Spatial Directionals for Robot Navigation
Andi Winterboer¹, Thora Tenbrink², Reinhard Moratz²

¹School of Informatics, University of Edinburgh, United Kingdom
²SFB/TR 8 Spatial Cognition, University of Bremen, Germany

Abstract. Previous research on spatial projective terms such as *to the left (of)* and *in front (of)* typically focuses on static (locative) usages. We address the usage of such expressions in dynamic contexts (i.e., as directionals). As part of our general aim of developing a speech interface for allowing intuitive control of a mobile robot in navigation tasks, we implemented a carefully selected subset of interpretations of directional terms in a robotic system. The system was validated and gradually improved by experiments involving spontaneous directional instructions by users who were not informed about the robot's capabilities.

1. Introduction
Previous research on spatial projective terms such as *to the left (of)* and *in front (of)* typically focuses on static (locative) usages. In these approaches it is often assumed that dynamic (directional) usages, i.e., those expressing motion into a direction specified by an expression such as *to the left or forward*, can be (more or less) directly derived from insights gained on the interpretation of the locative expressions (e.g., Herskovits, 1986; Levinson, 2003; Eschenbach, 2005). This assumption goes back to a proposal by Miller and Johnson-Laird (1976) who state that dynamic usages are closely interrelated to static ones, as reflected by the fact that the same basic expressions can often be used in both kinds of contexts.

Without doubt, there is a high degree of overlap between these two kinds of usages of spatial terms. In fact, the interpretation of dynamic utterances potentially involves similar complexities as those identified in the literature for static usage. For example, in the sentence 'Put the cup behind the plate', an underlying relative reference system (cf. Levinson, 2003) can be identified: since the plate does not have any intrinsic sides, the term *behind* needs to be interpreted relative to an observer's perspective. In 'Put the rucksack behind you', in contrast,
the reference system is intrinsic because the addressee's intrinsic back side is used for reference. These distinctions are well-known from the investigation of static usage of projective terms.

However, directionals\(^1\) also involve aspects that do not directly mirror static usage. For instance, static usage always involves an explicit referent (such as the cup in 'the cup is to the right of the plate') as well as an (implicit or explicit) relatum (here the plate). In contrast, in a very common usage of directionals it is not necessary to refer to an explicit reference object or a relatum, as in 'turn left'. Moreover, this utterance may be interpreted either as a rotation or as a movement instruction. In both cases, the quantity of the movement needs to be determined; this cannot be derived directly from knowledge about the static usage of projective terms. Thus, the analysis of the acceptability features and the interpretational scope of directional terms is an important research field in its own right. In this paper, we focus on a restricted scenario in which a particular subset of directionals is used regularly and spontaneously by speakers, namely, linguistic movement instructions to a robot. This kind of usage does not involve a further entity other than the addressee (the mover), who is not expressed linguistically in instructions taking the imperative form. Accordingly, there is no conflict of reference frames.

One of the aims of the research project SFB/TR 8 on Spatial Cognition (Bremen/Freiburg; funded by the German Science Foundation DFG) is to enable fluent and intuitive communication between humans and robots about spatial issues. Our basic scenario involves asking users who are not informed about the robot's capabilities to instruct the robot to move towards one of several similar objects present in a configuration. This scenario is essential for a broad range of service robot application contexts (Moratz et al., 2001). While it could be expected that users spontaneously refer directly to the goal object by using static locative terms, as in 'go to the box on the left', users unfamiliar with a system relatively quickly switch to low-level strategies such as 'go left' when advising a robot, especially if the goal-based strategy fails for some reason (Moratz & Tenbrink, 2001).

---

\(^1\) In this paper, following Eschenbach (2005), we use the term "directional" for dynamic usage of projective terms only. This term stands in contrast to the term "locative" for static usage.
Thus, speakers frequently use projective terms dynamically, indicating directions in which a robot might move, avoiding the mention of objects. Therefore we decided to complement our previous research on static projective terms by an investigation of a selected subset of directionals, leading to excellent performance results for instructions given spontaneously by users without the need of listing possible commands. Our robotic system starts from the interpretation of directional terms in specific ways that are motivated on theoretical grounds; its iterative development and evaluation complements these findings by showing whether the decisions are pragmatically adequate in the given human-robot interaction context.

