for directions, where it has to specify how information can be given in accordance to the robot's abilities. The *Giving Directions* phase is initiated when the human begins to describe the route and it ends with the implicit or explicit declaration of the goal location. The *Confirmation* phase should be conducted by the robot which conveys to the human how the route description was interpreted. In this phase the robot should give feedback about the perceived information by displaying the perceived route in a bird's eye perspective including depictions of salient landmarks, solving the analogon problem presented in Section 2.4. This enables the human to see instantaneously that the robot has perceived the information correctly and if necessary to correct it. Lastly the *Conclusion* phase is initiated by the robot disclosing that no more information is needed. The robot should thank the human for the help and move on.

Heuristic Rule 2. *The robot must solve the coordination problem, by asking the human to specify the first direction with a pointing gesture.*

At the beginning of the *Giving Directions* phase it is not immediately clear whether directions are given relative to the origo, i.e. the personal reference system, of the human or the robot. As outlined in Fig. 2, a point of interest (PoI) is described as being located in opposite directions relative to the reference system of the human O_h or the robot O_r, rendering the direction information ambiguous.

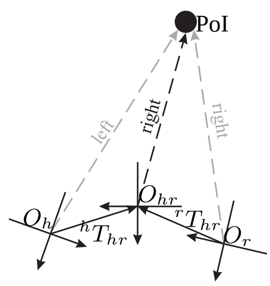

Figure 2. A PoI is referred to differently from the personal reference systems, or origines, of two dialog partners O_h and O_r. After coordinate transformations, ${}^rT_{hr}$ and ${}^hT_{hr}$ respectively, both partners use a common reference system O_{hr} to refer to a PoI in an unambiguous way.

To solve this problem the human is asked to point in the first direction of the route, as the gesture is unambiguous. At that point further directions are given relative to this direction. In mathematical terms, this pointing introduces a common reference system, O_{hr}, which is oriented towards the indicated direction, and specifies the necessary transformations, ${}^rT_{hr}$ and ${}^hT_{hr}$ of the reference systems O_h and O_r.

From then on it is understood that further directions are given relative to the same orientation. This rule solves the coordination problem, and is crucial for the success of the whole dialog.

Heuristic Rule 3. *The robot must break the perceived route description down into route segments and save the information as a route graph.*

The robot needs an internal representation of the route information to be able to reason about it and to navigate by it. Therefore the perceived route description must be broken down into route segments and stored as a route graph $G\,\langle\text{nodes, edges}\rangle$. Nodes $N_i(t_j)$ of the route graph represent landmarks of type t_j in the route description. Edges $E_i(a_i, N_k, N_l)$ represent actions a_i connecting the respective nodes, i.e. landmarks, N_k and N_l. An implementation of a system building a topological route graph from human direction information is presented in [3].

Heuristic Rule 4. *The robot must interpret the basic directions 'left', 'right', 'straight', and 'back'.*

The direction deictics 'left', 'right', 'straight', and 'back' establish the directions of actions starting at the last landmark. These deictics are the most important and basic ones in an asking-for-directions dialog. Thus the minimal requirement for a dialog system to interpret route directions is to be able to interpret these four deictics.

A possible definition for the deictics 'left', 'right', 'straight', and 'back' is to interpret every deictic as an angular range of $90°$, as depicted in Fig. 3. An example of this definition is given in [8]. The definition could be refined for example by including expressions like 'diagonally left'.

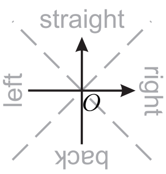

Figure 3. The deictics 'left', 'right', 'straight', and 'back' as angular ranges of $90°$.

Heuristic Rule 5. *The robot must identify landmarks in the route description.*

Landmarks structure the route description, provide starting and end points of route segments and they must be identified when they occur in the description. A node $N_i(t_j)$ must be inserted into the route graph for every landmark in the route description. There are four different ways to identify landmarks in a route description. Firstly, they can be identified through previous knowledge saved in the dictionary of the dialog system. Secondly, when the human points to a landmark it is recognized from gestural and visual clues. Thirdly, it can be identified from a non-deictic description of the landmark, e.g. 'the yellow building'. Fourthly, a landmark can be identified by its relationship to previous landmarks, as in 'the third intersection', where the next one or more landmarks are of the same type as the explicitly mentioned landmark. In this case the robot must insert the corresponding number of nodes and edges into the route graph.

