Is it Really that Simple? The Complexity of Object Descriptions in Human-Computer Interaction

Vivien Mast (viv@tzi.de)
SFB/TR8 Spatial Cognition, I5-[DiaSpace]
University of Bremen

Evelyn Bergmann (ebergman@uni-bremen.de)
SFB/TR8 Spatial Cognition, I6-[NavTalk]
University of Bremen

Abstract

In the literature on verbal human-computer interaction there is general consent that humans' preconceptualisations of the machine's capabilities lead to conceptual and syntactic simplifications of the language used. We present a Wizard of Oz / Confederate study where humans communicate with either a *system* or an *expert* in a localization task in a complex building, in a setting which encourages them to give as much information as possible. We analyzed the syntactic complexity of object descriptions. Although we did find differences concerning the complexity of object descriptions on the clausal level, there were no significant structural differences on the subclausal level.

Keywords: human-computer interaction; syntactic variation; complexity; dialogue systems; Wizard of Oz; object descriptions

Introduction

Imagine you were lost in a building, but a dialogue system offered help if only it could locate you by a description of your surroundings. What kinds of information would you believe to be comprehensible to the system? And how would you shape your language in order to be understood?

According to prior research, humans communicating with artificial agents tend to use language that is both conceptually and syntactically simpler than when talking to other humans (Amalberti, Carbonell, & Falzon, 1993; Tenbrink, 2005; Moratz, Fischer, & Tenbrink, 2001). On the other hand, we know that humans adapt to the needs of their communication partners (Clark & Wilkes-Gibbs, 1986; Clark & Bangerter, 2004). In interaction with artificial agents, humans often do not know what level of knowledge and competence to expect from their interlocutor, and their expectations are influenced by different sources, such as preconceptualizations, domain, robot appearance, dialogue situation, and the course of the dialogue itself (Fischer, 2011).

When designing a system for user localization in complex buildings, it is central to determine what kinds of utterances should be expected, and whether findings from human-human interaction (HHI) research serve as a good basis for system design. In this paper, we will present a comparative study of human-computer interaction (HCI) vs. HHI in a user localization scenario designed to encourage the assumption of high cognitive and linguistic system capacities. Our analysis focuses on the number and syntactic complexity of object descriptions, as they are a central part of localization dialogues. Based on the literature we expected participants' language to be more complex when talking to the *expert* than when talking to the *system*. Finally, we will discuss the consequences of our findings for research in human-computer interaction and system design.

Human-Computer Interaction

Amalberti et al. (1993) summarize early *Wizard of Oz* research which found that in HCI participants tend to use fewer dialogue control acts, less structured dialogue, more "standard" forms, and simpler linguistic structures than in HHI. Linguistic simplifications include fewer referring expressions, less variation of syntactic structures, shorter verbal complements and a smaller vocabulary. For example, in a study comparing typed conversations, Kennedy, Wilkes, Elder, and Murray (1988) found that participants in HCI relied on a reduced lexicon, minimized usage of pronominal anaphor, and used shorter utterances, as compared to HHI.

A number of studies also report conceptual simplifications in HCI: when giving route instructions to a system in a map-based task, speakers mainly rely on turn-by-turn instructions, as opposed to the more complex goal-oriented descriptions usually used by humans (Tenbrink, Ross, Thomas, Dethlefs, & Andonova, 2010). In an experiment by Moratz et al. (2001), when instructing a robot to interact with objects, users tend to use fine-grained, path-based instructions, micromanaging the robot's movements; unlike known findings in HHI, they also consistenly use the robot's perspective.

Influences on Expectations and Behaviour

The studies mentioned here seem to give a clear picture, indicating that humans use conceptually and linguistically simpler language when speaking to an artificial agent, as compared to humans. However, linguistic behaviour depends on a number of influencing factors, and the nature of the communication partner (human vs. machine) is only one of them.

When communicating with an artificial agent, humans do not know what degree of linguistic, cognitive, and sensorimotor capacities to expect from their interlocutor, be it a robot or an information-based computer system (Moratz et al., 2001; Fischer, 2011). Therefore, they are bound to form a hypothesis based on the information available. Fischer argues that both conceptual and linguistic behaviour of humans in HCI

depend on the user's conceptualization of the agent's affordances (Fischer, 2011). She shows that this conceptualization can be partially influenced by the *physical appearance* of the artificial agent, but more strongly so by *users' preconceptions* and the dialogue flow (Fischer, 2011, 2008).

