The Role of Users' Preconceptions in Talking to Computers and Robots

Kerstin Fischer
University of Bremen, Germany
kerstinf@uni-bremen.de

Abstract

Communication with artificial interaction partners differs in many ways from communication among humans, and often so in the very first utterance. That is, in human-computer and human-robot interaction users address their artificial communication partner on the basis of preconceptions. The current paper addresses the nature of speakers' preconceptions about robots and computers and the role these preconceptions play in human-computer and human-robot interactions. That is, I will show that a) two types of preconceptions as opposing poles of the same dimension of interpersonal relationship can be distinguished, b) these types can be readily identified on the basis of surface cues in the users' utterances, b) these preconceptions correlate with the users' linguistic choices on all linguistic levels, and d) these preconceptions also influence the speakers' interactional behaviour, in particular, with respect to which their linguistic behaviour can be influenced, that is, in how far speakers align with the computer's and robot's linguistic output.

1 Introduction

When we look at the literature available on how people talk to computers and robots, it soon becomes clear that people talk to artificial communication partners differently from how they talk to other humans. This has led to the proposal that speech directed at artificial communication partner constitutes a register, so-called *computer talk* [27, 14]. When we keep looking, however, it turns out that in fact we know very little both about the exact nature of users' preconceptions about artificial communication partners and the effect these preconceptions have on human-computer, or human-robot, interaction situations.

In this paper I will propose that there are two prototypes of users' preconceptions, which can be reliably identified on the basis of linguistic surface cues and which have systematic effects on the linguistic properties of users' utterances. Thus, I show that the speakers' recipient design, i.e. their choosing of linguistic properties on the basis of their concept of their communication partner, is pervasive and plays a central role in the formulation both of every single utterance and with respect to all linguistic levels.

Previous research has shown that recipient design [22, 23] and audience design [1] play a major role in the communication among humans. Recently, there is an ongoing debate about how much knowledge about the communication partner exactly speakers take into account [9, 10] and under what circumstances; however, it is clear that speakers take their communication partners into account to some degree [24]. How such models are being built up, what exactly speakers take into account when building up such models, and how these models influence the speech produced for the respective partner is so far an unresolved issue (see also the contributions by Branigan and Pearson, this volume; Wrede et al., this volume; Andonova, this volume). Thus, particularly in human-computer and human-robot interaction, we yet don't know much about the preconceptions on the basis of which users tailor their speech for their artificial communication partners and in which ways.

Moreover, users in human-computer interaction are usually treated as a homogeneous group (see, for example, the studies in [14] or Gieselmann and Stenneken, this volume; Kopp, this volume; Batliner et al., this volume; Porzel, this volume). If at all, external sociolinguistic variables, such as age or gender, domain knowledge or familiarity with computers are being considered: "Explicit data capture involves the analysis of data input by the user, supplying data about their preferences by completing a user profile. Examples of explicit data captured are: age, sex, location, purchase history, content and layout preferences." [2], where implicit data elicitation is taken to involve the examination of server logs and the implementation of cookies for the identification of users' "different goals, interests, levels of expertise, abilities and preferences" [12]. User modeling should however not be restricted to factors related to the task or domain, since, as I am going to show, the users' preconceptions about such interfaces themselves cause considerable differences in users' linguistic behaviour.

Another open issue is the influence of the speakers' preconceptions on the interactional dynamics; the question is whether, besides influencing the users' linguistic choices, their recipient design also determines the discourse flow. I am going to demonstrate that such concepts have considerable influence on the users' alignment behaviour [20, 19], see also Branigan and Pearson, this volume.