The last landmark in the route description is the goal location, or if the route is very complex, it may be a preliminary goal.

Heuristic Rule 6. *The robot must interpret 'here' and 'there' depending on when they occur in the route description.*

If the deictics 'here' or 'there' accompany a pointing gesture at the beginning of a route description, these words define the first landmark, namely the spatial structure that the gesture points to.

If these deictics occur later on in the route description they are usually linked with landmarks, i.e. they either stand in the place of landmarks or accompany them. Thus they structure the route description just as landmarks do. When one of these deictics occurs a new route segment begins in the description, therefore a new edge $E_i(a_i, N_k, N_l)$ must be inserted into the route graph.

Heuristic Rule 7. *The robot must identify movement verbs.*

Movement verbs define the actions that connect the landmarks and must be identified by the robot. For every movement verb an edge $E_i(a_i, N_k, N_l)$ must be inserted into the route graph. All movement verbs can all be interpreted as the one neutral movement verb 'move'. In a route description a movement verb may even be omitted as it may be clear from the context.

Heuristic Rule 8. *Deictics must be classified according to their semantic attributes.*

The deictics input must be classified by the robot according to semantic attributes in order to simplify the interpretation of the expression. To this end the semantic attributes of deictics must be stored in the dictionary of the dialog system. Relevant semantic attributes of deictics are the *deictic dimension*, the *distance range*, the *delimitation*, and whether or not a *relatum* is needed. The deictic *dimension* is either *spatial*, *temporal*, or *personal* and denotes whether a deictic points to a region in space, in time, or to a person. The *distance* of a deictic denotes how far the pointed region is away from the origo. It can be sub-classified into *origo-inclusive* and *origo-exclusive*, where the latter can be *close* to the origo or *far* away from it. The *delimitation* expresses whether or not the region a deictic points to has clearly defined boundaries. The last attribute indicates whether a deictic needs a *relatum* or whether it can occur without one. Deictics which can not be classified into origo-inclusive or origo-exclusive express only an uncertain proximity or distance, and therefore require a relatum to define the location in the deictic space [10]. As a general rule all personal deictics and deictics which are delimited do not necessarily require a relatum. Table 1 gives an overview of the most common deictics with the relevant semantic attributes, it does not provide an exhaustive list of deictics, but gives the reader an idea of how to classify deictics.

As a consequence of saving the attributes the interpretation of deictics is facilitated. The deictic dimension reveals whether or not a deictic is part of a route description, in the case of spatial and temporal deictics, or not, in the case of personal deictics. The distance provides an approximation of the spatial location of the described feature, as does the delimitation. Finally a deictic that needs a relatum must be interpreted together with the relatum.

Heuristic Rule 9. *To simplify the interpretation of deictics, the temporal domain can be mapped to the spatial domain.*

In a route description the passing of time corresponds to traveling along the route. When giving directions the 3-dimensional space is projected to a simple (1-D) path, where landmarks confine route segments according to heuristic rule 3. The temporal domain is also 1-dimensional where landmarks confine certain periods of traveling the path. Thus time and space have a similar structure in a route description and can be mapped to one another and the number of expressions that have to be analyzed further is reduced. Temporal expressions in

Table 1. Common deictics with relevant semantic attributes.

DEICTIC	DIMENSION			DISTANCE			DELIMITATION		NEEDING RELATUM	
	spatial	temporal	personal	origo-inclusive	close	far	yes	no	yes	no
'here'	×			×				×		×
'there'	×				×	×		×		×
'near'	×				×			×	×	
'far from'	×					×		×	×	
'left'	×							×	×	
'right'	×							×	×	
'in front of'	×							×	×	
'behind'	×							×	×	
'this'	×				×		×			×
'that'	×					×	×			×
'now'		×		×				×		×
'then'		×			×	×		×	×	
'soon'		×			×			×		×
'later'		×				×		×		×
'I'			×	×			×			×
'you'			×		×	×	×	×		×

route descriptions occur only in present and future tense, the most common of these are listed below.