It has also been widely demonstrated that speakers adapt to their partner during the course of a dialogue (Clark & Wilkes-Gibbs, 1986; Clark & Bangerter, 2004; Garrod & Pickering, 2007). This also holds for HCI. For example, speakers show linguistic adaptation to improve understandability (Oviatt, Bernard, & Levow, 1998). Concerning the differences between HHI and HCI, while Kennedy et al. (1988) failed to manipulate the language of the user towards a more HHI-like style by more polite machine output, Amalberti et al. (1993) show that differences between HHI and HCI decrease over time, if the interlocutor's behaviour is identical in both conditions. Also, the mode of communication influences discourse behaviour. Generally, in oral communication speakers produce longer utterances than in written communication, and use a less normative style (Chafe, 1985).

In our opinion, crucial factors in influencing user's linguistic style are the domain and dialogue task. Early *Wizard of Oz* studies centered on problems that could be solved by exchanging relevant information in short question-answer pairs, like requiring information about which cells contain which geometrical shapes (Kennedy et al., 1988). Also, it was usually very clear which kind of information would be required in order to succeed in solving the task.

An extreme example of a different domain and dialogue task is *ELIZA*, an early conversational agent that took the role of a Rogerian psychotherapist and was designed to draw "his patient out by reflecting the patient's statements back to him." (Weizenbaum, 1976). Though mainly intended as a demonstration gimmick, people who conversed with *ELIZA* became "deeply [...] involved with the computer and [...] unequivocally [...] anthropomorphized it." (Weizenbaum, 1976).

In the following, we present the setup of our study which was aimed at comparing HCI and HHI in a scenario designed to encourage participants to form high expectations of their interaction partner.

Setup of the Study

In the study presented in this paper, we relied on two strategies to create a dialogue situation that would encourage participants to speak naturally to the system.

Firstly, the setting itself was chosen to be one where the precise nature of the information needed could not be easily guessed. Participants were brought to different positions in a complex building, and engaged in a remote spoken language dialogue with either the so-called "Infocenter expert" or "Infocenter system" whose supposed task it was to locate the participants in the building. No information was given to participants about the kind of information the *sytem/expert* had or could process, and it is evident that such a task does not provide a clear and easy solution. Any number of objects and their features or relations to each other could be relevant, and there are numerous ways to describe these. Therefore the task and setting itself encouraged the participants to describe as much as possible so that they could be localised.

Secondly, participants were encouraged to give detailed descriptions by employing feedback methods (see section *Dialogue Flow* below). This is closer to natural discourse behaviour than just shaping questions more politely, as Kennedy et al. (1988) did.

Procedure

We conducted the study in GW2, a complex building at the University of Bremen. The building has four floors with different layouts consisting of one or two main areas. Five positions in the building with different spatial layouts (t-intersections, open spaces, and an irregular intersection) were chosen for the experiment, making sure they were sufficiently far apart to make the dialogue situation plausible.

Before the task, participants filled in a questionnaire regarding basic demographic facts, prior knowledge of the building, and the Questionnaire on Spatial Strategies by Münzer and Hölscher (2011). They were then told that they would talk to either the "Infocenter system" (*system condition*) or the "Infocenter expert" (*expert condition*) that would try to locate them in the building and ask them questions. They were instructed to answer as well as they could. In order to enable inherently plausible dialogues about the physical environment, participants were told that the use of room numbers was not allowed.

Participants were brought to each point in ascending order. They were instructed to initiate the dialogue at each position with a predefined phrase, ***Ich bin bereit.*** *(I am ready.)*. If participants asked the experimenter about the kind of expressions or information they should use, he/she repeated that they could say whatever they wanted, except for room numbers. No further information about the task was given.

Participants

Overall, we tested 33 participants. One participant had to be excluded from evaluation due to technical problems. Of the remaining 32 participants, 17 interacted with the *system*, and 15 with the *expert*. All participants were students at the University of Bremen and reported native or near-native competence of German. There were 26 female (13 per condition) and 6 male participants (*expert condition*: 2, *system condition*: 4) aged 18 – 31 years (mean: 22). Prior knowledge of the building was intermediary: On a 7-point Likert scale, scores ranged from 2 to 5 in both conditions, with means of 3.18 in the *system condition* ($sd = 1.07$) and 3.6 in the *expert condition* ($sd = 0.91$) There was no significant difference between conditions (two-sample t-test: $t = -1.194$, $df = 30$, $p = 0.2418$).

