Journal of Visual Languages and Computing 21 (2010) 292–309

Contents lists available at ScienceDirect

Journal of Visual Languages and Computing

journal homepage: www.elsevier.com/locate/jvlc

Route instructions in map-based human–human and human–computer dialogue: A comparative analysis

Thora Tenbrink*, Robert J. Ross, Kavita E. Thomas, Nina Dethlefs, Elena Andonova

FB 10, Universität Bremen, Postfach 330440, 28334 Bremen, Germany

ARTICLE INFO

Article history:
Received 13 July 2010
Accepted 15 July 2010

Keywords:
Perspective
Granularity
Human–computer interaction
Communicative strategies

ABSTRACT

When conveying information about spatial situations and goals, speakers adapt flexibly to their addressee in order to reach the communicative goal efficiently and effortlessly. Our aim is to equip a dialogue system with the abilities required for such a natural, adaptive dialogue. In this paper we investigate the strategies people use to convey route information in relation to a map by presenting two parallel studies involving human–human and human–computer interaction. We compare the instructions given to a human interaction partner with those given to a dialogue system which reacts by basic verbal responses and dynamic visualization of the route in the map. The language produced by human route givers is analyzed with respect to a range of communicative as well as cognitively crucial features, particularly perspective choice and references to locations across levels of granularity. Results reveal that speakers produce systematically different instructions with respect to these features, depending on the nature of the interaction partner, human or dialogue system. Our further analysis of clarification and reference resolution strategies produced by human route followers provides insights into dialogue strategies that future systems should be equipped with.

1. Introduction

In our work we are concerned with the development of naturalistic modes of communication for situated applications. Here, situated applications refers to systems that (a) are embedded in a real or virtual environment and (b) can perceive, reason on, and act within that environment. Examples of situated systems include in-vehicle information technologies [12], spatially aware assistance applications [20], as well as cognitive and service robotics (see [22] for an introduction). For such applications, exclusively graphical, restricted textual, or tactile communication is rarely feasible or satisfactory, and so natural language based dialogue is an appealing interaction mode.

For instance, today's GPS-equipped navigation systems are well suited to conveying information about routes both visually and verbally. Although many current route-providing systems are configured to react dynamically to a pre-defined range of user requests, flexible natural language-based dialogue with such a system is still impossible at present, and open questions remain as to what kinds of phenomena such a dialogue system would have to cover. A range of further possible application scenarios for truly dialogue-enabled route information systems are conceivable, encompassing not only user-centric outdoor route navigation but also various indoor settings in which either a human or a system instructed by a human, such as a mobile autonomous service robot for home usage, needs route information. In such scenarios, spatially aware systems need to respond flexibly to user requests, using natural language, preferably in relation to visualizations such as map depictions. Unfortunately,

* Corresponding author. Tel.: +49 421 21864212; fax: +49 421 2189864212.

E-mail addresses: tenbrink@uni-bremen.de (T. Tenbrink), robert.ross@dit.ie (R.J. Ross), kathomas@uni-bremen.de (K.E. Thomas), dethlefs@uni-bremen.de (N. Dethlefs), andonova@uni-bremen.de (E. Andonova).

1045-926X/$ - see front matter © 2010 Elsevier Ltd. All rights reserved.
doi:10.1016/j.jvlc.2010.07.001

dialogic spatial interaction remains under-represented in the literature, both with respect to dialogues between two people and even more so where dialogue systems are involved.

Verbal interfaces to both situated and non-situated applications are traditionally modelled in terms of 'Spoken Dialogue Systems'. These systems are architectures of computational components which handle the varied responsibilities of natural language processing [19]. While the general composition of spoken dialogue systems is relatively well understood, any interaction with the majority of prototype dialogue applications quickly reveals a multitude of limitations. Generic low-level problems such as poor speech recognition quality often undermine the success of spoken dialogic interaction considerably; further limitations are due to a lack of knowledge concerning the kinds of linguistic and cognitive phenomena users both employ and expect from these dialogue systems.

In this paper, we address a scenario in which a human user tells a robot – in this case a robotic wheelchair – to move to a particular location in an indoor office setting, using typed language. The environment and the wheelchair are depicted schematically on a screen in order to provide a shared basis for spatial communication. In order to gain insights about the impact of the interaction partner, we use the same scenario twice—comparing the linguistic choices made in unconstrained human–human interaction (HHI) with those made in a human–computer interaction (HCI) situation that involves interacting with a dialogue system that has various language and dialogic limitations.

Our main interest concerns the ways in which speakers tackle the complexity of the task via communicative strategies that lead to a joint minimization of effort (see [6]). In our scenario, human speakers contribute to the dialogues in three interactional roles: as a *route giver for another person* or *for the system*, and as a *route follower* of human route descriptions. Each of these roles leads to a range of available communication strategies such as general description levels and instruction styles, clarification requests, and reference resolution strategies. Fundamental variations of spatial conceptualization reflected in language concern perspective choice and references to places across levels of granularity. Although these phenomena have been discussed widely in the literature, very little is known so far about how speakers minimize collaborative effort in map-based route dialogues with regard to these phenomena, and particularly how their choices of conceptual reference systems and their linguistic representations are influenced by the interaction situation and develop within a dialogue (but see [8]). Specifically, it is an open question how users intuitively act in this regard towards a dialogue system equipped to deal with spatial settings.

Generally it is well-known that speakers react systematically to the real or imaginary biases, needs and requirements of artificial interaction partners, both with respect to linguistic choices [1,23] and high-level decisions [16]. Even small changes in the experimental setting, including the robot's reactions, may be crucial in this regard [21], along with users' preconceived mental models and expectations that are equally decisive for users' conceptualization of the dialogue and their ensuing linguistic reactions [5]. Unfortunately, in the area of human–robot interaction it is not very common to carry out evaluations of newly developed systems without restricting in advance the language that may be adopted by users. Existing evaluations of this kind have been shown to yield disappointing results simply because the actual language used lies outside of that supported [37]. It is therefore essential for human–robot and human–computer interaction to be based on realistic assessments of the kind of language that users will produce, and their likely reactions to the system's output. In our approach we combine established psycholinguistic experimentation with qualitative empirical discourse analysis of unrestricted dialogic contributions, using both human–human baseline-establishing experiments and genuine human–computer interactions. For the latter, the dialogue system is progressively augmented as the empirical results are integrated.

In the next sections, we provide the background needed to interpret the conceptual aspects addressed in our analysis of spatial dialogue, and briefly present our dialogue system. Then we introduce our general procedure, and subsequently present two studies of human–human and human–computer interaction with their respective results, which are compared in a single discussion section. We conclude by outlining the ensuing consequences for the further development of spatially aware dialogue systems.

2. Conceptual aspects in communicating routes

Consider a situation in which you need to communicate information about a spatial goal to an interaction partner, and you are required to do that via a computer interface such as the one depicted in Fig. 1. Here, a two-dimensional map is shown on the screen together with a chat interface to be used for communication with the agent—in this case a wheelchair. In our scenario, we used two versions of this task. In study 1 (HHI), one person (the route giver) was asked to imagine sitting in a wheelchair that was indicated by an avatar (symbol) on the screen. The goal was to navigate to a destination marked on the map. In order to reach this goal, they were asked to give route directions to their interaction partner via the chat interface. The interaction partner (the route follower) could move the wheelchair by using a joystick, and could respond verbally via the chat interface. In study 2 (HCI), the human user acted as route giver similar to the HHI scenario; the chat interface was coupled to a dialogue enabled agent capable of travelling along a described route, and of reacting verbally via the chat interface. Such a scenario may be used, for example, in order to demonstrate a service robot's future path or to visualize a route in reaction to a request made by a user. In both cases, the simulated wheelchair moving around in the scene could be observed by both participants (more precisely, the dialogue system had access to this information, which the human participants could see on

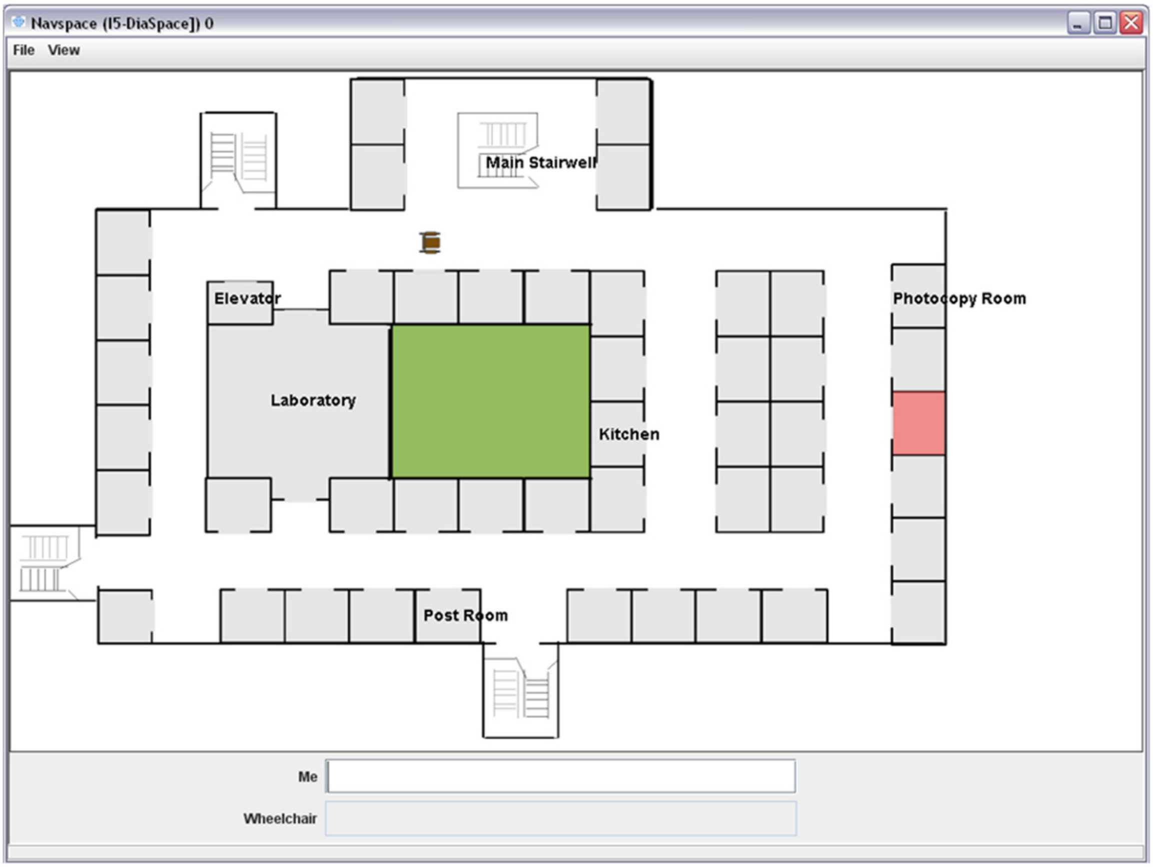

Fig. 1. The Navspace user interface with map. The robot wheelchair is shown in the top hallway, facing towards the right hand side of the picture. The goal location is marked for the route giver only; the location labels are always visible. In the HHI case, joystick movements by the route follower make the wheelchair move on the screen; in the HCI case, the wheelchair movements are handled by the system.

the screen)—but only route givers could see the location of the next destination (marked on the map).