'now' $\longrightarrow$ 'here'
'then' $\longrightarrow$ 'there'
'soon' $\longrightarrow$ 'near'
'later' $\longrightarrow$ 'far'
t $\longrightarrow$ $x = \frac{v}{t}$, with walking velocity v

If a temporal deictic cannot be mapped to a spatial equivalent, it can still be interpreted as a start or end point of an action.

Note that 'then' following 'if' is not temporal but causal and must not be mapped to the spatial domain.

Heuristic Rule 10. *The delimitation problem must be solved by modeling the distance ranges of various expressions such as 'here', 'near', and 'far' depending on the environment.*

It is clear that the deictic 'here' denotes the position of the actual or shifted origo however it is unclear how far the region 'here' stretches. The deictics 'near' and 'far' denote regions in space that do not include the origo and therefore do not coincide with the region 'here', where it is clear that the first expression denotes a region closer to 'here' and the latter expression one that is farther away. The reaches of the spatial deictics, such as 'here', 'near', and 'far', depend on subjective perception, the personal viewpoint and on the context, such as the used mode of transportation and the extension of the described route. The probability density functions (pdf's) of the deictics 'here' $p_{\text{'here'}}(d)$, 'near' $p_{\text{'near'}}(d)$, and 'far' $p_{\text{'far'}}(d)$ are estimated based on the results of a user study, presented in [4]. Fig. 4 depicts those pdf's depending on the distance d in an urban environment. The pdf's are modeled as

$$p_{\text{'here'}}(d) = \frac{1}{\mu} e^{\frac{d}{\mu}} \text{ , with } \mu = 148.7 \text{ ,}$$

$$p_{\text{'near'}}(d) = \frac{1}{d\sigma\sqrt{2\pi}} e^{\frac{-(ln(d)-\mu)^2}{2\sigma^2}} \text{ , with } \mu = 5.8 \text{ , } \sigma = 0.51 \text{ ,}$$

$$p_{\text{'far'}}(d) = \alpha\,\beta\,d^{\beta-1}\,e^{-\alpha\,d^b} \text{ , with } \alpha = 952.1 \text{ , } \beta = 5.85 \text{ ,}$$

where $p_{\text{'here'}}(d)$ is an exponential function, $p_{\text{'near'}}(d)$ is a lognormal function, and $p_{\text{'far'}}(d)$ is a Weibull function.

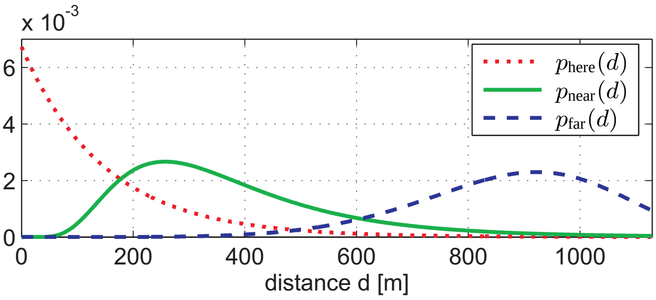

Figure 4. Estimated pdf's for the terms 'here', 'near', and 'far' over the distance d [m] from the origo in an urban environment.

The deduced heuristic rules present a guideline to researchers implementing the ability of asking for directions in a robotic system. Depending on the hardware and other constraints they may be altered or expanded to meet the requirements. The first 4 basic rules have been implemented on a robot and tested during a HRI experiment.

4 Experiment

In a first experiment previously described in [5] 4 of the proposed rules (heuristic rules 1 to 4) were included in the dialog system of a robot that had to ask passers-by for directions in order to reach an unknown goal location in an urban environment, without the use of GPS or previous map knowledge. Only the first 4 general heuristic rules were included as the robot was not equipped for natural language communication.