2 Methods and Data

The procedure taken here is to analyse first speakers' preconceptions of their artificial communication partners as they become apparent in several corpora of human-computer and human-robot interaction. There are various possibilities to study speakers' concepts about their communication partner; one is to elicit speakers' ideas about their communication partner by means of questionnaires; this method is used, for instance, by Andonova, this volume, and by Wrede et al., this volume. In contrast, the methodology used here is essentially ethnomethodological; that is, I focus on speakers' common sense reasoning underlying their linguistic behaviour by orienting to their own displays of their understanding of the affordances of the situation. For instance, speakers will produce displays of their concepts about the communication partner in their clarification questions, but also in their reformulations. For example, the question directed at the experimenter *does it see anything?* shows that the user suspects the robot to be restricted in its perceptual capabilities and, moreover, that the speaker regards the robot as an *it*, a machine, rather than another social interactant. The reformulation in example (1) shows that the speaker suspects the robot to understand an extrinsic spatial description if it doesn't understand a projective term:

(1) S: go left
R: error
S: go East

From such displays, especially if they turn out to be systematic and recurrent both between speakers as well as within the same speaker's speech through time, we can infer what preconceptions the speakers hold about their artificial communication partner and what strengths and weaknesses they ascribe to it.

In addition, I use quantitative analyses to identify differences in distributions of particular linguistic properties as effects of the speakers' differing preconceptions about computers and robots.

The corpora I use were elicited in Wizard-of-Oz scenarios in order to ensure that all users are confronted with the same computer or robot behaviour. That is, the linguistic and other behaviour of the artificial system is produced by a human wizard but on the basis of a fixed schema of behaviours. In this way I can control for inter- and even intrapersonal variation [6]. Speakers (just) get the impression that the system is not functioning well. Besides comparability, another advantage is therefore that the repeated use of system malfunction encourages the users to reformulate their utterances frequently and thus to reveal their hypotheses about their artificial communication partner.

Human-Computer Appointment Scheduling Corpus This corpus consists of 64 German and 8 English human-computer appointment scheduling dialogues (18-33 min each). The corpus was recorded in a Wizard-of-Oz scenario in the framework of the Verbmobil project [26]. Speakers are confronted with a fixed pattern of (simulated) system output which consists of sequences of acts, such as messages of failed understanding and rejections of proposals, which are repeated in a fixed order. The fixed schema of sequences of prefabricated system utterances allows us to identify how each speaker's reactions to particular types of system malfunctions change over time. It also allows the comparison of the speakers' use of language interpersonally. The impression the users have during the interaction is that of communicating with a malfunctioning automatic speech processing system, and the participants were indeed all convinced that they were talking to such a system. The data were transcribed and each turn was labelled with a turn ID that shows not only the speaker number, but also the respective position of the turn in the dialogue. Subsequently, the data were annotated for prosodic, lexical, and conversational properties. <P>, <B>, <L> stand for pause, breathing, and sylable lengthening respectively.

Human-Robot Distance Measurement Corpus The second corpus used here was elicited in a scenario in which the users' task was to instruct a robot to measure the distance between two objects out of a set of seven. These objects differed only in their spatial position. The users typed instructions into a notebook, the objects to be referred to and the robot being placed on the floor in front of them. The relevant objects were pointed at by the instructor of the experiments. There were 21 participants from all kinds of professions and with different experience with artificial systems. The robot's output was generated by a simple script that displayed answers in a fixed order after a particular 'processing' time. Thus, the dialogues are also comparable regarding the robot's linguistic material, and the users' instructions had no impact on the robot's linguistic behaviour. The robot, a Pioneer 2, could not move either, but the participants were told that they were connected to the robot's dialogue processing system by means of a wireless LAN connection. Participants did not doubt that they were talking to an automatic dialogue processing system, as is apparent from their answers to the question: "If the robot didn't understand, what do you think could have been the cause?". The robot's output was either "error" (or a natural language variant of it) or a distance in centimeters. Since by reformulating their utterances the users display their hypotheses about the functioning of the system (see above), error messages were given frequently.

The user utterances are typed and thus transcription was not necessary; typos were not corrected. The turn IDs show the speaker number, for instance, usr-20, and the number of the turn in the dialogue.

Human-Robot Spatial Instruction Corpus This corpus was elicited with three different robots, Sony's Aibo, Pioneer, another commercially available robot, and Scorpion, built by colleagues at the University of Bremen [25]. Since we used a Wizard-of-Oz scenario, we were able to confront all users again with identical non-verbal robot behaviours, independent of the users' utterances. We elicited 30 English dialogues, using the same speakers, scheduling the recordings at least three months apart, and 66 German dialogues, in which we recruited naive users for each scenario. Here, we elicited 12 dialogues with Aibo, 33 with pioneer and 21 with scorpion.