2. The interpretation of projective terms in static vs. dynamic situations
Using a projective term involves indicating a spatial direction within a certain region of acceptability and with respect to an underlying reference system (Levinson, 2003). Static projective terms denote spatial relations between two objects. One object serves as relatum, and the other (the referent or locatum) is positioned within a region surrounding a half axis (top, bottom, left, right, front and back) with respect to the relatum (Vorwerg, 2001). The underlying reference system (intrinsic, i.e., feature-based, or relative, i.e., viewpoint-based) determines how the directions are allocated. The size of the region depends on contextual factors (Carlson & Logan, 2001) but is at all times limited to a half plane (Herskovits, 1986:181f., Retz-Schmidt, 1988). With unmodified projective terms the most likely position is on the half axis itself; with increasing distance from the axis acceptability decreases. These effects have been treated formally in terms of 'spatial templates' (Logan & Sadler, 1999); they are reflected linguistically by increased use of modifiers or combinations of projective terms (Zimmer et al., 1998). However, they also depend crucially on the discourse task. Tenbrink (2007) shows that, in a situation in which an object needs to be identified, speakers adhere to a number of principles (see also Herrmann & Deutsch, 1976) such as minimal effort, maximum contrast, and partner adaptation (with an imaginary partner). Thus, in contrastive reference, spatial terms are preferred that are discriminative without linguistic modifiers if possible under consideration of the other reference candidates present. In other discourse
tasks where an object's location needs to be described with respect to another one, graded acceptability plays a much greater role (Vorwerg & Tenbrink, 2007).

Directional expressions are often viewed as similar (and secondary) to the corresponding locative terms. Eschenbach (2005) proposes the following description:

'The directional use of a preposition refers to a path that leads into a region as characterized by the locative use of the same preposition. Combinations of nach ('to') or von ('from') with one of the locative adverbs form directional adverbial expressions. (…) [T]he spatial condition expressed by the adverb (e.g., oben) specifies the goal region (nach oben) or the origin (von oben) of the path the composite expression refers to.'

Thus, goal (or source) regions are defined in a similar way as regions in static situations. For instance, it is possible to define a goal (or source) region on the grounds of different reference systems, using an explicit relatum. Furthermore, directionals are often used without an explicit relatum, as when an entity is moving autonomously in a direction specified by a directional, as in turn left. Such utterances are essentially non-relational, since no (explicit) spatial relation between entities is involved. They can be interpreted either as a rotation on the spot (see below), or they can be interpreted as a change of movement into the specified direction. Example (1) below would typically be interpreted using the external regions as defined by the addressee's internal sides (although different interpretations are possible if a different relatum is assumed). The movement to the right is then a movement into the goal region on the right hand side of the addressee, as described by Eschenbach.

(1) Move to the right!

It can be assumed that the region of acceptability in such a situation is similar to the regions encountered in static usage, i.e., the most likely direction is a movement on (or to) the half axis itself. Similarly, a forward motion may in the standard case describe a motion at a zero degree angle with respect to the moving entity's orientation. However, there are other options. In a context containing a path (such as a street with curves) it may need to be interpreted to mean something like follow the path in a more-or-less forward direction (e.g., Gryl et al., 2002). And if somebody who is already in a forward motion is addressed by now to the right, depending on context this
might involve a motion toward, say, a 45° angle rather than 90°, since the forward motion is merged with the rightward motion. In a route instruction context, again, turn left induces a search for a path on the left hand side of the moving entity; in particular, the future direction is determined by the first intersection of the current path with another path situated on the left of the mover (Gryl et al., 2002).

Thus, depending on the discourse situation it may or may not be feasible to apply the notion of 'spatial template' in a similar way as for static usage; in fact, with respect to some contexts this notion seems to be rather irrelevant. Also, since directional usages often only give the goal direction without a clear end position, the exact distance that should be covered is unclear.