4.1 The Autonomous City Explorer

The *Autonomous City Explorer (ACE)* robot as depicted in Figure 5 was equipped with an active-stereo-camera head for human tracking and gesture recognition, an animated mouth, a loud speaker, and a touch screen for HRI, as well as a differential wheel mobile platform and laser range finders for navigation.

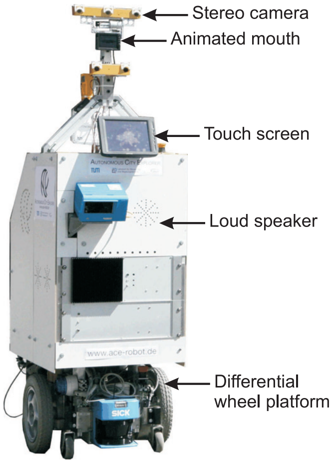

Figure 5. The robot *ACE* with principal hardware components.

The robot communicated with the human firstly by speaking over a loud speaker (English or German language) using MaryTTS [19] for speech synchronization. A mouth displayed on a small monitor was animated synchronously to the speech. Additionally the robot could present images and text to the human on a touch screen. As robustness to environmental disturbances such as noise was an important requirement for the system, the touch screen, was also used as the main means of input from the human. The robot was able to recognize and interpret pointing gestures, as described in [5].

4.2 Heuristic Rules Included in the HRI

As the completion of the task, i.e. to navigate to an unknown goal location in an urban environment, depends on a successful interaction, the design of the dialog was a crucial component and the first 4 of the proposed heuristic rules were included. The included rules are all general guidelines for asking for directions and not specific for natural language communication, therefore they could be applied to the robot that does not understand natural language but presents the human with options for answers.

Firstly when the robot addresses a human it introduces itself according to heuristic rule 1 and the dialog is structured in four phases, where the *Giving Directions* phase coincides with the *Confirmation* phase. The robot depicts every direction information on the touch screen whenever it is given. This allows the human to check whether the robot has interpreted the input accurately and if necessary to correct it. At the beginning of the *Giving Directions* phase the robot asks the human to point in the direction it has to go into first. So that at this point human and robot align their personal reference systems, i.e. origines, and it is understood that further directions are given relative to the common reference system. The human is then ask to give further directions through buttons on the touch screen, where buttons for the basic directions are provided, according to rule 4.

An exemplary dialog is given in the following:

ACE: ”Hello, my Name is ACE. I need to get to Marienplatz.”
Human confirms by touching screen
ACE: ”Please look into my eyes and point in the direction I should take now, as shown on the screen.”
Human points
ACE turns its head, takes an image of the indicated direction, and displays it on the touch screen
ACE: ”Please indicate the exact position in the image.”
Human touches the screen in the corresponding place
ACE: ”Please tell me how I should proceed then.”
Human gives direction by pressing buttons representing directions and indicates when the route instruction is complete
ACE: ”Thank you very much for your help.”
ACE: ”Please step back, I will go now.”

Following heuristic rule 3 the robot builds an internal representation of the route as a topological route graph with nodes representing intersections and edges representing actions connecting the intersections.

4.3 Experimental Results

The *ACE* robot found its way to a designated goal location without the help of GPS data or previous map knowledge, solely by asking passers-by for directions. The robot managed to travel the distance of $1.5\,km$ between the campus of Technische Universität München and Marienplatz (the central square of Munich) in 5 hours, interacting with 38 passers-by. The average duration of an interaction was 1.5 minutes. An example of an interaction between *ACE* and a passer-by is shown in Figure 6, where a passer-by points into the first direction the robot has to go while ACE follows the gesture with the camera head to take an image and present it to the human.

Figure 6. *ACE* following the pointing gesture of a passer-by with the camera head.

Figure 7 depicts the route graph, built by the robot, along with a corresponding occupancy grid in a satellite image of the real environment. The robot was currently positioned at node N_2, where it had retrieved topological knowledge of the route that lied ahead (white). The direction from N_2 to N_3 was given through a gesture, further directions were given through buttons on a touch screen.