The users' task was to instruct the respective robot to move to objects which were placed on the floor in front of them and which were pointed at by the experimenter. All robots moved between the objects in the same, predefined, way (there was no linguistic output).

The dialogues were transcribed and analysed with respect to their linguistic properties. Each turn ID shows whether the robot addressed was Aibo (A), Scorpion (S), or Pioneer (P). Transcription conventions are the following: (at=prominent) word (/a) means that the word is uttered in a prosodically prominent way, + indicates a word fragment, - means a short pause, – a longer pause, and (1) indicates a pause of one second; punctuation indicates the intonation contour with which the utterance was delivered.

Human-Aibo Interaction with and without Verbal Feedback For the comparison with the human-Aibo dialogues from the previous corpus, we elicited another corpus in the same scenario as before, just that Aibo also replied with verbal behaviours. The robot utterances were pre-synthesized and were played in a fixed order. The utterances were so designed as to give no clue as to what may have gone wrong in order to avoid prompting particular error resolution strategies from the users. However, in these utterances, three design features were used which previous studies [15, 3, 6] had revealed to be quite rare in human-robot interaction if the robot does not give feedback: First, we made the robot ask for and propose spatial references using object naming strategies. Second, we made the robot use an extrinsic reference system. Third, as an indicator of high linguistic capabilities, the robot made extensive use of relative clauses.

The robot's utterances are, for instance, the following: Ja, guten Tag, wie geht es Ihnen? (*yeah hello, how do you do?*) Soll ich das blaue Objekt ansteuern? (*do you want me to aim at the blue object?*) Soll ich mich zu dem Objekt begeben, das vorne liegt? (*do you want me to move to the object which lies in front?*) Meinen Sie das Objekt, das 30 Grad westlich der Dose liegt? (*do you mean the object that is 30 degrees west of the box?*) Ich habe Sie nicht verstanden. (*I did not understand.*) Entschuldigung, welches der Objekte wurde von Ihnen benannt? (*excuse me, which object was named by you?*) Ich kann nicht schneller. (*I can't go faster.*)

The corpus comprises 17 German human-Aibo dialogues recorded under circumstances exactly as in the corpus described above, just that the fixed schema of robot behaviours was paired with a fixed schema of robot utterances, both independent of what the speaker is saying.

3 Concepts about Computers and Robots

There are some beliefs about computers and robots that surface frequently and in all of the corpora under consideration. The first one is the concept of the computer or robot as linguistically restricted. This view of the artificial communication partner is in fact only encouraged in the human-computer interaction corpus when the system produces *I did not understand*. In the corpora in which the robot does not produce any speech, no such clues are given. Similarly, in the distance measurement corpus, only *error*-messages are produced, and thus the idea that the robot could be linguistically challenged is likely to stem from the speakers' preconceptions. Even more crucially, also in another corpus in which the linguistic capabilities of the robot were actually very good and in which communicative failure resulted from mismatches in instruction strategies [15], not in restricted linguistic capabilities, speakers overwhelmingly suspected the problem to have been that they weren't able to find those words that the robot would have been able to understand.

This preconception of artificial communication partners as linguistically restricted can turn out to be very problematic in the future; if our systems are getting better and the interfaces more natural, yet users continue to expect great linguistic problems, the interactions with such systems may turn out very strange, as can be seen in the following example:

(2) R: yes, hello, how do you do?
A031: (4) oh okay. - um - um go forward, to, -

Here, the user does not react at all to the polite interaction proposed by the system. The rejection of such speech acts has to be attributed to the user's preconceptions, since at that point there is no evidence of miscommunication or communicative failure. This corresponds to findings by Krause [13] as well as to observation regarding politeness by [16, 21] and [11].