Technical Setup

Three experimenters took turns as wizard or confederate, each experimenter playing both roles. Great care was taken to

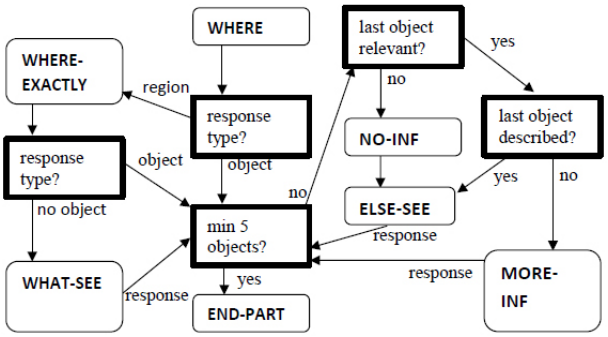

Figure 1: Overview of the dialogue flow for a given position.

provide a technical setup that ensured equivalent behaviour.

In the *system condition* we used a modified Wizard of Oz setup. The participant interacted with the *system* via spoken language, using a headset connected to a laptop. The participant's speech was sent via a one-way skype connection to the wizard laptop. The wizard classified the user utterance using an interface implemented for this purpose, whereupon the system response was automatically determined by the interface according to the dialogue flow described below. This ensured that the wizard and confederate behaved equivalently. The system response was sent as text via a socket connection to the user laptop where it was converted to speech using the MARY text-to-speech system (Trouvain & Schröder, 2003).

In the *expert condition* a computer-aided confederate setup was created. The confederate used the same interface as the wizard in order to determine the next response. Instead of text-to-speech conversion, a two-way skype connection was used. The response text was shown on the wizard interface, and was then spoken by the confederate. Confederates were instructed to conform as closely to the wording of the response utterance as possible while still maintaining natural speech rather than reading out loud.

Dialogue Flow

Care was taken to create a dialogue flow where participants felt an expectation to give as much information as possible without rendering the dialogue unnatural. Therefore, system/expert responses were designed to be as open as possible, and to give the general impression that the previous utterance(s) had been understood. In this respect, our dialogue flow was partly inspired by *ELIZA*. Although our wizard interface did not perform transformations on user utterances, it did use object names users uttered (which were typed in by the wizard/confederate) for generating replies, giving the impression that the user had been understood.

Additionally, in order to increase the naturalness of the dialogue and give the impression of high verbal capacity of the system, we produced a number of variant utterances for each system response type. The response variants were checked by two independent coders for equivalence. Responses that were semantically or pragmatically more constrained than the desired response were discarded.

Code	Example utterance
WHERE	Bitte sage mir, wo du gerade stehst. *Please tell me where you are currently standing.*
WHERE-EXACTLY	Wo genau in diesem/dieser <NAMED ELEMENT> bist du gerade? *Where exactly in this <NAMED ELEMENT> are you now?*
WHAT-SEE	Beschreibe mir bitte, was du von deiner momentanen Position aus sehen kannst. *Please describe to me what you can see from your current position.*
NO-INF	Ich habe leider keine Informationen über Einrichtungsdetails wie <NAMED OBJ> oder <RANDOM IRRELEVANT OBJ>. *Unfortunately I don't have information about details of interior furnishing such as <NAMED OBJ> or <RANDOM IRRELEVANT OBJ>.*
MORE-INF	Was kannst du mir noch über diese(s/n) <NAMED OBJ> sagen? *What else can you tell me about this <NAMED OBJ>?*
ELSE-SEE	Erzähle mir, was du dort noch sehen kannst. *Tell me what else you can see there.*
END-PART	Okay, ich habe herausgefunden wo du bist. Du kannst jetzt zum nächsten Standpunkt gehen. *Okay, I have found out where you are. You can go to the next position now!*

Table 1: Codes and example utterances for the different utterance types of the expert/system.

Before the first position, the participant was greeted by the system/expert, and after the last position, the system/expert thanked them for their participation. As a general rule, the system/expert gave a success or failure message after the user had named at least 5 types of objects. Success and failure were predetermined and did not depend on the performance of the user. Position 2 led to failure, all others to success.

If less than 5 types of objects had been named, the system asked questions to elicit further information, depending on the course of the interaction. The detailed dialogue flow can be seen in Figure 1. Examples for each type of system/expert utterance are given in Table 1; the beginning of a dialogue between the *system* and a participant is shown in Table 2.