As already indicated, movements into a newly specified region need to be differentiated from rotational movements, in which an expression like left does not specify a future direction to move into, but only a re-orientation towards the left side. This may not always be obvious: depending on context, a brief utterance like to the right or rechts may be intended to mean either of both. How rotational descriptions should be interpreted is, for instance, addressed in Habel (1999). Here, also, the expressions are underspecified with respect to the quantity of the movement; this may concern either the distance to be covered into a specific direction as well as the angle of rotation. Both of these may be influenced by contextual factors which require further empirical investigation.

Terms such as vorwärts / geradeaus (forward / straight ahead) carry a dynamic element already in their semantics, in contrast to the projective terms to the left / front etc. While it could be assumed that, in dynamic contexts, these are approximately synonymous to nach vorne (to the front), there are in fact systematic differences in usage, as illustrated by the following:

(2) Ich gehe nach vorne. [I am going to the front.]
(3) Ich gehe vorwärts / geradeaus. [I am going forward / straight ahead.]

If uttered on a train, example (2) would probably be interpreted to mean that the speaker intends to go towards the front section of the train, regardless of whether the speaker is currently oriented towards the train's front or
happens to look towards a different direction. But example (3) can only mean a forward motion on the part of the speaker (defined by the speaker's orientation), which may or may not coincide with the forward direction of the train. With respect to the latter type of expressions, Eschenbach (2005) notes:

'The adverbs vorwärts, rückwärts, and seitwärts ('forward', 'backward', 'sideways') specify the alignment of a path relative to the intrinsic reference system of the figure. Vorwärts ('forward') expresses that the direction of motion is in accordance with the intrinsic orientation of the body. Thus, the reference system is bound to be intrinsic to the figure and cannot be specified differently by contextual influences. The geometric condition can be described as the alignment of the object order of the path with the intrinsic access order of the figure. The lexeme rückwärts is morphologically related to the noun Rücken (the body-part 'back') and seitwärts to the noun Seite ('side'). Rückwärts ('backward') expresses that the backside of the moving figure (...) is leading, i.e., precedes the center. Correspondingly, seitwärts ('sideways') can be used to say that a lateral side of the moving figure is leading.'

The lateral axis does not offer such a distinction between only-intrinsic versus more flexible expressions in German, except for seitwärts which is unspecified for direction on the axis. In English, leftward(s) and rightward(s) seem to be available though used infrequently.

(4) Ich gehe nach rechts. [I am going to the right.]

The interpretation of (4), uttered on a train, would probably depend on the speaker's orientation as in (3), in spite of the fact that the surface form corresponds to that in (2). But this intuition may be due to the fact that the internal front and back regions of trains are much more prominent than their right and left sides. A different situation is provided, for example, in reference to the regions within an opera house, which are often even explicitly marked as 'left' and 'right'. Furthermore, it is likely that the interpretation of nach vorn (to the front) is influenced by the availability and relevance of background entities with internal regions, such as the train in example (2). Without such a mutually agreed-on background entity, a forward motion of the speaker may be more relevant, rendering the utterance synonymous to example (3). In English a forward motion can only be expressed by forward, straight (ahead) and perhaps ahead, but not to the front; for the German nach vorne, the case is less obvious. Clearly, targeted empirical investigations are necessary to shed further light on these phenomena. Our experimental study described in the next section contributes to this issue by showing in how far speakers in a human-robot movement instruction context spontaneously use nach vorne.
3. Human-Robot Interaction and Directionals

In this section, we examine and point out potential difficulties in a real-world application, adopting a computational perspective. While we do not attempt to account for the full range of interpretational options sketched in this paper so far, we have implemented the most fundamental subset, namely, the usage of a directional to provide a future direction for a moving entity (a robot) without reference to an external relatum. Our aim was to achieve pragmatic adequacy with respect to the envisioned human-robot interaction scenario.