In such a situation, as in all spatial communication tasks, a number of strategies are available to the dialogue partners [33]. Here we will focus on two distinctions crucial for navigation: *granularity* and *perspective*. Both of these aspects involve a substantial potential for rendering dialogue more or less efficient, leading effortlessly to communicative success only if the chosen variant is readily understood and accepted by the addressee.

2.1. Granularity

The notion of granularity refers to the conceptual level of specification in the representation of a particular situation, event, or object. The identification of relevant granularity levels is of central importance for issues of cognition and communication, particularly in the spatial domain [11,38]. While issues of spatial scale ranging from coarse to fine degrees of detail are often regarded as most relevant, granularity levels may also favour particular semantic aspects of a situation over other aspects, independent of scale.

Here we address the level of granularity used naturally by humans when providing information about how to reach a goal location. Such a communicative aim may be achieved either by referring directly to the goal location by using destination descriptions as described by Tomko [40], or by incrementally guiding the traveler to the goal by using turn-by-turn directions [25]. While this may appear to be a binary distinction, in actual fact speakers combine and vary their descriptions flexibly along these lines [36]. This behaviour may reflect their assumptions about efficiency just as well as further communicative aspects that influence such decisions.

The resulting high degree of diversity in spontaneously formulated spatial descriptions clearly needs to be accounted for in the design of adaptive dialogue systems. Discrepancies between human expectations and strategies with respect to the preferred choices of granularity in instruction settings could lead to communication failure, especially if the system is not equipped to deal with the level of granularity chosen by the user—or lacks the information necessary to infer the relevant spatial relationships [35]. In principle, the investigation of natural human–human interaction can provide a gold standard for the joint negotiation of spatial goals in this regard. However, humans may use different strategies depending on the nature of their interaction partners; in the case of automatic systems, this effect is certainly mediated by the users' personal experience with such systems as well as the (perceived) current state of the art concerning their capacity [1,7,16]. Therefore, we investigate both in comparison.

What do we expect concerning granularity considering our two settings, HHI and HCI? Two general previous findings are relevant. On the one hand, dialogue partners aim to minimize the collaborative effort required to reach a shared communicative goal [6]. However, it is not easy to assess in general whether turn-by-turn or rather destination descriptions actually involve more effort, as this may depend crucially on the situation at hand. In the present scenario, a purely destination based description is complicated since none of the pre-defined goals has a label. Complex spatial descriptions using a labelled room as a basis for reference, such as "the room that is two rooms below the photocopy room" (to describe the marked location in Fig. 1), are available in principle, but they may not always appear to be a feasible solution, since only very few locations on the map are labelled at all. This leads to an increased degree of complexity involved in destination based descriptions in our scenario. On the other hand, especially with increasing distance between starting point and destination, incremental turn-by-turn directions may appear to be awkward, leading to a potential trade-off between the two levels of granularity in the instructions.

On the other hand, previous research in our group has consistently shown that goal-based spatial reference – or indeed any reference to a spatial entity, rather than a spatial direction – is conceived of as particularly difficult for automatic systems [10,21,43]. Therefore, the users in our human–computer interaction scenario may not consider destination based descriptions or spatial location references as options at all, regardless of efficiency. However, these earlier studies concerned human–robot interaction contexts with real-world entities rather than maps. Conceivably, users might be somewhat less reluctant to refer to locations on a map which, as they know, is accessible to the system.

Taken together, our prediction is that route givers in the HHI setting will mix levels of granularity, possibly tending towards turn-by-turn directions with short distances and towards destination based descriptions with longer distances. In the HCI setting, in contrast, we predict a systematic general preference for turn-by-turn directions.

2.2. Perspective

In this scenario, there are two main kinds of perspective available, similar to other task scenarios as in [17,32]. One of these, which we here call the *survey* perspective, involves looking at the map from outside the scene or above, the way a reader actually perceives the map. Using this perspective, in order to reach the room marked as the destination in Fig. 1, the wheelchair could move *to the right and then down into the hallway all the way on the right; then the goal is the third room on the right hand side.* The other type of perspective, referred to here as *route* perspective, represents the view of the route-travelling agent. Using this perspective, the wheelchair could move *straight on and then take the second hallway on the right; then the goal is the third room on the left hand side.* Note that in this categorization, unlike some previous accounts, both perspectives allow for both static and dynamic descriptions. They therefore both enable the counting of rooms: however, whether the goal is described as being on the left or right side is a matter of perspective choice. This fact can obviously lead to communicative misunderstandings and a need for clarification.

As with granularity, neither of the two perspective choices is obviously simpler and more efficient in the given scenario. Route perspective matches the actual movement visualized by the symbolized wheelchair in the map, corresponding to the known tendency for an actor's perspective to be adopted when action is involved in a spatial description [42]. However, at least for a subset of the movements involved, route perspective requires mental rotation, increasing the effort for both dialogue partners in the present scenario. Survey perspective involves no mental rotation but may be confusing in the present route instruction scenario, particularly since the symbolic wheelchair actually possesses an intrinsic front which then does not correspond to the externally based turn directions.

Much earlier research has addressed speakers' actual choices across various situations, establishing, for instance, that perspective choices are flexibly adapted to various kinds of contextual influences [41]. Moreover, speakers react subtly and systematically to their interaction partners' ability [29]. In a recent series of studies directly related to our endeavors reported here, Andonova and Coventry [3] used a restricted experimental setting focusing on simple two-leg routes (rather than complex route descriptions). In that setting, route perspective dominated overall while survey descriptions averaged only about a third of cases. The naïve speakers' choices of spatial perspective were influenced systematically by the perspective used by a confederate on the preceding trials. However, astonishingly little is known about how speakers actually handle perspective choice, and shifts of perspective, in the dynamic course of spatial interaction. Some findings by Garrod and Anderson [13], Schober [28], Filipi and Wales [8], and Healey and Mills [15] suggest that perspective choices and shifts are systematically related to communication problems: speakers offer a new conceptual perspective on a situation when there seems to be trouble interpreting the previous one. However, these findings also show very clearly that such interactional strategies are highly dependent on the interaction situation, which in each case differed fundamentally from our current design. While Schober's task design did not involve a map, the other studies used either one of the well-known Maze Game [13] or Map Task [2] paradigms—both of which involve a situation where the visual map information differs for the interactants and where there is no shared visual feedback concerning the current position within the map. These facts clearly (and intentionally) result in a high need for explicit spatial clarification.

Our scenario, in contrast, involves a relatively simple shared spatial context with direct visual feedback concerning the interpretation of a route instruction. The setting allows for two distinct perspectives, route and survey as described above, both of which are frequently used spontaneously by speakers as shown in a previous study [14]. Using a simplified setting allows for a better understanding of perspective strategies, as well as for a direct comparison with HCI. The analysis of HHI provides information about how human speakers act naturally in dialogue in a situation like this; accordingly, a dialogue system that operates in an optimally natural way should be capable of supporting similar speaker behavior. The analysis of HCI then illuminates the extent to which current users of dialogue systems expect such a support to be already available, as reflected by their spontaneous perspective choices when addressing the system.

One way of addressing how understanding is achieved by participants in dialogue is to take a closer look at clarification questions, which may indicate when and how understanding failures occur. They can be analyzed across various levels of communication [24], ranging from auditory problems (in spoken interaction) to reference resolution and other types of conceptual, semantic, and pragmatic issues that need to be clarified. In our current scenario, the role of the route follower is most decisive with respect to the potential need for clarification. Unlike route givers, route followers did not know where the goal was, and they had to make decisions since they had to steer the wheelchair. Route givers could simply watch the wheelchair's movement on the screen, and provide further instructions in case of errors, resulting in a reduced need for clarification. The route follower's role in our HHI scenario is particularly interesting with respect to the requirements for dialogue systems, which in the future should be capable of asking similar, intuitively suitable clarification questions.

2.3. General instruction strategies

Previous results on the conceptual strategies and hierarchies available for speakers in route giving [10,25,32] provide ample reasons to assume two basic options for our scenario: route-based turn-by-turn strategies with little need for reference to spatial entities, vs. destination-based survey perspective descriptions that directly represent the target location. While other options and variants are possible, we thus expect that survey perspective based descriptions should be associated with a higher amount of spatial references to locations. Moreover, direct reference to the destination location, rather than the route towards it, should result in a lower number of instructions, leading to an expected association of survey-based descriptions with fewer instructions. Taken together, the comparative analysis of granularity and perspective choices and switches and the associated clarification patterns highlight speakers' negotiation strategies in a route interaction scenario in HHI as opposed to HCI.

In the next section, before turning to our comparative study, we introduce some of the modelling decisions made in the design of the employed dialogue system. We do so for two reasons: first, to indicate explicitly how issues such as perspective and granularity in route dialogues are handled by the artificial route follower and second to illustrate some typical limitations of computational language interpretation systems which can give rise to miscommunication in route instruction contexts.

3. Computational modelling

To investigate route instruction dialogues in the human–computer domain, we made use of a computational system that was created to handle the interpretation of spatial information in route instructions. The system, named Navspace, plays the role of route follower in the scenario introduced earlier, and as such is responsible for: (a) interpreting the verbal information given by the route giver with respect to the environmental and contextual models; (b) acting on the information given by moving towards the destination as appropriate; and (c) composing confirmations or queries as necessary with respect to the route instructions given. Some of the key notions behind the interpretation process have been presented in [26], while the complete computational system as used here has been detailed in [27]. Therefore, in the current section, we limit ourselves to a brief overview of the Navspace system.