Sp.	Code	Utterance
S	**WHERE**	Wo stehst du gerade? *Where are you standing now?*
U		Ich stehe vor einem Gang, der durch eine Glastür geht und links und rechts von mir ist jeweils auch ein Gang. *I am standing in front of a corridor that goes through a glass door and left and to my left and right is also a corridor each.*
S	**MORE-INF**	Beschreibe mir bitte diesen Gang etwas genauer. *Please describe this corridor to me in a bit more detail.*
U		Ähm, der Gang vor mir geht durch, äh, eine Glastür und ist noch sehr lang. Ähm, von dem Gang gehen viele Türen ab. Der Gang rechts von mir ist sehr kurz. Ähm, in diesem Gang befinden sich zwei Säulen und der Gang links von mir ist auch relativ lang, ähm. Geht aber ziemlich bald durch eine Glastür und auch in diesem Gang befinden sich viele weiße Säulen. *Erm, the corridor in front of me goes through, erm, a glass door and ist very long. Erm, from the corridor many doors go off. The corridor to my right is very short. Erm, in this corridor there are two pillars and the corridor to my left is also relatively long, erm. But goes also quite soon through a glass door and also in this corridor there are many white pillars.*
S	**ELSE-SEE**	Kannst du noch mehr sehen? *Can you see anything else?*
U		Hinter mir ist ein Zeichen für einen Feuermelder, verschiedene Informationsplakate. *Behind me is a sign for a fire-alarm, different information posters.*

Table 2: Beginning of a dialogue between a user (U) and the system (S).

Analysis

One major strategy for localization dialogues is a description of the current view, the spatial scene that surrounds the user. When asked where they were standing, participants answered with lists of concrete physical objects and optionally their relative position, but also with higher level descriptions of corridor constellations and region names that were inferred from signs or retrieved from knowledge. Another description strategy was to provide a route description to the current position. In our analysis, we focused on descriptions of objects. However, these could include descriptions of potential actions, as will be explained below.

The structure and complexity of object descriptions has been analyzed mainly from the point of view of how humans establish joint reference (Clark & Wilkes-Gibbs, 1986; Clark & Bangerter, 2004) and computational generation of referring expressions (Bohnet & Dale, 2005). A referring expression has the structure of a more or less complex noun phrase, modified with adjectives of colour or form, or prepositional phrases that indicate parts or spatial location. However, in the scenario presented here, people give descriptions of objects to an interlocutor without knowing the amount and nature of information he/she has available. The goal is to provide information about the scene, therefore the syntactic structures are more complex than those of referring expressions.

On the other hand, people usually rely on two main strategies when describing complex multi-object scenes like a room, a city, or a desktop array. Either the discourse is organized sequentially as a mental tour or the objects are described in lines as a parralel structure (Linde & Labov, 1975; Ullmer-Ehrich, 1982). Object configurations are referred to sequentially or in clusters, and usually the object's location is specified and not its orientation (Tenbrink, Coventry, & Andonova, 2011). Although it could be expected in the current task that participants relied on the well-documented strategies for room descriptions, interestingly, they did not. For the purpose of localisation, it seems, the participants did not aim for completeness, hence no systematical discourse organisation.

In our analysis, we used *object descriptions* as the main unit of analysis, taking into account structures more complex than referring expressions, but below the level of full scene descriptions. Based on transcriptions of the original audio recordings, coders identified all object descriptions. The beginning of an object description was identified as follows: Any object that was introduced by a noun in a main clause in rheme position: *Es gibt/ ich sehe/ ich stehe vor einer* ***Treppe****. (There is/ I'm looking at/ standing in front of a* ***staircase****)* or in an elliptic main clause: *[5] Und dann noch so ne* ***Treppe****.*[1] *(and also a* ***staircase****)*, was regarded to be the beginning of an object description, unless the clause in question was the continuation of a prior object description.

Once a new description had been identified, all parts of the utterance that preserved anaphoric reference to the described object were considered continuations of the given object description. Clauses containing repetitions or reformulations of the object name were considered continuations of the object descriptions only if they did not introduce a new object in rheme position.