In our research program within the project SFB/TR 8 Spatial Cognition we aim to enable efficient and intuitive communication between human users and robots. In our current target setting, a robot is instructed to move to a particular place; the users achieve this by relying on their own intuitions rather than a list of commands. Nowadays, direct control devices, e.g., joysticks or graphical user interfaces, can achieve near optimal results without linguistic modules (Tsuji & Tanaka, 2005). Such a direct control system can benefit from multimodal interaction methods combining gestures and verbal commands from a predefined list (Trouvain et al., 2001). However, such systems are less suitable than language-based control systems for more advanced generic (for example, conditional) tasks and interaction scenarios in which humans and robots are not co-present. For the first steps towards speech-based human-robot interaction, it makes sense to start by enabling direct natural language control in individual tasks in face-to-face scenarios, even if these scenarios from an engineering point of view could be solved by non-linguistic means more easily. Any comprehensive robotic system capable of interpreting generic natural language instructions would certainly be equipped to deal with direct speech-based commands as well. Generally, it is advantageous to enable simple control in face-to-face scenarios before moving on to more challenging generic instructions, for example, in order to familiarize new users with the speech interface (Moratz & Tenbrink, in press). Although this kind of linguistic motion control may not be a technological novelty, we do not know of accounts in the literature reporting of an evaluation of a similar implementation with naïve users, which is our particular focus.
In a human-robot interaction scenario in which a human controls a physically embodied agent like a mobile robot, not only static objects can be referred to, but the entire (dynamic) physical environment. Bos et al. (2003), for instance, present a system capable of interpreting goal-based (place-related) instructions such as 'go to the kitchen'. Kruijff et al. (2007) and Spexard et al. (2006) describe robotic systems able to learn relationships and locations in the environment with the help of a human tutor using natural language. However, one major previous result of our own empirical work (Moratz & Tenbrink, 2006) is that participants spontaneously produce incremental (step-by-step) rather than object-based descriptions. Thus, in a scenario where users are not informed about the robot's capabilities and are asked to instruct a robot to move to one of several similar objects indicated by the experimenter, they tend to use directionals such as 'move forward and then to the right' rather than goal-based static spatial instructions such as 'move to the object on your right'. Since this was an unexpected result, previous versions of our system did not account for the former kind of instruction. In Winterboer (2004), an implementation of directionals for the same kind of task was successfully accomplished. In the following, we describe the main aspects of this system, which was developed in several iterations on the basis of the results of experimentation. We discuss problem areas encountered during the development process and present the solutions found in the current implementation.

3.1. The robot system architecture

Our aim in the present work was to develop a speech interface for allowing intuitive control of a mobile robot in navigation tasks. The deployed system consisted of an AIBO robot (Figure 4), a speech recognition and natural language understanding module using Nuance tools, and a robot motion control module. Among other possible behaviours, AIBO robots can move in several directions as well as rotate on the spot. Nuance's speech recognizer allows specifying a speech recognition package on the basis of their grammar specification language (GSL), which is both used for language modelling and parsing, and requires a careful design process taking into account the linguistic knowledge of the domain. The recognized utterances trigger predefined actions which are sent to the robot by the robot motion control module, AIBOControl, via WLAN. To carry out the predefined
actions, a navigation component based on AIBOControl, acting as a compact version of the powerful SimGT2003 (Burkhard et al., 2002), was implemented.

For the experiments, we enabled the AIBO to perform forward and backward motions, to stop a current movement, to turn on the spot, and to skew, i.e., to move into a direction of approximately 45 degrees to the left or right for a distance of one meter (see Figure 5 below). The skewed movement was implemented in order to combine simple forward movements with turns, as could be intended, for example, by an utterance like 'go to the right'. The decision to use a 45° angle was largely arbitrary, but motivated by the idea that a default forward direction is combined with a partial reorientation to the left or right. Though one may argue that a 90° angle for such instructions might be more intuitive, we hypothesized that restricting the angle of such a skewed movement would support the user in approximating the goal in small steps.

For every movement type, different linguistic variations could be uttered. To define the content of our lexicon, containing approximately 100 words, we took into account the theoretical considerations described above as well as the variability of users' linguistic choices that we observed in earlier experiments (e.g., Moratz
The user study described in this paper addresses our experience with a system that was specifically designed to deal exclusively with incremental (i.e., not goal object based) utterances.