Fig. 2 presents a simplified description of the system's components. The user is provided with an interface that includes text-based components for interacting with the system. This same interface also provides a visual

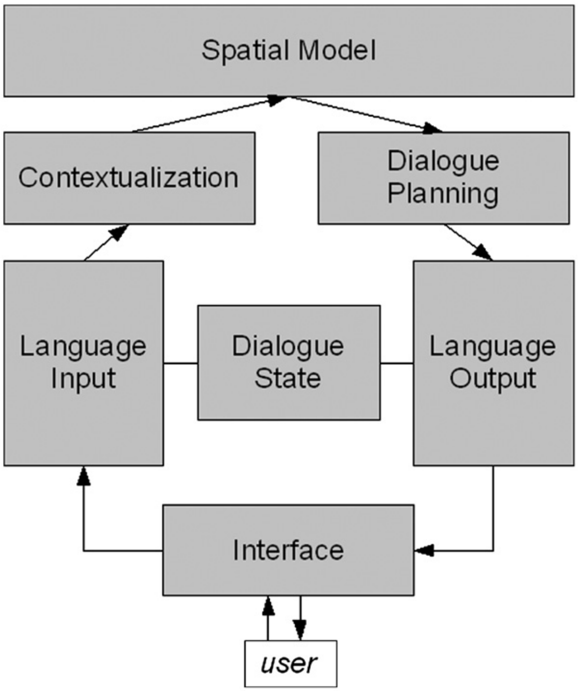

Fig. 2. Overview of the Navspace dialogue system components.

depiction of the agent's environment, as well as its pose in that environment. User utterance processing can be broadly split into a two-stage analysis. The first stage (Language Input) involves the assignment of a shallow linguistic semantics to each user utterance. This is achieved through the application of a Combinatory Categorial Grammar [31] which produces semantics given in terms of a well-defined linguistic ontology for spatial expressions [4]. The second stage of user utterance processing involves taking such logical descriptions, and applying discourse and situational context to construct the deep contextual meaning of the instruction. Here we achieve this construction through the application of a functional contextualization process. The process makes use of semantic category specific augmentation and resolution functions to compose quantitative interpretations of spatial meaning against the agent's spatial model. The result of this contextualization process is a fully specified logical description of the user's utterance where references have been resolved, elided content recovered, and a likelihood distribution of the agent's resultant location computed.

The production of system utterances follows a similar though reversed process. Namely, Dialogue Planning determines when the system can produce an utterance, and in turn constructs a shallow linguistic semantic specification for that utterance. This specification is then passed through language production grammars to derive a surface form which is then presented to the user through the interface. Dialogue modelling for the Navspace system is constructed around an information-state based account [39]; however, for the purposes of this paper, we applied a rudimentary dialogue strategy which simply accepted understood user instructions and statements, while producing appropriate error statements in the event of language analysis or contextualization failure.

Resources for language analysis and contextualization were configured for a range of spatial language phenomena observed in verbal route descriptions. Both simple motion expressions that denote movement based on directional information (e.g., *fahr nach links* [move to the left]), as well as path-descriptions which make reference to landmarks in the environment (e.g., *geh zu dem ersten Raum über der Küche* [go to the first room above the kitchen]) were handled by the Navspace system. Moreover, the interpretation models supported the modification of directional and path-descriptive terms with both quantitative and qualitative extent information (e.g., *beweg dich ein bisschen weiter* [move a little bit further]). Notably, in addition to handling fully qualified individual descriptions, the language interpreter could also process a range of highly elided instructions (e.g., *links* [left]), as well as the complexing of such individual instructions into instruction sequences. Furthermore, in addition to handling explicit movement instructions, the system could also handle statements that explicitly defined the target's location with respect to other entities in the environment or to the agent (e.g., *es ist der zweite Raum rechts* [it is the second room to the right]).

The above range of input language types ensured that the Navspace system facilitated the interpretation of route instructions that varied both in terms of granularity and perspective use. With respect to granularity, the system's ability to interpret directional and path descriptive instructions, along with its ability to accept instructions given in a relatively free order, allows a range of granularity strategies to be adopted by users. Meanwhile, with respect to perspective handling, the interpretation model could handle directional terms defined with respect to either the route or survey perspectives. For projective terms such as *left* and *right*, a default route-based interpretation was applied, but this default behavior was subsumed by more specific behaviors if either the affordances offered by the environment prevented such a default route interpretation, or if a user explicitly marked a perspective choice in the surface form.

4. Route communication in HHI vs. HCI: comparative study

Our aim in the present studies was twofold. On the one hand, we wished to gain more knowledge about speakers' intuitive strategies with respect to granularity and perspective when interacting dialogically about routes in a spatial setting. On the other hand, we aimed at comparing these strategies with respect to the current interaction partner. An important constraint was to provide a scenario that offered a meaningful communicative goal both for human–human and human–system interaction, enabling spontaneous and natural dialogue in both settings. Such a goal was provided by asking the participants to give route directions to a pre-defined location for a wheelchair avatar shown on a map, which could be steered by a person using a joystick, or automatically by the Navspace system (see Appendix A for an example instruction given to the participants). Our chat-based scenario in the human–human condition ensured that the communication mode resembled typed human–computer interaction as closely as possible.

In both scenarios, participants started out with a test run followed by ten trial routes. The tasks always appeared in the same order. They differed with respect to the distance to be covered between start (initial location of the depicted wheelchair) and goal location. The shortest route towards the goal was always clearly identifiable. The shortest routes for five of the trial tasks (trials nos. 1, 3, 4, 8, and 10) were below 500 programming internal units and were categorized as short; the other five routes were above 600 of such units and were categorized as long.

After the study, all of the participants filled in a questionnaire eliciting a range of impressions about the communication situation as well as the speakers' intuitive perspective preferences. In particular, the questions concerned the extent to which a route-based perspective and a survey-based perspective were considered helpful in a given situation, and, if forced to make a choice between the two types of description, what preference would emerge as the dominant one (see Appendix B). This questionnaire was identical for both studies, and allowed us to gain further insights about the extent to which participants were influenced by the communicative situation concerning their perspective preferences.

5. Study 1: Human–human interaction

5.1. Participants

The HHI study involved 15 pairs or *dyads* of two naïve participants (21 female and 9 male participants; ages ranged from 20 to 40, mean age 24.6 years). They were students at the University of Bremen (native speakers of German), who were compensated with 7.50 € per hour for their efforts.

5.2. Procedure

The participants were given written instructions containing a depiction of the map used in the study similar to the one shown in Fig. 1. These instructions ensured that the participants understood the goal of the study, namely to move the depicted wheelchair through the map towards the location marked in red, based on verbal instructions typed by the route givers into a chat interface. Only route givers could see the red markings. All participants were told that the first trial was a test run after which they were still allowed to ask questions, and that there would be ten more trials following the test run.

The two participants of a dyad were seated at two computer terminals in separate rooms, which showed the schematic map giving the simulated wheelchair's position. One of them was asked to imagine sitting in the wheelchair and to give instructions (using the chat interface) for the other to navigate towards a goal predefined by colour marking on their screen (but not on their partner's). The other participant was asked to: (a) imagine that their partner was sitting in the wheelchair; (b) steer the wheelchair with a joystick towards the goal according to their partner's instructions; and (c) communicate via the chat interface. The chat interface used by both participants made use of two typing activity indicators that signaled to a given participant when their partner was composing a dialogue contribution: a visual indicator in the form of a rotating egg-timer and an audible indicator in the form of the sound of typing. These activity indicators were intended to keep communication overlaps between the route giver and the route follower to a minimum. The messages given by the participants in the chat line disappeared after several seconds (depending on the message length).

5.3. Analysis

To address the research questions outlined above, we employed qualitative and quantitative data analysis identifying various dialogue features. Purely quantitative analyses such as duration of trials could be derived automatically. For the analysis of directional and other spatial term use, perspective choice and shifts, granularity, and clarification requests, we developed a coding schema as follows.

5.3.1. Directional and other spatial terms

For each instruction by the route giver, we coded whether it contained a directional (projective) term such as *straight ahead, right*, or *left*, as opposed to other types of spatial descriptions that did not project a direction from the wheelchair's current position, or non-spatial utterances such as *wait* or *yes*.

5.3.2. Perspective choice

We checked carefully whether each utterance could be identified as being based on either route or survey perspective in a two-stage filtering process. The first stage involved filtering for perspective cues; as laid out in [14], a range of German terms could be identified semantically as directly reflecting either survey or route perspective in the given setting, such as those listed in Table 1. Such unambiguous allocations are essential for the development of dialogue systems, since they can be used to support the identification of underlying perspectives in an interaction situation; however, they need to be adapted to the scenario at hand.

The majority of utterances, however, did not use these particular perspective cues. The remaining utterances were labelled ambiguous in the first pass, and then further analyzed in the second filtering stage. In that stage we additionally considered the current spatial situation and the given goal in order to interpret potentially ambiguous expressions. In particular, the orientation of the wheelchair played a crucial role in determining whether ambiguous terms like "left" or "right" were based on route or survey perspectives. When the wheelchair was facing towards the bottom of the screen, these terms could disambiguate spatial perspective use by considering where the goal was relative to the wheelchair and interpreting instructions according to the intended goal. Furthermore, subsequent movement of the wheelchair in one direction or the other would be quickly corrected if the perspective of the instruction had been misinterpreted. This second filtering stage involved the annotators watching screenshot videos of the realtime

Table 1
Direct linguistic indicators of perspective in the given setting.

Route perspective	Survey perspective
vom Rollstuhl/Fahrer aus [from the wheelchair/driver]	*von dir/mir aus gesehen* [from my/your point of view]
wieder (links / rechts) [again (left / right)]	*auf der Karte* [on the map]
hinter [behind]; *vor* [in front of]	*oben* [top]; *unten* [bottom]
vorwärts [forwards]; *rückwärts* [backwards]	*hoch* [up]; *runter* [down]

interaction of the participants, where the timing of instructions, movement and orientation all played a role in disambiguation. Nevertheless, not all of the utterances could be assigned a clear perspective. Altogether, our coding options were the following:

- route: the utterance can be clearly identified as using route perspective
- survey: the utterance can be clearly identified as using survey perspective
- mixed: the utterance contains spatial terms indicating both route and survey perspective
- unclear: the utterance contains spatial terms that cannot be interpreted using either survey or route perspective in the current situation
- conflated: the utterance contains terms that are compatible with both route and survey with respect to the current spatial situation
- without: the utterance contains no terms that could potentially indicate a spatial perspective.