Categories of Elaborations

To address the complexity of object descriptions as explained above, we analyzed the number of attributes and elaborative features (henceforth elaborations) directly relating to the target object on the subclausal level, and the number and type of clauses of the object description. Elaborations and clauses were classified into 8 categories post-hoc on the basis of the data as described below. Annotations were carried out by 3 independent coders. Intercoder reliability was checked for by independent double coding for a subset of 10 % of the data. Levels of agreement were either good or very good: Krippendorff's alpha computed on each of the 8 categories ranged between 0.785 (Adverbial Attributes) and 0.945 (Pronominal Clauses).

Compound Name: A compound noun was counted, if the initial noun describing the object was modified by a morpheme, but not if it was a simplex noun: *[2510]* ***Künstler****-Büro (****artist's*** *office)*

Adjective Attribute: indicates the number of dependent adjective attributes of the object: *[5] so 'ne* ***blaue*** *Treppe (such a* ***blue*** *staircase)*

Prepositional Attribute: number of dependent prepositional phrase attributes: *[2320] eine Treppe* ***mit Glaswänden*** *(stairs* ***with glasswalls****)*

Genitive Attribute: number of dependent genitive attributes: *[2320] im Erdgeschoss, äh,* ***des GW2*** *(on the first floor* ***of the GW2)***

Adverbial Attribute: number of adverbial attributes: *[17]* ***draußen im Flur*** *(****outside, on the corridor****)*

Pronominal Clause: number of dependent pronominal clauses: *[33] Ähm, links von mir ist wieder so 'n Eingang,* ***wo die Haupttreppe zum Kunstbereich kommt****. (On my left is an entrance,* ***where you can enter the art department****)*

Conjunction Clause: number of elaborating subordinate clauses introduced by conjunctions: *[1152] Es ist ein Holzbrett davor,* ***um die Tür aufzumachen.*** *(There is a wooden piece in front of it* ***to open the door****)*

Main Clause: indicates the number of main clauses which elaborate on the aforementioned object. This includes the primary introductory clause, and further clauses connected via 1) anaphoric pronouns „der, die, das, es, da": *[9]Ähm, ich seh hier Zeitungen, Flyers–* ***Die*** *sind rechts von mir (I see magazines, Flyers–* ***They*** *are on my right)*; 2) they explicitly refer back to an object from the discourse history: *Ähm, ich seh hier Zeitungen, Flyers.* ***Die Flyers*** *liegen rechts von mir*

[1] Numbers in angled brackets indicate the utterance number in the original corpus.

(I see magazines, flyers. ***The Flyers*** *are on my right)* 3) or they elaborate on parts of the aforementioned object which appear in theme position. Elaborations via main clauses can also be connected via „also, und, oder“ (thus, and, or). In this case, they are only counted as elaborations if they give further information about the described object, and do not introduce new objects. *[1942] Und, ähm, man kann sich das vorstellen wie ein, wie ein Dreieck.* ***Also ich stehe gerade, ich kann quasi geradeaus gehen, nach links oder nach rechts.*** *(This is like a triangle.* ***So, I can walk straight, to the left, or to the right****)*

Results

In this section, we present our findings with respect to the number of object descriptions given and the usage of the different types of elaborations and clauses. The mean number of object descriptions produced by participants at each position, and the overall mean number of descriptions per position per speaker are shown in Table 3.

Position	Expert	System
1	8.87	7.59
2	11.07	6.65
3	11.13	6.82
4	10.33	7.18
5	10.13	7.59
overall means	10.31	7.16

Table 3: Mean number of object descriptions produced by speakers of both conditions at each position.

For the number of object descriptions given per position, we fitted a linear mixed model to the data, including fixed effects for Condition, Position and their interaction and a random intercept and random slope (with respect to Position) for each participant (Pinheiro & Bates, 2000). There was a significant main effect of condition (HHI vs. HCI) ($F = 5.37$ and $p = 0.028$), but not for position ($F = 0.29$ and $p > 0.05$). No interaction effect was found ($F = 1.16$ and $p > 0.05$). In summary, this indicates that participants gave more object descriptions when talking to an *expert* than when talking to a *system*, and that there was no significant change in the number of object descriptions given during the course of the interaction.

Complexity of Object Descriptions

Table 4 shows the frequency of the different types of elaborations and clauses in the corpus in each condition, and the mean frequency of each feature per object description. As can be seen from this table, all feature types are present in both system and expert condition, indicating that humans in both HHI and HCI use the full range of syntactic possibilities for describing objects. As Table 5 shows, while the relative frequency of subclausal elaborations per object description shows only a small difference between conditions, the relative number of clauses per description shows a larger difference.