The system interpreted utterances such as 'geradeaus (gehen)' [(go) straight on], 'vor' / 'vorwärts' [forward], and 'geh' / 'lauf' / 'fahre' [go] as a forward movement. Backward movements could be expressed by 'zurück' / 'rückwärts' (gehen) [(go) backward] and the like. Left and right rotational movements could be triggered by 'dreh links/rechts' [turn left/right], 'links' [left], 'nach links' [to the left], and similar terms; left and right skewed movements by 'geh links/rechts' [go left/right], etc. Finally, a stop could be expressed by 'stop' / 'halt' [stop]. The full lexicon can be found in the appendix.

Thus, a range of semantically similar expressions were treated as if they were synonyms. For example, apart from directionals indicating a forward movement, the forward direction was treated as a default for underspecified indications of movement (go). Directionals indicating the lateral axes are generally treated as instructions to turn but not move; only instructions that additionally contain a verb of movement are interpreted as a skewed movement. In this case, it was interesting to find out whether the 45° angle was suitable to enable successful movement in the intended direction. In general, although the interpretation decisions do not necessarily account for subtle differences in the expressions' semantics, the experimental results will show
whether the deployed procedure is pragmatically adequate for the purpose at hand. This is a sensible approach especially in light of the fact that a number of issues are still unresolved in the literature, including the preferred angle for a skewed movement or a turn. This question is only relevant in scenarios where no additional information can be derived from the scenario itself, as for example information provided by a street network. In accord with the findings reported above, the expressions *nach vorne* (to the front) and *nach hinten* (to the back) were not implemented; it was assumed that these expressions would not occur in the given context, since internal reference systems were less likely to be employed (cf. Section 2 above).

### 3.2. Experimental study

We asked participants to instruct the AIBO robot to move from a given start to a goal position by using natural language instructions. Based on our previous work we could expect that naive users, who were not informed about the robot’s abilities, would spontaneously choose an incremental instruction strategy by using directionals to control the robot. To test this hypothesis, we did not tell the users what kind of instruction they should use in order to find out about their intuitive strategies. Furthermore, there were a number of open questions that needed to be addressed in order to allow for effective and intuitive instructions using directionals. For example, prior to the study we could not know whether the participants would prefer continuous robot movements until a definite verbal “stop” command, or whether limited movements would be preferable until a specific distance was covered. In addition, the pragmatic adequacy of the above interpretation decisions needed to be examined.

After an evaluation phase, we performed a revision of the existing system. To achieve a functional speech interface several further experiments were conducted with the revised system. Using this iterative approach of alternating model building and empirical phases, a direct feedback between simulation and experiment was achieved.
3.3. Procedure

The experiment was conducted in rooms of the University of Bremen. 21 participants (15 male; 6 female) were asked to navigate the AIBO robot to particular objects or locations pointed at by the experimenter, using German language instructions. Two participants took part twice (at the beginning and at the end of the experimental study). The mean age of the participants was 29 (range: 19 - 44). 13 of the 21 participants had a computer science background. The experiment took approximately 15 to 20 minutes per participant. Altogether, 93 navigation tasks using various configurations were completed.

The participants sat in front of a desk and were equipped with a headset for instructing the robot. They were requested to deal with several scenes which consisted of a start and a goal position, plus up to four objects (identical cardboard boxes) arranged in a configuration (cf. Figure 6). The experimental setting was carefully designed to minimize the high variety of factors that may influence the performance of a speech based navigation task. For example, markings on the floor guaranteed that the positions of the robot, the obstacles as well as the goal could be precisely replicated for each participant. In addition, to avoid effects of order, the order of the particular navigation tasks was randomized.
Each time the robot arrived at the intended goal position, the configuration of the objects was changed. The participants did not get a response if their instruction was not understood by the speech recognition; in fact, the robot did not talk at all. If the user's instruction could be interpreted by the robot, the robot started to move, otherwise nothing happened. Thus, in accord with the methodology proposed by Fischer (2003) the test participants did not receive any hints concerning the implemented computational model or the linguistic abilities of the robot. If the participants' instructions were not successfully recognized, they had no indication regarding the reasons, and therefore developed their own intuitive strategies for achieving successful communication.