5.3.3. *Perspective shifts*

We identified utterances as indicating shifts of perspective if they contrasted directly with respect to a previous utterance between clearly identified route and survey perspectives, produced by the same speaker on one and the same trial. This conservative way of defining perspective switches ensured that no ambiguous utterances got to be over-interpreted as representing speaker flexibility. Motivated by findings in the previous literature as described in Section 2.2 above indicating a lack of knowledge about when and how speakers switch, we furthermore searched for possible reasons for perspective switches qualitatively.

5.3.4. *Granularity*

We identified for each spatial instruction by a route giver whether it contained a reference to a spatial location, described by a concrete noun such as *kitchen, room, intersection*, etc. Further information about the level of detail in these location descriptions was gained by annotating whether these locations were anchored spatially. For instance, an utterance such as ”go past the kitchen” does not provide any information about the kitchen's location, whereas an utterance such as ”go past the kitchen on the right hand side” is more informative spatially.

Such location descriptions may refer to start locations, subgoals, or the destination itself. However, destination descriptions do not necessarily contain nouns referring to locations; they may remain implicit in elliptical phrases such as in *jetzt der vorletzte links* [now the second last on the left]. Since both of these inter-related aspects concern the issue of granularity, we next addressed destination descriptions independently of noun usage. We identified for each utterance whether it indicated a subgoal, the final destination, or none of these.

5.3.5. *Clarification requests*

All questions by the route follower were examined as to whether they asked about a prior utterance or action, or something implied from a prior instruction, or specific information about the action to be performed in the current turn. If one of these was the case, the question was labelled as a clarification request. In contrast, forward-looking questions like ”what next?” were not interpreted as clarification requests. Next, we investigated the types of information that the clarification requests queried. In our scenario, reference resolution attempts are particularly informative as they indicate problems with the spatiotemporal matching of an instruction to an action by the route followers. These included clarification questions asking which of several similar location-types the route giver referred to, e.g., ”which door?”, asking where a location which was referred to in a prior instruction was, e.g., ”opposite the lab?” or ”where do you mean?”, as well as deictic clarification questions such as ”this door?” or ”this one?”.

5.3.6. *Reliability*

We assessed inter-coder reliability for our annotations by getting a quarter of the data randomly selected by dyad, trial, and location in trial annotated by a second annotator who used the annotation manual. Scott's Pi, Cohen's Kappa and Krippendorff's Alpha were calculated for each of our annotation categories. The results can be seen in Table 2. Following the inter-coder reliability tests annotations were further revised and improved on the basis of annotators' agreement in order to achieve optimal reliability.

5.4. *Results*

5.4.1. *General results*

The collected corpus of 150 task dialogues contains 951 utterances in total, 870 of which were produced by

Table 2
Results of inter-annotator agreement tests on annotation categories in the HHI scenario.

Measure	Directional and other spatial terms	Nouns referring to spatial entities	Destination or subgoal descriptions	Perspective choice	Perspective shifts
Scott's Pi	0.61	0.79	0.68	0.57	0.17
Cohen's kappa	0.63	0.79	0.68	0.57	0.18
Krippendorff's alpha	0.61	0.79	0.68	0.57	0.17
Agreement (%)	76.2	89.5	85.9	70.7	89.1

route givers. 656 of these utterances were spatial descriptions by route givers; the other 214 utterances by route givers were reactions to route followers' actions such as ''exactly'', ''yes'', or ''no'', immediate movement instructions such as ''stop'', and the like. Unlike an earlier study [14,34], the participants in this study did not normally know each other and so did not engage in off-topic discourse. The average was 6.34 utterances per task dialogue and 63.4 per dyad. Here is one typical example of a HHI dialogue from our corpus:

Example 1.

Route giver:	*jetzt gerade hoch, dann links*	[now straight up, then left]
Route follower (moves the wheelchair until the first intersection of hallways, then asks):	*jetzt links?*	[now left?]
Route giver:	*nein weiter*	[no go on]
Route follower (moves the wheelchair further)		
Route giver:	*bis zum treppenhaus, das 3. zimmer auf der linken seite*	[to the stairs, the third room on the left side]
Route follower (moves the wheelchair directly to the destination)		

This short dialogue illustrates the range of phenomena that we are addressing. The initial instruction by the route giver contains mixed perspectives: the survey perspective indicator ''up'' is followed by the projective term ''left'', which is true for both survey and route perspectives in the current situation. Therefore this utterance is coded as ''mixed''. After having started moving the wheelchair avatar, the route follower asks a clarification question concerning the precise location of the left turn. Since the clarification question is spatiotemporally anchored as indicated by the deictic term ''now'' and the projective ''left'', it can only be interpreted by reference to the immediate spatiotemporal context. The question is answered by the route giver by a simple instruction to go on, followed by a fairly detailed and complex spatial description containing both a reference to a subgoal (''stairs''), and a description of the spatial location of the final destination (''third room on the left side''). This instruction directly leads to success. In the following we will present the results of our analysis of the whole corpus of route dialogues with respect to these phenomena.

5.4.2. Perspective

The corpus contains 514 utterances (54.05% of all 951 utterances) that indicated spatial perspective. 497 of these were produced by route givers (57.13% of all 870 route giver utterances). Of these, 297 (59.76% of 497) were clearly based on route perspective, 79 (15.90%) on survey, 83 (16.70%) conflated, 25 (5.03%) unclear, and 13 (2.62%) mixed. Thus, 376 (43.22% out of all 870) utterances by route givers clearly indicated a single perspective, either route or survey. Of the 17 utterances produced by route followers that indicated a perspective, 10 were clearly based on route perspective, and 7 were unclear.

Perspective choices changed with time—there was a negative correlation between trial number and the mean percent use of the route perspective, i.e., in later dialogues, route givers tended to produce more instructions in the survey perspective and fewer route perspective instructions ($r=.-162$, $p=0.048$). This result replicates previous findings in [14]. An analysis on the correlation between the percent use of route perspective and the number of utterances produced by each dyad showed that the number of utterances produced by the route giver was positively correlated with percent use of route perspective ($r=0.41$, $p<0.001$), indicating that route-based strategies may require more instructions than survey-based instructions in this scenario, corresponding to our expectations. Finally, we compared participants' questionnaire responses about perspective preferences (see Appendix B) with the verbal data measures in order to establish if their conscious choices are those found in their dialogic contributions. Our correlation analyses showed a high degree of inconsistency between the participants' binary choices of perspectives and their assessments of relative usefulness; moreover, their subjective ratings of spatial perspective generally did not reflect their real choices in this corpus, as indicated by a lack of correlation between the route givers' responses on the helpfulness questions and their mean percent use of the route perspective.

We identified 41 cases of same-speaker perspective switches (defined in the conservative way described above), produced by 9 of the 15 route givers. As in earlier findings [28], one reason for a perspective switch may be that speakers change their conceptual strategy after a misunderstanding. For example, an erroneous action by the route follower could be followed by a perspective switch by the route giver, as in:

Example 2.

Route giver:	*und zwar der rechtere*	[namely the one on the right]	SURVEY PERSPECTIVE
Route giver:	*nein*	[no]	
Route giver:	*zurück*	[back]	ROUTE PERSPECTIVE

However, our data do not support any claims as to perspective switch being a standard strategy for dealing with misunderstandings, since of the 21 explicit rejections (''no'') by the route giver, only three were followed by a perspective shift (the others were typically followed by a different wording of the previously misunderstood instruction, or by providing more details). In order to shed more light on these interactional issues, we investigated the relationship between perspective use by route givers and clarification requests by route followers, to which we turn next.

5.4.3. Clarification requests

Of the 81 utterances produced by route followers, 49 (60.49%) were identified as clarification requests. Of these, 28 (57.14% of 49) were identified as reference resolution attempts. Here is an example of a reference resolution attempt that aims to clarify the intent of the instruction by rephrasing it using a label given in the map:

Example 3.

Route giver:	dritte tür rechts ist das ziel	[third door on the right is the goal]
Route giver:	nein	[no]
Route giver:	*in fahrtrichtung rechts meinte ich*	[I meant right from the moving direction]
Route follower:	also küche?	[kitchen then?]
Route giver:	nein	[no]
Route giver:	eins weiter als küche	[one further from the kitchen]

In this case the route giver clarifies a previous instruction, and the route follower requests clarification of the route giver's self-clarification. There are 9 such cases (32.14% of all reference resolution attempts) in which clarification requests mention labelled locations, apparently representing their best guesses as to which room was currently being referred to by the route giver. So one spatial strategy of route followers is to suggest labelled locations as possible stepping stones along the route to the final goal. Two of these suggestions were directly confirmed by the route givers, and in two other cases route givers reacted by refining their original instruction using the labelled location as a subgoal, as in Example 3.

Clarification requests which were not reference resolution attempts included requests for confirmation such as "like this?", and requests for repetition. A particularly *spatial* type of clarification request is one where the perspective choice of the route giver is explicitly or implicitly questioned, e.g.: "to the right in the direction of motion?", or "from the driver's perspective, yeah?". Sometimes they even mention both perspectives explicitly, as in: "right from my perspective or that of the wheelchair?". These perspective clarification requests can also be subtle, e.g. "where left? In the direction of the mail room?", or "which side? copy room?", where the spatial context needs to be considered in order to determine whether these clarification requests ask for perspective to be clarified. For example, downward-facing orientation lends itself to perspective ambiguities in "left"/"right" instructions, as these are resolved to opposite directions based on whether route or survey perspective is assumed. Note that some of these perspective clarification requests concern reference resolution attempts, since spatial location references are based on perspective just in the same way as motion directions. Altogether we found seven (explicit and implicit) perspective clarification requests, i.e., these constitute 14.29% of the clarification requests found in the data.

One might then wonder whether these clarification requests occur as a consequence of the preceding spatial turn communicating mixed or unclear perspective. There were 38 cases of unclear or mixed perspective produced by route givers in the data; of these, 10 (26.32%), were followed by clarification requests before a different instruction containing a clear perspective was given. Not all of these clarification requests were perspective clarification requests; however, the possibility remains that ambiguous perspective usage adds to the uncertainty of the route follower in their interpretation and results in a clarification request.

Correspondingly, one might also wonder whether clarification requests in general predict perspective change in the route giver in their subsequent utterance, where the switch can be seen as a collaborative response to the route follower's lack of understanding. In our corpus, we identified two individual cases where route givers switched perspective in the subsequent spatial turn following a clarification request.