For the number of clauses per object description, we fitted a linear mixed model with fixed effects for Condition, Position and their interaction and a random effect of Participant on the intercepts. We found a statistically significant effect of condition ($F = 10.55$ and $p = 0.003$), but no effect for position ($F = 2.17$ and $p > 0.05$), and no interaction ($F = 0.43$ and $p > 0.05$), indicating that participants speaking to the *expert* used significantly more clauses per object description than those speaking to the *system*, regardless of the position. For the number of subclausal elaborations per object description, we fitted a linear mixed model with fixed effects for Condition, Position and their interaction and a random effect of Participant on the intercepts. No effect was found for condition ($F = 0.49$ and $p > 0.05$), but a significant effect for position ($F = 6.41$ and $p < 0.0001$), and no interaction ($F = 1.31$ and $p > 0.05$), showing that on the subclausal level there was no systematic difference between HHI and HCI in our study. Using contrasts to break down the effect of position, a significant linear trend was found ($t = 3.32$, $p = 0.001$), indicating a linear increase in subclausal elaborations in the course of the interactions.

Elaborative Feature	Expert			System		
	(in 773 OD)	mean per OD	% of E	(in 609 OD)	mean per OD	% of E
Compound Names	276	0.36	10.57	219	0.36	12.01
Adjective Attributes	305	0.40	11.69	199	0.33	10.91
PP-Attributes	324	0.42	12.41	277	0.46	15.19
Genitive Attributes	7	0.01	0.27	3	0.00	0.16
Adverbial Attributes	251	0.32	9.62	173	0.28	9.48
Main Clauses	1147	1.48	43.95	816	1.34	44.74
Promoninal Clauses	209	0.27	8.01	87	0.14	4.77
Conjunction Clauses	91	0.12	3.49	50	0.08	2.74
Total	2609		100.00	1824		100.00

Table 4: Frequency of the different elaboration or clause types (E) when speaking to the *expert* vs. the *system*. The table shows total frequency in the corpus, and mean frequency per object description (OD).

Feature Type	Expert			System		
	total	mean	% E	total	mean	% E
Clausal	1446	1.87	55.42	953	1.56	52.25
Subclausal	1163	1.50	44.58	871	1.43	47.75
Total	2609		100.00	1824		100.00

Table 5: Number of clauses and non-clausal elaborating features (E) in each condition. The table shows total frequency in the corpus, and mean frequency per object description (OD).

Discussion

In this paper, we have examined the differences between human-human und human-computer interaction in a user localization scenario designed to encourage participants to develop high expectations of the linguistic and cognitive capacities of an artificial communication partner. We have analyzed the number of object descriptions and their complexity as represented by 8 types of elaboration in HHI and HCI. Although the number of object descriptions given overall, and the number of clauses within these descriptions was higher for HHI than for HCI, participants in the HCI scenario showed the

full range of syntactic variability. They used all types of clauses and subclausal elaborations in sufficiently high frequency that they cannot be discarded as exceptions. Particularly, for subclausal elaborations, no significant difference in frequency between HHI and HCI could be found. These findings support our assumption that, given the appropriate scenario, HCI can be fairly natural and more similar to HHI than may be expected. In our opinion, this contradicts strong claims of *computer talk* as a separate register which is per se distinct from HHI (Zoeppritz, 1985). Rather, HCI shows parallels with *intercultural communication* where a number of individual and situational factors come together to shape (linguistic) behaviour, mediated by the interactant's conceptualizations (Fischer, 2011, 2007).

With regard to system design, the broad variability of the user's utterances shows that there is no way around developing systems with high verbal skills, which includes grammatical as well as conceptual competence. Future research could focus on further determining the influences and boundaries for shaping humans' linguistic behaviour towards artificial agents. Focusing on system design, this could help answer the question of how to frame human-computer interactions in a way that the users' expectations and the competence of the system are well matched.

Acknowledgments

This research was supported by the SFB/TR 8 Spatial Cognition (Deutsche Forschungegemeinschaft, DFG). We would also like to thank the I5-[DiaSpace] and I6-[NavTalk] project groups, and especially Thora Tenbrink for support and insightful discussions, and Daniel Couto Vale and Mohammed Fazleh Elahi for their support with the design, technical setup, and data collection. We also thank our student assistants, Nadine Hagemann, Denise Rathjen, Gesa Schole, Christina Freihorst, Jonathan Burke and Verena Seidel for their reliability and speed with data collection and coding.