Another aspect is the suspected formality of artificial communication partners. In the following example, the speaker reformulates her utterance by using exact measurements:

(3) A003: nun zu den, zwei, Dosen, – links. (5) (now to the, two, boxes, – left)
R: Ich habe Sie nicht verstanden. (I did not understand.)
A003: (1) links zu den zwei Dosen circa 30 (at=lengthening) Grad(/a) Drehung (22) (left to the two boxes about 30 degrees turn)

In the appointment scheduling dialogues, often the year is added:

(4) e4012101: what about Monday, the fourth of January? <P> from eight <P> till fourteen-hundred.
s4012102: blurb appointment right blurb mist. [*nonsense*]
e4012102: okay. what about Tuesday, the fifth of January? <P> from<L> <P> eight to fourteen-hundred?
s4012103: please make a proposal.
e4012103: <Smack> <P> okay. <;low voiced> do you have time on Monday, the eleventh of January nineteen-ninety-nine?
s4012201: this date is already occupied.
e4012201: what about Tuesday, the twelfth of January nineteen-ninety-nine?

These preconceptions seem to be very common in HCI and HRI. In [6], I furthermore show that speakers generally believe that robots can be easily disturbed by orthographical matters, that they have problems with basic level and colloquial terminology and metaphorical concepts, and that they have to learn skills in the same order as humans do. Besides these generally shared ideas, users also seem to have very different concepts of their artificial communication partner and the situation, e.g. in the human-robot dialogues:

(5) P075: I was g+ I was wondering, whether it whether it understood English. - (laughter)

(6) S037: scorpion, - turn - ninety - left. (2) turn left (at=prominent)ninety(/a). - - now is that one command or two, - -

(7) A001: good (at=laughter)dog(/a), (1) now pee on 'em (laughter) – sit, (laughter) –

(8) A004: go on, - you are doing fine,

Such utterances indicate two fundamentally different attitudes towards robots, one in which the robot is treated as a mechanical device that needs commands and which is not expected to understand natural language, and the other in which the robot is expected to function like an animal or needs positive encouragement. Similar differences can be found in the distance-measurement corpus:

(9) usr1-2: wie weit entfernt ist die rechte Tasse? (*how far away is the right cup?*)
sys:ERROR
usr1-3: Tasse (*cup*)
sys:ERROR 652-a: input is invalid.
usr1-3: die rechte (*the right one*)

(10) usr3-3: wie heißt du eigentlich (*what's your name, by the way*)

(11) usr4-25: Bist du für eine weitere Aufgabe bereit? (*are you ready for another task?*)

Examples from the appointment scheduling corpus are the following:

(12) e0045206: können Sie denn Ihre Mittagspause auch erst um vierzehn Uhr machen? (*could you take your lunch break as late as 2pm?*)

(13) e0387103: Sprachsysteme sind dumm. (*language systems are stupid*)

An important observation is that these different attitudes towards the computer or robot correspond to different ways of opening the dialogue with the artificial communication partner. These different dialogue openings reveal different preconceptions about what the human-computer or human-robot situation consists in. For example, one such first move is to ignore the contact function of the system's first utterance completely and to start with the task-oriented dialogue immediately:

(14) S: ja, guten Tag, wie geht es Ihnen? (*yes, hello, how do you do?*)
e0440001: ich möchte gerne einen Termin einen Arzttermin mit Ihnen absprechen. (*I want to schedule an appointment a doctor's appointment with you.*)

This group of speakers only minimally reacts to the interpersonal information provided by the system or even refuse communication at that level. Instead they treat the computer as a tool, at best, in any case not as a social actor. I refer to this group as the *non-players*.

In contrast, the players will take up the system's cues and pretend to have a normal conversation. I call these speakers *players* because the delivery of the respective utterances show very well that the speakers find them unusual themselves, as in the following example where the user breathes and pauses before asking back:

(15) S: ja, guten Tag, wie geht es Ihnen? (*hello, how do you do?*)
e0110001: guten Tag. danke, gut. <B> <P> und wie geht's Ihnen? (*hello, thanks, fine. <B> <P> and how do you do?*)

Thus, it is not the case that these users would mindlessly [18, 17] transfer social behaviours to the human-computer situation. For them, it is a game, and eventually it is the game system designers are aiming at. Thus, these users talk to computers *as if* they were human beings.