5.4.4. Granularity

5.4.4.1. Noun usage. 312 (47.56% of all 656 spatial instructions) utterances contained at least one noun referring to a spatial entity, such as *hallway, room, intersection* and the like. The locations of 171 (54.81%) of these were specified. Most of the utterances containing references to spatial locations were fairly complex; they were either descriptions of the precise location of the final destination (which was never labelled), or they used the locations as landmarks during the wayfinding process, as in *nach dem Treppenhaus rechts* [to the right after the staircase]. There were only 15 cases in which the preposition *zu* [to] was used in the spatial instructions together with a location label given in the map (such as *zum Treppenhaus* [to the staircase]), representing simple directions to a labelled subgoal. 13 of these contained no indication that this was not the final goal. 9 further cases referred to the direction of a labelled subgoal, as in *Richtung Küche* [in the direction of the kitchen]. There were no cases in which a labelled subgoal was mentioned without a spatial preposition (as we will see later, this contrasts with the HCI situation).

5.4.4.2. Destination descriptions. Altogether, 175 utterances by route givers (26.68% of all 656 spatial instructions) contained a destination description (18 of these without noun usage). Many of these were linguistically complex, such as *jetzt bitte den ersten raum nach der küche, rechts, direkt an der ecke* [now please the first room after the kitchen, on the right, directly at the corner]. Most destination descriptions occurred after directing the route follower incrementally towards the goal, as in the following sequence of instructions (to which the route follower only reacted nonverbally, i.e., by moving the wheelchair):

Example 4. *links — dann wieder rechts — und rechts in den nächsten gang — dann die zweite tür links*

[left — then again right — and to the right into the next hallway — then the second door on the left]

However, some of the instructions directly started out with a destination description. Of the 150 *first* instructions in the collected dialogues, 25 (16.67%) already referred to the final destination. Remarkably, although no dyads consistently used this method as a strategy throughout all of the tasks, only three of the 15 dyads never used it at all. Such dialogues did not require much negotiation; in eight of these 25 cases, there was no further verbal exchange at all once the destination was described in the first utterance, contrasting with the average of 6.34 utterances per task mentioned above. Consistent with this observation, our correlation analyses revealed that higher numbers of nouns referring to a spatial entity, as well as destination descriptions produced by the route givers, were associated with fewer utterances within a task dialogue ($r=-0.60$, $p<0.001$ for mean percent use of nouns referring to a spatial entity, and $r=-0.61$, $p<0.001$), and also with a lower total amount of words within a trial. In other words, the more route givers used destination references on a trial, the fewer words they said on a trial ($r=-0.26$, $p=0.001$).

5.4.5. General instruction strategies

The analysis of the data revealed two sets of variables that were positively correlated within each set and negatively correlated across sets. This kind of pattern underlines two general strategies that information givers used in their instructions. A more incremental step-by-step strategy relied on extensive use of projective terms and the route perspective: as expected, route perspective was associated with less use of destination reference ($r=-0.35$, $p<0.001$) and nouns referring to a spatial entity ($r=-0.27$, $p=0.001$) but increased use of projective terms ($r=0.32$, $p<0.001$), and vice versa, the more goal-based strategy was related to more use of destination reference and (naturally related to this) nouns referring to a spatial entity, and to less use of route perspective descriptions. Furthermore, while instructions in a more goal-based strategy were on average shorter in terms of number of route giver utterances, instructions in the route perspective and instructions with a higher percentage of perspective shift generally required a higher number of utterances ($r=0.20$, $p=0.016$, and $r=0.41$, $p<0.001$, respectively). Route length did not have a significant effect on percentage use of destination descriptions, nouns referring to a spatial entity, or perspective choice.

6. Study 2: Human–computer interaction

The second study made use of our dialogue system described in Section 3 above to play the role of route follower in a scenario which otherwise matched that presented in Section 5.2 for human–human interaction. The dialogue system itself was equipped with the capacity to interpret simple spatial movement instructions which facilitated variation in both granularity and perspective choice. The study was carried out for two reasons. On the one hand, we wished to evaluate the current status of the dialogue system by investigating how easy it would be for users unfamiliar with the system to navigate the virtual wheelchair to a pre-defined destination. As part of this purpose, we improved the system after testing a first group of participants and then tested the system further with a limited number of new users. On the other hand, we aimed to compare the linguistic data collected in this study with the HHI data as described above.

6.1. Method

6.1.1. Participants

21 participants (ages ranging 20–33; mean age 23.9 years; 13 male and 8 female) took part in this study. They were students at the University of Bremen (native speakers of German), who were compensated with 7.50 € per hour for their efforts. All participants were naïve in that they did not know how the dialogue system worked; 2 described themselves as having studied linguistics previously, while 7 described themselves as having some experience of robotics or dialogue systems. Group A consisted of 15 individuals who participated in a first study. Group B, consisting of the remaining 6 participants, took part in a second study phase where slight modifications of the system – particularly the interpretation grammars – had been made on the basis of the first phase of the study. Since the two groups were tested using different versions of the system, we report the results separately. Statistical analyses on user strategies were only run on the first (larger) data set, as the second was primarily added to evaluate the next increment in the system development.

6.1.2. Procedure

Individual participants were placed at a terminal running the Navspace system. As with the human–human study, each participant was given written instructions informing them of their objectives (see Appendix A). These instructions were almost identical to those given to the route giver in Study 1 (HHI). The one notable exception was that instead of making explicit reference to their partner as performing the route follower role, it was made clear that an artificial agent would be performing that role. No priming of the user's language was allowed. Thus, no examples were given in the instructions, nor did experimenters suggest language usage to participants.

Participants took part in 11 trials, the first trial being a test run after which questions could be directed to the experimenter. For each trial the wheelchair avatar was positioned within a room in the map and a target location was highlighted in red for the participant to see. The same 11 start and end points were used as in the human–human study, in the same sequence. However, unlike in the human–human trial, a time-out mechanism was included in the experiment scenario to (a) encourage participants to complete the task in a timely fashion and (b) to prevent participants becoming stuck in a trial due to failure of the dialogue system. The time-out was set for 3 min (apart from the test trial), after which the trial was aborted and the next trial loaded. It should be noted however, that no system crashes occurred for any of the 21 participants in this study.

6.2. Analysis

The analysis was carried out in the same way as in Study 1 (HHI); however, no analysis of clarification requests was done since these requests were uttered by the route follower, which in the HCI study was represented by the dialogue system.

6.3. Results

6.3.1. General results

In Group A consisting of 15 participants, we collected 150 task dialogues with 1818 utterances in total. On average, this yields 12.12 utterances per dialogue, 6.02 of which were produced by the human users. 903 of the 1818 utterances were produced by the users; 825 of these were spatial instructions (the remaining 78 utterances were mainly ”stop” and ”go”). In Group B consisting of 6 participants, we collected 60 task dialogues with 708 utterances in total, 355 of which were produced by users, with 267 of these spatial instructions. This yields 11.8 utterances per dialogue, 5.92 of which were produced by the users.

All 21 participants completed all tasks successfully. The average time taken per task was 63.27 s, with an average time for the initial task per user of 150.57 s, and an average time per final task of 48.81 s. This time compared favourably with the times taken for the human–human dyads completing the same tasks (70.5, 114.73, and 37.57 s, respectively). In particular, the system improvement appeared to be successful, with task duration dropping from 65.53 s for Group A to 58.21 s per task for Group B. Even more strikingly, the rate of language processing error fell from 0.95 errors per task for Group A to 0.59 errors per task for Group B.

User satisfaction rates with the task were moderate to good based on the post-study questionnaire. 19 of 21 participants agreed, or agreed strongly that the system understood their instructions, while 14 of 21 believed that the system behaved in a way corresponding to what they expected.

The user utterances in this study were overwhelmingly sparse, containing repetitions of simple movement instructions such as *links* [left], *rechts* [right], *geradeaus* [straight ahead], *stop* [stop] and the like. In fact, 756 (83.72%) of the user utterances in Group A and 212 (59.72%) of the user utterances in Group B consisted of only one single word, although the system could understand longer utterances as explained above, and in fact successfully handled commonly used instructions such as *dreh dich nach rechts* [turn to the right]. Here is one typical example of an HCI dialogue from our corpus:

Example 5.

User:	*los*	[go]
System:	*O.K.*	[O.K.]
User:	*right*	[right (in English)]
System:	*Entschuldigung, koennen Sie das vielleicht umformulieren?*	[Sorry, could you perhaps re-formulate?]
User:	rechts	[right]
System:	*O.K.*	[O.K.]
User:	*umdrehen*	[turn around]
System:	*O.K.*	[O.K.]
User:	*links*	[left]
System:	*O.K.*	[O.K.]

6.3.2. Perspective

In Group A, 666 utterances (73.75% of all 903 user utterances) were based on some kind of perspective. Of these, 523 (78.53% of 666) were clearly based on route perspective, 8 (1.20%) on survey, 130 (19.52%) conflated, 5 (0.75%) unclear, and none mixed. Thus, 531 (58.80% out of all 903) user utterances in this corpus clearly indicated a single perspective, either route or survey, and here route perspective clearly dominated with an average of 97.17%. The questionnaires were consistent with this preference; we found a positive correlation between the mean percent use of route perspective and the preference for the route example on the binary question (r=0.18, p=0.029). We identified 12 cases of perspective switches, produced by 5 of the 15 route givers.

In Group B, 262 utterances (73.8% of all 355 user utterances) were based on some kind of perspective. Of these, 200 (76.34% of 262) were clearly based on route perspective, 6 (2.29%) on survey, 54 (20.61%) conflated, 1 (0.38%) unclear, and one mixed. Thus, 206 (58.03% out of all 355) user utterances in this corpus clearly indicated a single perspective, either route or survey. We identified 8 cases of perspective switches, produced by 2 of the 6 route givers.

6.3.3. Granularity

6.3.3.1. Noun usage. In Group A, 35 utterances (4.24% of all 825 spatial instructions) contained at least one noun referring to a spatial entity. The location of these entities was never specified, and the instructions containing nouns were never complex; in the vast majority (28 of 35) of these cases, the complete instruction only consisted of one of the location labels given in the map, without a spatial preposition or anything else (e.g., *Küche* [kitchen]). In three cases the preposition *zu* [to] was used together with a location label, and the remaining four contained instructions for the wheelchair to leave the room. So most of these utterances using nouns referring to a spatial entity represented simple directions to a labelled subgoal, without indication that this was not the final goal. There were no references to the direction of a labelled subgoal, as in *Richtung Küche* [direction of the kitchen].