The experiment was carried out in five stages (experiment parts) consisting of a varying number of single experiments. In between, the system was modified according to the experience gained in each stage.

3.4. Results
Altogether, we collected 1536 instructions, 1181 of which were successfully recognized and carried out by the robot, yielding a recognition rate of 76.9%. The following general results pertain to all experiment parts.

- Our hypothesis that participants would primarily use incremental instructions (i.e., directionals and motion verbs such as go) to instruct the robot was confirmed. In fact, only one single participant directly referred to the goal position in four instructions before turning to incremental instructions, too. Note that, in our previous experiments described in Moratz & Tenbrink (2006), those participants whose initial incremental instructions were not successful typically did not spontaneously switch to the goal-based strategy. If they started out using a goal-based strategy and their instruction failed for some reason, they usually directly switched to the (non-implemented) incremental strategy. In the present experiment, the users did not attempt to use a different level of instruction, such as goal-based instructions, after unsuccessful attempts. Instead, a typical reaction was a modification of the utterances concerning lexical or syntactical choice. This result corresponds to earlier findings according to which users tend to switch to lower, but not higher level strategies in case of failure (Fischer & Moratz 2001).
The range of expressions we expected from previous experiments as well as theoretical considerations corresponded fairly exactly to the instructions actually used by the participants. Only 47 instructions contained an expression that was not contained in the lexicon; therefore, 96.9% of all utterances were theoretically interpretable (the remaining failures were due to speech recognition rather than system coverage). This is an impressively high proportion especially in light of the fact that the participants were not previously informed about the robot's capabilities.

As expected, the specific directional terms *nach vorne* (to the front) and *nach hinten* (to the back) were almost never used, confirming our expectation that they were not typical in the given context. However, the fact that one participant did use *nach vorne* four times (before switching to *geradeaus*) shows that this usage is not entirely ruled out.

The participants did not exhibit fundamental problems with the interpretations of their instructions. They seemed to be surprised about the skewed movement; the simple turning behaviour appeared to be easier to handle. This seems to be related to the fact that this was the only compound movement, consisting of a turn on the spot and a movement into a new direction. In general, however, the robot appeared to behave in an expected way, apart from a number of problems described below.

Therefore, the initially implemented lexicon turned out to be suitable to a very high degree. The few modifications we carried out in the course of the iterative process concerned adding a 180 degree turn to the lexicon and removing the term *weiter* (which was interpreted as a forward motion to begin with, but, due to the deployed keyword spotting method, turned out to be problematic in connection with instructions indicating other directions of movement). Since there was no clash between the users’ intentions and the robot’s reactions, there was no need for further modifications at this point. The other terms that were uttered by the participants and that were not contained in the lexicon, were only used once or twice. In order to keep grammar and lexicon as concise as possible for obtaining the best possible recognition results, we did not add such exceptional expressions to the lexicon.
However, a high influence of individual differences could be observed. For instance, some participants easily lost their temper when the robot did not react as quickly as expected. Sometimes there were delays between an instruction and the corresponding robot movement, caused by the high demands of working memory required by the speech recognizer. Then, the participants repeated their instruction instead of waiting for the processing of the last utterance. Furthermore, some utterances were not correctly recognized because the instructions were uttered too quietly. An appropriate adjustment of the headset and a clear articulation supported the recognition. Finally, those participants who were not confident and afflicted by self-doubts had the worst recognition rates and needed the highest number of instructions. They blamed themselves for the bad experiment progression instead of putting the blame on the speech recognizer or other technical modules. Therefore, their problems only increased when they noticed that the AIBO did not act as expected. In general, those participants were most successful who acted self-confidently and pronounced their utterances with a clear articulation.