The fact that the robot reached its goal solely with the help of instructions of passers-by who were not previously instructed on how to interact, leads to the conclusion that the interactions were successful and therefore intuitive for the humans. Problems arose, where the human partners had too high expectations of the abilities of the robot. For example many users expected the robot to be able to understand speech at first and tried to answer through natural language until they realized that they had to use the touch screen to communicate. Also some humans had problems making a pointing gesture that was recognizable by the robot, as the robot could only recognize gestures where the arm was fully extended and therefore instructed the humans on how to point. The robot was sent in the wrong direction

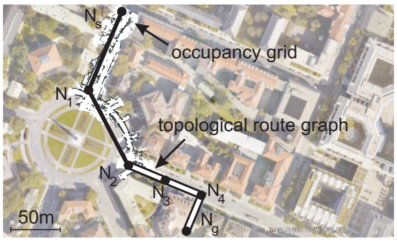

Figure 7. Example of a route graph retrieved from human instructions.

once by a passer-by, but this wrong information was corrected by the next interaction partner after the robot had to stop because the path was blocked. Otherwise no conflicting information occurred.

To summarize most of the problems in HRI arised when it was not entirely clear to the human, what abilities the robot possessed and what the constraints were. Therefore the robot must clarify those points at the beginning of every interaction. The next step is to make the communication more natural, in specific improve the gesture recognition and include a robust speech recognition system incorporating all of the introduced heuristic rules for interpreting deictics.

5 CONCLUSION

In future applications we see the necessity for robots to be able to retrieve the missing information through HRI. Human-human communication was used as a role model for human-robot communication. To this end theories from linguistics research focussing on asking for and giving directions were presented. Heuristic rules for a robot asking a human for the way were deduced. Additionally we presented an experiment where a robot used 4 of the proposed rules to successfully asks passers-by for directions and find an unknown goal location in Munich, Germany.

The heuristic rules that have been implemented so far proved useful in rendering the interaction more intuitive for the human, helped avoiding ambiguities in the route descriptions, and enabled the robot to build an internal representation of the route description. The experiment showed that it is necessary to explain exactly to the human partners how to interact with the robot and what the constraints are. Our future work will be to include all of the proposed heuristic rules in a text based dialog system in order to evaluate them.

ACKNOWLEDGEMENTS

This work is supported in part within the DFG excellence initiative research cluster *Cognition for Technical Systems – CoTeSys*, see also `www.cotesys.org` and within the European Commission as part of the *Robot@CWE project*, see also `www.robot-at-cwe.eu`.

REFERENCES

[1] S. Anderson and E. Keenan, *Language typology and syntactic description, Vol. 3: Grammatical categories and the lexicon*, chapter Deixis, 259–308, Cambridge University Press, 1985.

[2] H. Asoh, Y. Motomura, F. Asano, I. Hara, S. Hayamizu, N. Vlassis, and B. Kröse, ‘Jijo-2: An office robot that communicates and learns’, *Intelligent Systems*, **19**(5), 46–55, (2001).

[3] A. Bauer and M. Buss, ‘Information retrieval system for human-robot communication - asking for directions’, in *ICRA*, (2009).

[4] A. Bauer and M. Buss, ‘Modeling probability distributions of deictic expressions ’here’, ’near’, and ’far”, Technical report, Inst. of Automatic Control Eng., Technische Universität München, (2009).

[5] A. Bauer, K. Klasing, G. Lidoris, Q. Mühlbauer, F. Rohrmüller, S. Sosnowski, T. Xu, K. Kühnlenz, D. Wollherr, and M. Buss, ‘The autonomous city explorer: Towards natural human-robot interaction in urban environments’, *International Journal of Social Robotics*, (2009).

[6] K. Bühler, *Theory of language. The representational function of language*, Benjamins, 1934/1990.