In Group B, 7 utterances (2.62% of all 267 spatial instructions) contained at least one noun referring to a spatial entity. All of them were produced by the same user, and they were actually slightly more complex than those seen in Group A, encompassing both *nach rechts in den raum fahren* [go to the right into the room] and *in richtung küche* [in the direction of the kitchen].

6.3.3.2. Destination descriptions. In Group A, two utterances contained a destination description, albeit an invalid one, as the user was explicitly told that the system was unable to perceive ”the red room” to which the user referred. Both of these utterances were initial descriptions. All other references to spatial entities pointed to subgoals. In Group B, three of the utterances containing a reference to a spatial entity as just mentioned were non-initial destination descriptions, and the other four referred to subgoals. There were no destination descriptions without noun usage in either of the groups.

6.3.4. General instruction strategies

Strategies did not change across the time span of the interaction; there was no correlation between trial number and any of the granularity (percent use of nouns referring to a spatial entity, percent reference to destinations) or other spatial strategy (percent route perspective, percent projective term use) measures. Along with the strong preference for route perspective (M=97.17%), there was a high percentage of projective terms (M=93.94%), and a low percentage of nouns referring to a spatial entity (M=4.77%) as well as destination reference (M=0.27%) in route givers' utterances throughout the interaction. Furthermore, higher helpfulness ratings on the route perspective example in the questionnaire were associated with more use of projective terms (r=0.47, $p<0.001$), less use of nouns referring to a spatial entity ($r=-0.44$, $p<0.001$) and fewer instructions generally ($r=-0.28$, p=0.001), though not with more use of route perspective during the interaction. In addition, higher helpfulness ratings on the survey perspective example were negatively correlated with more use of nouns referring to a spatial entity ($r=-0.20$, p=0.015).

6.3.5. Test runs

We hypothesized that participants may have gained experience with the dialogue system already in the initial test run, which may then have led them to ignore other communicative options in later trials. To gain insights about their initial ideas, we considered the test runs separately. The test runs of both groups comprised a corpus of 258 user utterances. Of these, 102 had clearly identifiable perspectives. While the majority (94 utterances) were route perspective, 8 (7.84%) were survey. 202 of the user utterances were spatial instructions. Of these, 39 (17.33%) contained a noun referring to a spatial entity, the location of which was described in the utterance in four cases. 6 utterances were final destination descriptions, and 30 referred to a subgoal. Thus, the variety appeared to be slightly higher than in the analyzed trials, particularly with respect to granularity.

7. Comparison of user strategies in HHI vs. HCI

From our separate results of the two dialogue situations, it is obvious that fundamentally different communication types emerged; human speakers reacted systematically to the nature of their interaction partner (human or system). While task completion times were comparable to the HHI case, the contents of the users' descriptions differed considerably from those of the route givers in the HHI situation, and the utterances produced for a human partner were generally more complex than those produced for the system. Here we provide a direct comparison.

A series of two-tailed t-tests for independent samples examined the HHI and HCI data (only the larger Group A) averaged per trial and per dyad for each of the variables. Responses in the two tasks differed significantly on most of these measures except in terms of the average number of route givers' utterances. Generally, there was less variability in the HCI than in the HHI data. Route givers in the HHI task produced fewer projective term utterances (72.97% vs. 93.94%, $t(28)$=8.55, $p<0.001$), more nouns referring to a spatial entity (56.75% vs. 4.77%, $t(28)$=19.55, $p<0.001$), more destination reference utterances (36.56% vs. 0.27%, $t(28)$=16.27, $p<0.001$), fewer route perspective utterances (67.71% vs. 97.19%, $t(28)$=8.04, $p<0.001$), and more perspective shift utterances (7.03% vs. 1.37%, $t(28)$=3.96, $p<0.001$).

Table 3 gives an overview of the main results of the relative frequency analysis for the HHI and HCI data sets, comparing a subset of linguistic features related in each case to the relevant subset of utterances. For the purposes of this comparison we combine Groups A and B of the HCI study, since the distinct analyses of these groups reported above consistently revealed similar distributions of linguistic features both before and after the system improvement. Thus, the main outcome of the modification was the reduced language processing error rate and

Table 3
Main features of the human–human and human–computer interaction data.

	HHI		HCI (Groups A and B combined)	
	No. of cases	%	No. of cases	%
Perspective based utterances (out of all utterances produced by route givers)	497 of 870	57.13	928 of 1258	73.77
Route perspective (out of all perspective-based utterances produced by route givers)	297 of 497	59.76	723 of 928	77.91
Survey perspective (out of all perspective-based utterances by route givers)	79 of 497	15.90	14 of 928	1.51
Single-word utterances (out of all utterances by route givers)	304 of 870	34.94	968 of 1258	76.95
Utterances containing nouns referring to a spatial entity (out of all spatial instructions by route givers)	312 of 656	47.56	42 of 1092	3.85
Utterances containing a spatial description for a noun referring to a spatial entity (out of all utterances with nouns referring to a spatial entity)	171 of 312	54.81	5 of 42	11.90
Destination descriptions (out of all spatial instructions by route givers)	175 of 656	26.68	5 of 1092	0.46
Initial destination descriptions (out of all initial instructions by route givers)	25 of 150	16.67	2 of 210	0.95

task duration in Group B reported above, with no discernible effects on user strategy choices.

Finally, we also compared the ratings given by route givers in the human–human and human–computer interaction on the three questionnaire items concerning spatial perspective choices. The mean helpfulness rating differed for the route perspective example, $t(27)=7.23$, $p<0.001$, the survey perspective examples, $t(27)=5.57$, $p<0.001$, and the binary choice between route and survey perspective, $t(26)=8.72$, $p<0.001$. Route givers in the human–human interaction task gave higher helpfulness ratings to the survey perspective example than those in the human–computer interaction (3.36 vs. 2.60 on average), and lower helpfulness ratings to the route perspective example (2.14 vs. 3.00 on average). They also generally preferred the survey perspective example to the route perspective one, more so than the route givers in the human–computer interaction task (60% vs. 15%).

8. Discussion

In order to investigate speakers' communication strategies and conceptual choices of perspective and granularity levels when interacting with dialogue systems and with human partners, we carried out two studies involving map-based linguistic interaction. Our results show systematic patterns on various levels, emerging particularly clearly in the comparison of the two settings. The human–human interaction scenario provides new insights concerning speakers' choices and negotiation processes in spatial dialogue. The human–system interaction scenario reveals how speakers deal with a somewhat unusual communication situation in which they are asked to convey a route instruction to a dialogue system with capabilities unknown to them. The contrastive analysis highlights the fundamental difference between speakers' concepts of these distinct communication situations, and their efficient strategies of dealing with each of these.

In both settings, route perspective (imagining being inside the scene and moving through the hallways with the wheelchair) was generally preferred throughout, corresponding to the task instructions given to the participants. However, in the HHI study most of the speakers occasionally switched perspective choices. Similar to earlier studies in spatial communication [8,15], perspective shifts could be triggered by misunderstandings and mistakes; in that case, the route giver may feel a need to re-represent the spatial description in a different way. The analysis of clarification requests further revealed that route followers asked for a clarification of perspective, albeit in sometimes subtle ways, several times in reaction to a previous spatial description that was in fact unclear with respect to perspective. In some cases, clarification requests as well as explicit rejections (''no'') by route givers (both indicators of communication problems) could trigger perspective shifts.

In the HCI case, miscommunication was also common. Although task-completion and satisfaction rates were high, the dialogue system was prone to coverage limitations which resulted in the user having to reformulate their instructions in many cases. As with related studies that have considered the verbal presentation of spatial information, e.g., MacMahon's monologic analysis of in-advance route instructions [18], the causes of miscommunication were typically due to participants using words or syntactic constructions unknown to the dialogue system. In part, these limitations can be overcome through an iterative approach to linguistic resource construction [43]. In our study, even a modest increment to language processing grammars led to improvement both with respect to the rate of language processing error and with respect to task duration.

In spite of the need for reformulations and clarification, switches to survey perspective were extremely rare in the HCI study. Instead, participants of both HCI groups – before and after the system improvement – quickly converged on a simple spatial information production strategy. As a result, the HCI and HHI data sets differed considerably in the area of granularity, not only with respect to the frequency of references to locations, but also with respect to the particular way in which these were employed. Speakers talking to human partners frequently employed references to labelled places embedded in more complex spatial references, often as part of direct destination descriptions which made communication very easy. This increased with later trials, possibly based on experience. Thus, the HHI dialogues add to earlier findings on route giver flexibility in monologic settings [36]. In the HCI situation, in contrast, users almost exclusively (and uniformly) relied on route-perspective based turn-by-turn instructions containing no nouns referring to a spatial entity. Consistent with this, they expressed a clearer preference for route perspective subsequent to the interaction. Mention of locations was restricted to simple references to subgoals as ''stepping stones'' in order to reach the goal incrementally. This indicates a constantly low level of granularity, similar to earlier findings on user strategies for spatial communication with robots (e.g., [10]). Interestingly, the analysis of reference resolution attempts by route followers in the human–human interaction situation revealed that using labelled locations as subgoals was for them, as well, a useful strategy supporting the efficient negotiation of routes.

The additional analysis of test runs showed that the frequencies of more sophisticated conceptual choices were higher in the test runs than the average in the actual trials, though still much lower than in HHI. So it stands to reason that some users actually tried out strategies in the test run that they quickly abandoned—in spite of the fact that the system was, in fact, capable of interpreting more complex descriptions. It seems that the inevitable coverage limitations of automatic systems systematically make users stick to rather simplistic linguistic strategies, which work well and thus reduce the effort required to figure out what else might work in the present scenario. This result is similar to earlier findings by Fischer and Moratz [10], who pointed to a conceptual hierarchy that system users apparently descend – from high level goal-based instructions to rudimentary movement commands – in the case of communication failure. The implication of this observation is that system users apparently expect systems to have only low-level comprehension capabilities. Therefore, if communication fails, the most likely explanation hypothesized by users will be that the conceptual level of instruction was too high, resulting in simplified discourse strategies. Similar to the findings by Fischer and Moratz, users in our scenario replaced references to goals and subgoals by (syntactically as well as conceptually) simpler turn-by-turn instructions in later utterances.