Also in the human-robot dialogues with written input in which the user has the first turn, the same distinction can be found:

(16) usr17-1: hallo roboter (*hello robot*)
sys:ERROR
usr17-2: hallo roboter (*hello robot*)
sys:ERROR
usr17-3: Die Aufgabe ist, den Abstand zu zwei Tassen zu messen. (*The task is to measure the distance between two cups.*)

In this example, the speaker proposes a greeting himself and even repeats it. Then, he provides the system with an overview of the task. In contrast, user 19 in the following example first types in the help command, which is current practice with unix tools; when he does not get a response, he starts with a low-level, task-oriented utterance without further elaboration or relation-establishing efforts:

(17) usr19-1: hilfe (*help*)
sys:ERROR
usr19-2: messe abstand zwischen zweitem becher von links und zweitem becher von rechts (*measure distance between second mug from left and second mug from right*)

The same two prototypes can be found in our human-robot dialogues in which Aibo uses the same initial utterance as in the appointment scheduling corpus:

(18) R: Ja guten Tag, wie geht es Ihnen? (*yes hello, how do you do?*)
A011: (1) äh, geradeaus gehen. (breathing) – (*uh, going straight*)
R: Welches Objekt soll ich ansteuern? (*which object should I aim at?*)
A011: (1) links. (7) (*left*)

In this example, the speaker immediately produces a very basic spatial instruction. The next utterance is not syntactically or semantically aligned with the robot's question. In contrast, in the next example, the speaker asks the robot back politely. Her next utterance takes up both the term and the syntactic construction of the robot's utterance, and thus her utterance can be understood as the second part of an adjacency pair:

(19) R: Ja guten Tag, wie geht es Ihnen? (*yes hello, how do you do?*)
A014: Mir geht es sehr gut und selbst? (laughter) (1) (*I'm fine and how about you?*)
R: Welches Objekt soll ich ansteuern? (*which object should I aim at?*)
A014: (2) das Objekt äh hinten links. (6) (*the object uh at the back left.*)

Further examples of dialogue beginnings illustrate the spectrum of possible dialogue openings. Thus, the two behaviours identified, the task-oriented response (by the non-players) and the polite complementary question about the system's well-being (by the players) constitute prototypes, which are located at the opposite poles of the same dimension of social relationship:

(20) R: Ja guten Tag, wie geht es Ihnen? (*yes hello, how do you do?*)
A009: (laughter) - guten Tag, - ähm, vorwärts, (2) losgehen? (1) (*hello, um, straight, start?*)

(21) R: Ja guten Tag, wie geht es Ihnen? (*yes hello, how do you do?*)
A022: (1)(at=quiet)gut?(/a) (1) (laughter) (1) (*fine?*)
R: Welches Objekt soll ich ansteuern? (*which object should I aim at?*)
A022: (1) äh vorne links? (4) stopp, - links, (*uh front left? stop, left,*)
R: Soll ich mich zu dem Objekt begeben, das vorne liegt? (*do you want me to move to the object which is in front?*)
A022: (2) nein, - weiter links, (2) (*no, - further left,*)

(22) R: Ja guten Tag, wie geht es Ihnen? (*yes hello, how do you do?*)
A012: (1) gut, danke, (2) (*fine, thanks*)
R: Welches Objekt soll ich ansteuern? (*which object should I aim at?*)
A012: (1) die Schale, - ganz links. (6) (*the bowl, very far left.*)

(23) R: Ja guten Tag, wie geht es Ihnen? (*yes hello, how do you do?*)
A025: (at=prominent)ja,(/a) (hnoise) ganz gut. (at=quiet) und du? - äh(/a) - so, getz, (*yes, quite fine. and how about you? - uh - so, now,*)
R: Welches Objekt soll ich ansteuern? (*which object should I aim at?*)
A025: (1) ähm dieses Müslischälchen was da ganz links steht. - da sollst du hingehen. (*um this muslibowl which is very much to your left - there you have to go to.*)

In general, then, irrespective of particular communication situations between humans and artificial communication partners, we can distinguish two different prototypes of preconceptions: the computer as a tool versus the computer as a social actor. These prototypes are easily classifiable with automatic means since they correlate with a set up surface cues [8].