Figure 7. Average number of instructions/successful instructions required per navigation task

There was an overall increase of success throughout the study, as could be expected due to the gradual improvement of the system. To illustrate this improvement we compare the first and the last experiment part. While it took the first eight participants on average about 83 verbal instructions per person to solve their navigation tasks, the last eight needed only about 58 for theirs. Figure 7 illustrates this result, showing the
average number of instructions needed per configuration along with the average of successful instructions (where the AIBO acted as expected). Note that instructions may be successful without leading directly to the goal position, which is why, in more efficient trials, speakers use fewer successful instructions to reach their goals. In addition, not only did the last eight participants require fewer instructions on average to arrive at the goal position, they also solved their tasks in less time. The average duration per configuration until the goal position was reached decreased from roughly 88 seconds (participants 1-8) to approximately 65 seconds (participants 15-23). Therefore, the revisions clearly enabled more effective robot navigation.

Although the relation between uttered instructions and successfully executed instructions improved only slightly during the experiments with the first and the last eight participants (76.1 to 80.0 %), the average number of instructions required for navigating the robot to the goal position decreased considerably from an average of 20.8 to 13.8. Thus, the instructions became increasingly effective throughout the study due to the modifications to the system. In the following, we give a more detailed account of the system’s iterative development process.

3.5. Iterative development process

One of the open questions prior to the experiment had been whether the robot should carry out a continuous movement when it was instructed to move in a certain direction or to turn, or whether it should stop after a certain distance or angle. We started from the former variant, assuming that a continuous movement would feel natural to users, since the users would not need to repeat instructions. Therefore, whenever a directional was recognized the AIBO performed the corresponding movements in a continuous way and only stopped when the user uttered an explicit instruction, such as 'stop'.

However, it turned out that the delays between the uttered instructions and the robot reactions caused problems. Participants did not anticipate the continuous robot movements and therefore the robot frequently overshoot the mark. To meet this problem, we first reduced the speed of the robot motions when a turn instruction was recognized in order to decrease the covered distance or angle. When it turned out that this was not sufficient, the movements were also given a restricted value. For instance, the turning movement was
restricted to 45 degrees, mirroring the implemented skewed movement. The forward and backward movements were restricted to one meter each. After these modifications, the users’ interactions with the speech interface were more efficient, which clearly improved the results. Crucially, the participants seemed to get used to the restricted quantity of the movement (distance or angle) rather quickly and could focus on the next movement to be accomplished, because they did not need to stop the current movement via a new instruction. After some further experimental iterations, we settled on turning angles of 30 degrees which seemed to be pragmatically optimal in our scenario (see Winterboer 2004 for details).

Since there were still some problems with delayed responses, we furthermore updated the prioritization of the 'stop' instruction within the GSL recognition grammar in order to let the speech recognizer always choose this instruction in the case of an ambiguous utterance. Further improvement could be obtained by decreasing the WLAN traffic volume by optimally reducing the AIBO sensor data (e.g., camera data) that were automatically transferred via WLAN from the robot to the AIBOControl robot motion control module. This effect was further supported by carrying out the experiments in a WLAN traffic free testing environment in which no other WLAN traffic could affect the connection, and where one router was exclusively allocated to transfer the data between the AIBO and the computer operating the robot motion control module.

Another modification affected the lexicon as well as the moving behaviour of the AIBO. In some of the experiments the participants tried to about-face the robot with the two phrases:

(5) *Drehe dich um 180 Grad* (turn 180 degrees)
(6) *Umdrehen* (turn around)

Therefore we inserted both phrases in the lexicon and included the corresponding movement in the robot action module.

3.6. Summary of the human-robot experiment

We have presented the iterative development of a speech interface for an AIBO robot, aiming at solving navigation tasks by intuitive natural language instructions. By evaluating the behaviour of the users in
connection with the robot's reactions and by carrying out several modifications, an empirical validation of the speech interface was obtained. Our results show that the initial interpretation decisions with respect to a range of linguistic expressions (less than 100 words in the lexicon) turned out to be pragmatically adequate. The users could, with a high degree of success, use the kind of language they intuitively expected to be successful. Thus, our expectations originating in theoretical (literature-based) considerations as well as previous experiments with a different system were confirmed. The remaining problems that were detected and addressed throughout the study primarily concerned other kinds of factors. Here, the most important revisions were the decrease of the turning angles as well as the speed, the prioritization of the 'stop' command within the GSL grammar, and the reduction of the data flow between the robot and the robot motion control module. These modifications resulted in a reliable and, even for uninformed users, well-controllable speech interface.