[7] W. Burgard, A. Cremers, D. Fox, D. Hähnel, G. Lakemeyer, D. Schulz, W. Steiner, and S. Thrun, ‘Experiences with an interactive museum tour-guide robot’, *Artificial Intelligence*, **114**(1-2), 3–55, (1999).

[8] B. Claus, K. Eyferth, C. Gips, R. Hornig, U. Schmid, and F. Wysotzki, ‘Reference frames for spatial inference in text understanding’, in *Spatial Cognition*, pp. 241–266. Springer, (1998).

[9] C. Fillmore, *Lectures on Deixis*, U. of Chicago Press, 1997.

[10] E. Fricke, ‘Origo, pointing, and speech: the impact of co-speech gestures on linguistic deixis theory’, *Gesture*, **2**(2), 207–226, (2002).

[11] H. Gross, H. Böhme, C. Schröter, S. Müller, A. König, C. Martin, M. Merten, and A. Bley, ‘Shopbot: Progress in developing an interactive mobile shopping assistant for everyday use’, in *SMC*, (2008).

[12] H.Hüttenrauch, K.Severinson Eklundh, and E.A.Topp A.Green, ‘Investigating spatial relationships in human-robot interaction’, in *International Conference on Intelligent Robots and Systems*, (2006).

[13] W. Klein, *Spatial Orientation: Theory, Research, and Application*, chapter Deixis and spatial orientation in route directions, 283–320, New York: Plenum, 1980.

[14] W. Klein, *Speech, Place, and Action: Studies in Deixis and Related Topics*, chapter Local Deixis in Route Directions, 161–182, New York: Wiley, 1982.

[15] S. Lauria, G. Bugmann, T. Kyriacou, and E. Klein, ‘Instruction based learning: how to instruct a personal robot to find hal.’, in *European Workshop on Learning Robots*, (2001).

[16] J. Lyons, *Linguistic Semantics: An Introduction*, Cambridge University Press, 1995.

[17] M. Michalowski, S. Sabanovic, C. DiSalvo, D. Busquets Font, L. Hiatt, N. Melchior, and R. Simmons, ‘Socially distributed perception: Grace plays social tag at aaai 2005’, *Autonomous Robots*, **22**(4), 385–397, (2007).

[18] R. Müller, T. Röfer, A. Lankenau, R. Musto, K. Stein, and A. Eisenkolb, ‘Coarse qualitative descriptions in robot navigation’, in *Spatial Cognition II*, pp. 265–276, (2000).

[19] M. Schröder and J. Trouvain, ‘The german text-to-speech synthesis system MARY: A tool for research, development and teaching’, *International Journal of Speech Technology*, **6**, 365–377, (2003).

[20] M. Skubic, D. Perzanowski, S. Blisard, A. Schultz, W. Adams, M. Bugajska, and D.Brock, ‘Spatial language for human-robot dialogs’, in *Transactions on Systems, Man and Cybernetics*, pp. 154–167, (2004).

[21] T. Spexard, S. Li, B. Wrede, J. Fritsch, G. Sagerer, O. Booij, Z. Zivkovic, B. Terwijn, and B. Kröse, ‘Biron, where are you? enabling a robot to learn new places in a real home environment by integrating spoken dialog and visual localization’, in *IROS*, pp. 943–940, (2006).

[22] D. Spiliotopoulos, I. Androutsopoulos, and C. Spyropoulos, ‘Human-robot interaction based on spoken natural language dialogue’, in *E. Workshop on Service and Humanoid Robots*, (2001).

[23] S. Thrun, M. Beetz, M. Bennewitz, W. Burgard, A.B. Cremers, F. Dellaert, D. Fox, D. Haehnel, C. Rosenberg, N. Roy, J. Schulte, and D. Schulz, ‘Probabilistic algorithms and the interactive museum tour-guide robot minerva’, *IJRR*, **19 (11)**, 972–999, (2000).

[24] D. Wunderlich, *Linguistische Berichte*, volume 53, chapter Wie analysiert man Gespräche? Beispiel: Wegauskünfte, 41–76, Helmut Buske Verlag, 1978.