References

Amalberti, R., Carbonell, N., & Falzon, P. (1993). User representations of computer systems in human-computer speech interaction. *International Journal of Man-Machine Studies*, *38*(4), 547 - 566.

Bohnet, B., & Dale, R. (2005). Viewing referring expression generation as search. In *Proceedings of the 19th international joint conference on artificial intelligence* (pp. 1004–1009). San Francisco, CA: Morgan Kaufmann.

Chafe, W. L. (1985). Linguistic differences produced by differences betweeen spaking and writing. In D. Olson, N. Torrance, & A. Hildyard (Eds.), *Language, literacy, and learning* (p. 105-123). Cambridge University Press.

Clark, H. H., & Bangerter, A. (2004). Changing ideas about reference. In D. Sperber & I. A. Noveck (Eds.), *Experimental pragmatics.* Hampshire, NY: Palgrave Macmillan.

Clark, H. H., & Wilkes-Gibbs, D. (1986). Referring as a collaborative process. *Cognition*, *22*(1), 1 - 39.

Fischer, K. (2007). Mensch-Computer-Interaktion als interkulturelle Kommunikation. In C. Sandten, K. Starck, & M. Schrader-Kniffki (Eds.), *Transkulturelle Begegnungen* (pp. 35–50). Trier: Inputs.

Fischer, K. (2008). The role of user's concepts of the robot in human-robot spatial instruction. In T. Barkowsky, M. Knauff, G. Ligozat, & D. R. Montello (Eds.), *Spatial Cognition V: Reasoning, Action, Interaction. International Conference Spatial Cognition 2006, Bremen, Germany, September 24-28, 2006, Revised Selected Papers. LNCS vol. 4387.* (pp. 76–89). Springer.

Fischer, K. (2011). How people talk with robots: Designing dialogue to reduce user uncertainty. *AI Magazine*, *32*(4), 31–38.

Garrod, S., & Pickering, M. J. (2007). Alignment in dialogue. In G. Gaskell (Ed.), *Oxford handbook of psycholinguistics.* Oxford University Press.

Kennedy, A., Wilkes, A., Elder, L., & Murray, W. S. (1988, oct). Dialogues with machines. *Cognition*, *30*(1), 37–72.

Linde, C., & Labov, W. (1975). Spatial networks as a site for the study of language and thought. *Language*, *50*(4), 924–939.

Moratz, R., Fischer, K., & Tenbrink, T. (2001). Cognitive modelling of spatial reference for human-robot interaction. *International Journal On Artificial Intelligence Tools*, *10*(4).

Münzer, S., & Hölscher, C. (2011). Entwicklung und Validierung eines Fragebogens zu räumlichen Strategien (Development and validation of a self-report measure of environmental spatial strategies). *Diagnostica*, *57*(3), 111–125.

Oviatt, S., Bernard, J., & Levow, G.-A. (1998). Linguistic adaptations during spoken and multimodal error resolution. *Language and Speech*, *41*(3-4), 419–442.

Pinheiro, J. C., & Bates, D. M. (2000). *Mixed-effects models in s and s-plus*. New York, NY [u.a.]: Springer.

Tenbrink, T. (2005). Identifying objects on the basis of spatial contrast: An empirical study. In C. Freksa, M. Knauff, B. Krieg-Brückner, B. Nebel, & T. Barkowsky (Eds.), *Spatial cognition iv. reasoning, action, interaction* (Vol. 3343, p. 124-146). Springer Berlin / Heidelberg.

Tenbrink, T., Coventry, K. R., & Andonova, E. (2011). Spatial strategies in the description of complex configurations. *Discourse Processes*, *48*(4), 237–266.

Tenbrink, T., Ross, R. J., Thomas, K. E., Dethlefs, N., & Andonova, E. (2010). Route instructions in map-based human–human and human–computer dialogue: A comparative analysis. *Journal of Visual Languages and Computing*, *21*(5), 292 - 309.

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

Ullmer-Ehrich, V. (1982). The structure of living space descriptions. In R. J. Jarvella & W. Klein (Eds.), *Speech, place, and action* (p. 219-249). Chichester: Wiley.

Weizenbaum, J. (1976). *Computer power and human reason: From judgment to calculation*. San Francisco: Freeman.

Zoeppritz, M. (1985). *Computer talk?* (Tech. Rep. No. 85.05). IBM Scientific Center Heidelberg.