One question that calls for further experimentation concerns the ways in which turning behaviour and movements into a non-straight direction (skewed movements) could be expressed linguistically and interpreted optimally by the robot. In the present solution, it turned out to be easiest to have the robot turn on the spot and then, with a separate instruction, let it move forward. But other solutions are conceivable, since the semantics of directionals like *rechts* and *links* are both ambiguous (because they can denote a rotation as well as a movement into a non-straight direction) and underspecified (because angles and distances are not predefined). The participants' slight surprise with respect to the skewed moving behaviour of the robot highlights this observation. Further experimentation could shed more light on this issue.

4. Conclusions and Outlook

The use and interpretation of (spatial) projective terms in natural discourse is influenced by a considerable variety of factors, both in static and in dynamic kinds of contexts. While in some dynamic contexts various underlying reference systems come into play, similar to those used in static scenarios, other usage contexts do not involve entities as relata and are therefore conceived of as non-relational. The motion instructions used in the presented human-robot interaction setting are cases in point. Such terms involve few problems; since the robot’s
intrinsic movements are the sole target of reference, the variability of interpretation is greatly reduced. This leads to a high pragmatic adequacy of a relatively simple system that interprets a range of different expressions in a predefined way, mapping them to suitable robot reactions. As our experiments have shown, incremental motion descriptions based on dynamic projective terms are an essential part of any efficient and robust motion command strategy for navigating mobile robots intuitively. Nevertheless, the ambiguity and underspecification of directional terms leaves room for different ways of interpreting the instructions. This needs to be carefully balanced with respect to the requirements of an actual scenario. For such a goal, an iterative system development starting from theoretical assumptions is particularly useful, as our example demonstrates.

Eventually, in order to cover a greater range of interaction settings, a number of system modifications will be necessary to account for the complexities involved in employing directionals for purposes that go beyond simple robot movements. Furthermore, a number of aspects concerning the use and interpretation of directionals still require empirical research, not only in a human-robot interaction situation but also with respect to psycholinguistic issues. Thus, while our robotic system starts from a simple scenario, the present research has outlined some of the problems and ambiguities involved in more complex kinds of situations that need to be dealt with in the future.

Acknowledgements

The experiments were conducted when the first author was at the Transregional Collaborative Research Center "Spatial Cognition", Faculty of Mathematics and Informatics, University of Bremen. Funding by the Deutsche Forschungsgemeinschaft (DFG) is gratefully acknowledged. We also appreciate support and many fruitful discussions with researchers in the SFB/TR 8 and the University of Edinburgh.
Appendix: Contents of the lexicon

Forward movement:

geradeaus (straight on); geradeaus gehen (go straight on)

vor / vorwärts (to the front)

geh / gehe / lauf / laufe / fahr / fahre (go); los / fahr los (start moving)

weiter (keep moving)

Backward movement:

zurück / rückwärts (backward); zurück gehen / zurück laufen / rückwärts gehen / rückwärts laufen (go backward)

Left turn [right turn is treated equivalently]:

dreh links / drehe links / drehe dich nach links / drehe nach links (turn left)

links (left)

nach links (to the left)

etwas nach links / ein bißchen links (a little to the left)

drehung links / links herum / links umdrehen (left rotation)

Left skewed movement [right skewed movement is treated equivalently]:

geh links / gehe links / fahr links / fahre links / links gehen (go left)

lauf nach links / fahr nach links (go to the left)

Stop

stop / halt (stop)
180 degrees turn
drehe dich um 180 Grad (turn 180 degrees); umdrehen (turn around)

References


Spexard, Thorsten, Shuyin Li, Britta Wrede, Jannik Fritsch, Gerhard Sagerer, Olaf Booij, Zoran Zivkovic, Bas Terwijn and Ben Kröse. 2006. *BIRON*, where are you? Enabling a robot to learn new places in a real home