Thus, speakers adapted to the automatic dialogue system as an interaction partner on several levels. They consistently employed simple syntax with reduced spatial content from the start, along with limited location references, and a reluctance to switch perspectives. Apparently, current users of such systems assume (typically correctly) that they cannot employ complex spatial descriptions resembling those used regularly in human–human interaction. As such, this result is not surprising given earlier results on users' adaptation to systems as interaction partners [1,16]—however, the specific impact on the crucial spatial issues of perspective and granularity has not been identified in this way before. While humans are known to be particularly flexible in these areas [36,41], the present study has identified the existence of simple default options that are apparently quite unanimously felt to be suitable for automatic dialogue systems. Such low-level strategies are in fact useful as they exclude misunderstandings due to perspective switches, or to clashes with respect to the chosen level of granularity. Speakers appear to use such simple linguistic problem avoidance strategies intuitively, even if they lack earlier experience with the system as in the present study. However, humans' natural interaction strategies in spatial settings allow for far more flexible communication, including strategies for clarification and adaptation that ultimately lead to enhanced efficiency (such as switches to direct destination descriptions). Such flexibility also corresponds more closely to human mental hierarchical structuring of environments [32]. Crucially for embedded navigation settings, such processes reduce the need to supervise and control the route follower's actions on a turn-by-turn basis.

The rapid developments in current dialogue systems research will hopefully render the strong differences found between the current study settings obsolete fairly quickly. Users will increasingly employ natural dialogue strategies intuitively, as soon as efficient negotiation strategies become established as an essential component of automatic systems.

9. Conclusion

We presented the results of two studies investigating route directions with a map, first with human–human dyads interacting via a chat interface (HHI), second with individual human users interacting with a dialogue system (HCI). Results showed systematic differences between these two cases concerning both choice of perspective and level of granularity. While speakers in the HHI situation employed a wide range of communicative strategies in order to reach their goals efficiently, users in HCI typically relied very consistently on simple turn-by-turn instructions. We conclude that, when confronted with an automatic system equipped with limited capabilities, speakers restrict their linguistic choices to a fairly limited subset of the options generally available to them. This affects not only language (syntactic and semantic range) but also the spatial and conceptual aspects of the navigational setting: this leads, for instance, to a re-interpretation of landmarks to become subgoals (in HCI) rather than orientation aids (in HHI). As an outcome, human–computer interaction remains artificial, awkward, and inflexible. Since the scope of our investigation was limited to the specific situation of map-based interaction, future investigations will need to address the extent to which such patterns might transfer to real-world navigation scenarios, or to route instructions that are based on a 3D virtual reality environment.

For natural interaction to run efficiently, the employment of suitable clarification strategies and feedback by the system [9,44] should encourage users to widen the scope of their linguistic strategies, gradually moving towards more flexible interaction. Formal dialogue models developed for this purpose are presented, for example, in [30]. Further ongoing work concerns the controlled investigation of alignment and misalignment, plus a better understanding of speakers' repair strategies in cases of communication failure based on mismatches of perspective and granularity levels [35].

Acknowledgements

Funding by the DFG, SFB/TR 8 Spatial Cognition, project I5-[DiaSpace], is gratefully acknowledged. The DFG provided financial support for the conduct of this research and the preparation of the article, but had no involvement with the contents and decisions regarding the study design, analysis, and publication. We thank our student assistants (Anna Strotseva, Jan Hinsch, Bastian Ohmstede, Aziza Nawazir, Lea Fremann) for their support, and extend our thanks to John Bateman, Shi Hui, and Juliana Goschler, who were involved in this endeavour at different stages.

Appendix A

Example instruction: Translation of the instruction given in German to the participants in the HCI study. The instructions in the HHI scenario corresponded to this instruction as far as possible; instructions differed only with respect to the different technical and practical implications of the HHI vs. HCI situations. Route givers and route followers (in the HHI study) were given separate instructions adapted to their roles.

Welcome to our study!

Today you will navigate with a wheelchair through a virtual building with corridors and floors. On the screen you will see a map of this building (Fig. A1).

Imagine you are sitting in a wheelchair which can drive independently and understand instructions. The small wheelchair in the picture below will indicate your present position.

A red area in the map will indicate your destination. This is the area to which you should navigate with the wheelchair.

The wheelchair can move independently on the screen. You are now asked to interact with the wheelchair to tell it where you should go. You do this by typing instructions through the computer's keyboard. These will be displayed at the bottom of the screen.

The wheelchair knows the map and its current position, and it can understand your instructions; however, it does not know anything about the red target area. You must explain to the wheelchair where you must go. The wheelchair can at any time respond to you or pose questions. This information will also appear at the bottom of the screen.

Please try to navigate as quickly as possible to the red area. When you reach the goal, the current map will disappear and a test pattern will appear on the screen. After this, a map with a new initial position and a new goal will appear. Once this appears on the screen, you may begin. In total you are required to travel to eleven goals.

The first trial is a test run without any time-out. You can ask the experimenter questions during this trial if anything is unclear.

During the following ten trials the experimenter cannot answer any questions. Also there will be a time-out of 3 min per trial. If the wheelchair does not reach the destination before the time-out expires, then the next trial will be started automatically.

Have fun!

Appendix B

The following three questions were given to the participants at the end of the study in order to assess their general perspective preferences.

In the diagram below your position is represented by the wheelchair. The target destination is denoted by the star. How useful do you find the following description of this situation?

"The target is all the way to your left." (Fig. A2)

Very inaccurate	Inaccurate	Moderate	Accurate	Very accurate

How useful do you find the following statement in the same situation?

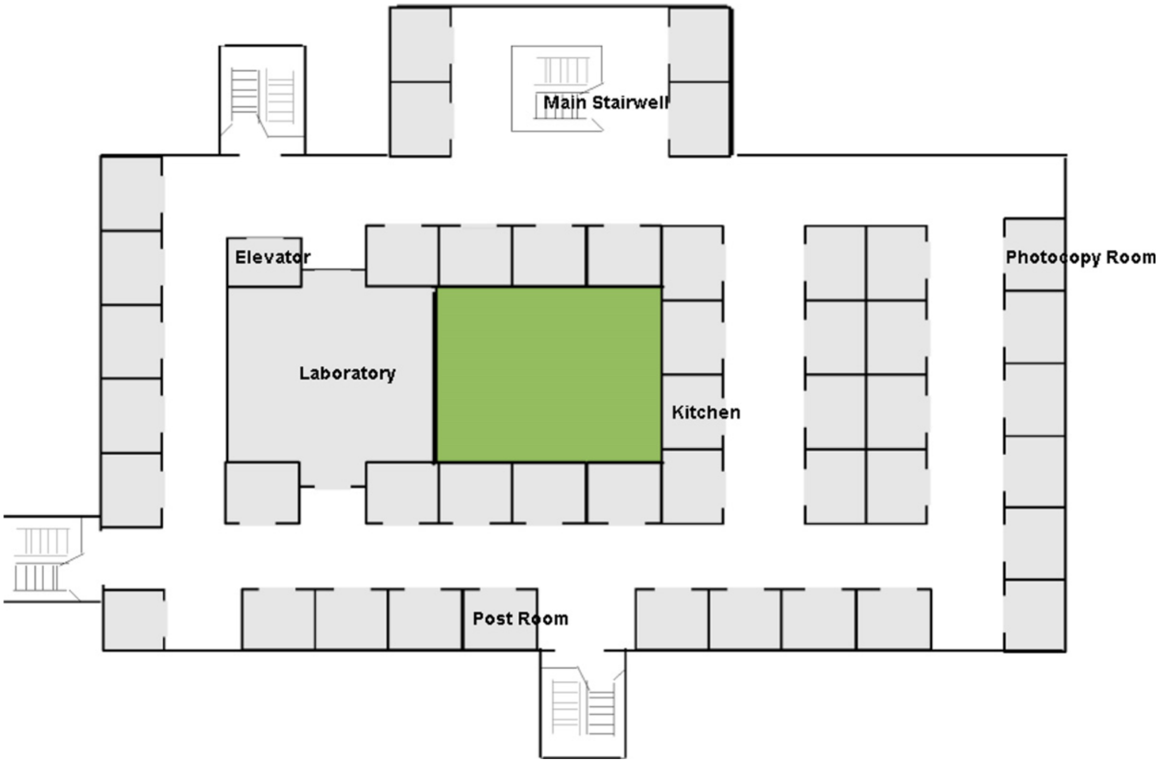

Fig. A1.

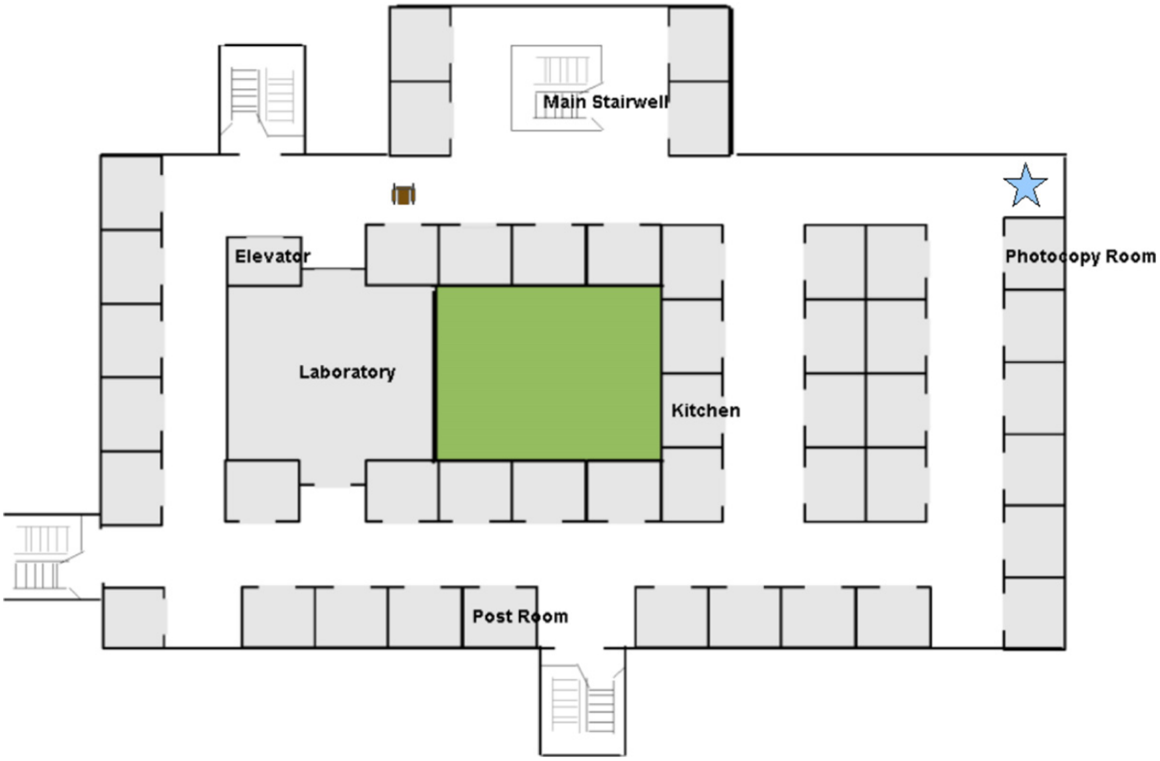

Fig. A2.

"The target is to the right in the image."

Very inaccurate	Inaccurate	Moderate	Accurate	Very accurate

If you had to decide: which of the following two descriptions would you find most useful to reach the destination?

(a) "The target is all the way to your left."
(b) "The target is to the right in the image."

References

[1] R. Amalberti, N. Carbonell, P. Falzon, User representations of computer systems in human–computer speech interaction, International Journal of Man–Machine Studies 38 (1993) 547–566.

[2] A.H. Anderson, M. Bader, E.G. Bard, E. Boyle, G. Doherty, S. Garrod, S. Isard, J. Kowtko, J. McAllister, J. Miller, C. Sotillo, H. Thompson, R. Weinert, The HCRC Map Task corpus, Language and Speech 34 (4) (1991) 351–366.

[3] E. Andonova, K. Coventry, Perspective priming in spatial descriptions, in: Proceedings of the 14th Annual Conference on Architectures and Mechanisms for Language Processing, Cambridge, UK, 2008.

[4] J. Bateman, J. Hois, R. Ross, T. Tenbrink, A linguistic ontology of space for natural language processing, Artificial Intelligence 174 (2010) 1027–1071.

[5] H.H. Clark, How do real people communicate with virtual partners? in: Proceedings of the AAAI-99 Fall Symposium, November 5–7, 1999. North Falmouth, MA, AAAI Press, Menlo Park, CA.

[6] H.H. Clark, D. Wilkes-Gibbs, Referring as a collaborative process, Cognition 22 (1986) 1–39.

[7] C. Doran, J. Aberdeen, L. Damianos, L. Hirschman, Comparing several aspects of human–computer and human–human dialogues, In: Proceedings of the Second SIGdial Workshop on Discourse and Dialogue, Aalborg, Denmark, September 1–2, 2001.

[8] A. Filipi, R. Wales, Perspective-taking and perspective-shifting as socially situated and collaborative actions, Journal of Pragmatics 36 (10) (2004) 1851–1884.

[9] Fischer, K. and M. Lohse., Shaping naive users' models of robots' situation awareness, in: Proceedings of the 16th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN 2007), Jeju, Korea, August 2007, IEEE, 2007.

[10] K. Fischer, R. Moratz, From communicative strategies to cognitive modeling, in: Proceedings of the First International Workshop on Epigenetic Robotics: Modeling Cognitive Development in Robotic Systems, vol. 85, Lund University Cognitive Studies, 2001.

[11] F. Fonseca, M. Egenhofer, C. Davis, G. Câmara, Semantic granularity in ontology-driven geographic information systems, Annals of Mathematics and Artificial Intelligence *36* (1–2) (2002) 121–151.

[12] A. Frisch, M. Stenberg, Navigo—an in-vehicle navigation dialogue system, Master's Thesis, University of Gothenburg, 2008.

[13] S. Garrod, A. Anderson, Saying what you mean in dialogue: a study in conceptual and semantic coordination, Cognition 27 (1987) 181–218.

[14] J. Goschler, E. Andonova, R. Ross, Perspective use and perspective shift in spatial dialogue, in: C. Freksa, N. Newcombe, P. Gärdenfors, S. Wölfl (Eds.), Spatial Cognition VI: Learning, Reasoning, and Talking about Space, Springer, Berlin, 2008, pp. 250–265.

[15] P.G.T. Healey, G.J. Mills, Participation, precedence and co-ordination in dialogue, in: R. Sun, N. Miyake (Eds.), Proceedings of the 28th Cognitive Science Conference, 2006, pp. 1470–1475.

[16] P.J. Hinds, T.L. Roberts, H. Jones, Whose job is it anyway? A study of human–robot interaction in a collaborative task, Human–Computer Interaction 19 (1/2) (2004) 151–181.

[17] W.J.M. Levelt, Cognitive styles in the use of spatial direction terms, in: R.J. Jarvella, W. Klein (Eds.), Speech, Place, and Action, Wiley, Chichester, 1982, pp. 251–268.

[18] M.T. MacMahon, Following natural language route instructions, Ph.D. Thesis, The University of Texas at Austin, 2007.

[19] M. McTear, Spoken dialogue technology: enabling the conversational user interface, ACM Computing Surveys (CSUR) 34 (1) (2002) 90–169.

[20] T. Misu, T. Kawahara, Speech-based interactive information guidance system using question-answering technique, In: Proceedings of the ICASSP IEEE International Conference on Acoustics, Speech and Signal Processing, vol. 4, 2007, pp. 145–148.

[21] R. Moratz, T. Tenbrink, Spatial reference in linguistic human–robot interaction: iterative, empirically supported development of a model of projective relations, Spatial Cognition and Computation 6 (1) (2006) 63–106.

[22] R. Murphy, Introduction to A.I. Robotics, The MIT Press, Cambridge, MA, 2000.

[23] J. Pearson, M. Pickering, H. Branigan, J. McLean, C. Nass, J. Hu, The influence of beliefs about an interlocutor on lexical and syntactic alignment: evidence from human–computer dialogues, in: Proceedings of the 10th Annual Conference on Architectures and Mechanisms of Language Processing, 2004.

[24] M. Purver, J. Ginzburg, P. Healey, On the means for clarification in dialogue, in: R. Smith, J. van Kuppevelt (Eds.), Current and New Directions in Discourse and Dialogue, Kluwer, Dordrecht, 2003, pp. 235–255.

[25] K.-F. Richter, M. Tomko, S. Winter, A dialog-driven process of generating route directions, Computers, Environment and Urban Systems 32 (3) (2008) 233–245.

[26] R.J. Ross, Tiered models of spatial language interpretation, in: C. Freksa, N. Newcombe, P. Gärdenfors, S. Wölfl (Eds.), Spatial Cognition VI: Learning, Reasoning, and Talking about Space, Springer, Berlin, 2008, pp. 233–239.

[27] R.J. Ross, Situated dialogue systems: agency & spatial meaning in task-oriented dialogue, Ph.D. Thesis, University of Bremen, 2009.

[28] M.F. Schober, How addressees affect spatial perspective choice in dialogue, in: P.L. Olivier, K.-P. Gapp (Eds.), Representation and Processing of Spatial Expressions, Lawrence Erlbaum, Mahwah, NJ, 1998, pp. 231–245.

[29] M.F. Schober, Spatial dialogue between partners with mismatched abilities, in: K. Coventry, T. Tenbrink, J. Bateman (Eds.), Spatial Language and Dialogue, Oxford University Press, Oxford, 2009, pp. 23–39.

[30] H. Shi, R.J. Ross, T. Tenbrink, J. Bateman, Modelling illocutionary structure: combining empirical studies with formal model analysis, in: Proceedings of the 11th International Conference on Intelligent Text Processing and Computational Linguistics (CICLing), Iasi, Romania, 2010.

[31] M. Steedman, The Syntactic Process, MIT Press, Cambridge, MA, 2000.

[32] H.A. Taylor, B. Tversky, Perspective in spatial descriptions, Journal of Memory and Language 35 (1996) 371–391.

[33] T. Tenbrink, K. Fischer, R. Moratz, Spatial strategies in human–robot communication, in: C. Freksa (Ed.), KI 4/02 Themenheft Spatial Cognition, 2002, pp. 19–23.

[34] T. Tenbrink, R.J. Ross, E. Andonova, J. Goschler, Spatial granularity and perspective in route descriptions for humans and dialogue systems, in: M. Tomko, K.-F. Richter (Eds.), Adaptation in Spatial Communication: Workshop held in conjunction with AGILE 2009, Hannover, Germany, SFB/TR 8 Report No. 019-05/2009, 2009, pp. 27–36.

[35] T. Tenbrink, H. Shi, Negotiating spatial goals with a wheelchair, in: S. Keizer, H. Bunt, T. Paek (Eds.), Proceedings of the Eighth SIGdial Workshop, Antwerp, Belgium, 2007, pp. 103–110.

[36] T. Tenbrink, S. Winter, Variable granularity in route directions, Spatial Cognition and Computation 9 (2009) 64–93.

[37] S. Thrun, Toward a framework for human–robot interaction, Human–Computer Interaction 19 (1/2) (2004) 9–24.

[38] S. Timpf, W. Kuhn, Granularity transformations in wayfinding, in: C. Freksa, C. Habel, W. Brauer, K.F. Wender (Eds.), Spatial Cognition III: Routes and Navigation, Human Memory and Learning, Spatial Representation and Spatial Learning, Springer, Berlin, Heidelberg, 2003, pp. 77–88.

[39] D. Traum, S. Larsson, The information state approach to dialogue management, in: R. Smith, J. van Kuppevelt (Eds.), Current and New Directions in Discourse and Dialogue, Kluwer Academic Publishers, 2003, pp. 325–353.

[40] M. Tomko, Destination descriptions in urban environments, Ph.D. Thesis, University of Melbourne, 2007.

[41] B. Tversky, Spatial perspective in descriptions, in: P. Bloom, M.A. Peterson, L. Nadel, M.F. Garrett (Eds.), Language and Space, MIT Press, Cambridge, MA, 1999, pp. 109–169.

[42] B. Tversky, B.M. Hard, Embodied and disembodied cognition: spatial perspective-taking, Cognition 110 (2009) 124–129.

[43] A. Winterboer, T. Tenbrink, R. Moratz, Spatial directionals for robot navigation, in: M. Dimitrova-Vulchanova, E. van der Zee (Eds.), Motion Encoding in Spatial Language, Oxford University Press, in press.

[44] E. Zoltan-Ford, How to get people to say and type what computers can understand, International Journal of Man–Machine Studies 34 (1991) 527–547.