4 Effects of the Users' Preconceptions

Now that we have established the prototypical preconceptions in human-computer and human-robot interaction, the question is whether and how these preconceptions influence the way users talk to their artificial communication partners.

4.1 The Predictability of Linguistic Features from Preconceptions

For the appointment scheduling dialogues, it was found that the occurrence of conversational and prosodic peculiarities is significantly related to the users' preconceptions as evident from the different dialogue openings [6]. That is, there are significant correlations between dialogue beginning and the use of linguistic strategies on the conversational as well as the prosodic level. The conversational peculiarities comprise reformulations, meta-linguistic statements, new proposals without any relevant relationship to the previous utterances, thematic breaks, rejections, repetitions, and evaluations. In contrast to, for instance, sociolinguistic variables, such as gender, the distinction between players and non-players has a consistent effect on the use of the above conversational strategies. Similarly, the occurrence of phonetic and prosodic peculiarities, in particular, hyper-articulation, syllable lengthening (e.g. Mon<L>day), pauses (between words and syllables, e.g. on <P> Thurs <P>day), stress variation, variation of loudness, and the variation of intonation contours, can be predicted by the dialogue beginnings [6].

Also in the distance-measurement corpus, the dialogue openings can be used to predict the linguistic strategies used. In this case, we have found a systematic relationship with the occurrence of clarification questions [7]. That is, whether speakers began dialogues with a greeting or some other kind of contact-establishing move, as in the following example, or whether they started the task immediately could be used to predict the occurrence of clarification questions, in particular questions concerning the recipient design, such as the robot's perception, functionality and linguistic capabilities, for instance:

(24) usr11-1: hallo# (*hello#*)
sys:ERROR
usr11-2: siehst du was (*do you see anything*)
sys:ERROR
usr11-3: was siehst du (*what do you see*)

Also for the three German human-robot corpora with Aibo, Scorpion and Pioneer, results show a very significant effect between dialogue opening and emotional expression, sentence mood, structuring cues, and reference to the robot. Emotional expression was coded by distinguishing interjections, e.g. *oh, ah*, contact signals, e.g. *hello*, and displays of relationship, e.g.*my friend*. Regarding structuring cues, we distinguish implicit, such as *now*, from explicit cues, e.g. *the first task*. For sentence mood, particularly relevant are imperative vs. infinitive vs. declarative mood vs. no verb at all. Finally, we coded whether speakers talked about the robot at all, and if so whether they referred to the robot as *he* or as *it*. For these linguistic features, significant differences could be found, depending on the first utterance in the dialogue [5].

To sum up, in the emotional HCI-corpus, the users' concept of the communication partner as tool or as a conversation partner was significantly related to the prosodic properties of their utterances and to the conversational strategies taken. In the distance-measurement corpus, the number of clarification questions asked can be significantly related to different dialogue openings as well [7]. In the three comparable HRI-dialogues, the effect of the conceptualisation of the robot as a tool or as a conversation partner (as apparent from the dialogue opening) was found to be the most important predicting factor for features as diverse as emotional expression, the sentence mood chosen, the kind of structuring cues used and the way the robot is referred to.

We can conclude that the preconception of the artificial communication partner as a tool versus as a social actor plays an important role in predicting the linguistic features empoyed by the users.

4.2 The Constraining Effect of Preconceptions for Alignment and Shaping

We may now want to ask how stable these preconceptions are and whether they may influence the course of the dialogue and particularly the users' alignment behaviour. This is not only theoretically interesting, but also of great practical interst since due to the restrictedness of current systems, it may be very useful to be able to subtly guide users into using those linguistic structures and strategies that the system can process best [28]. As predicted by [20], as well as Branigan and Pearson, this volume, speakers may align with the robot's output. In particular, we find:

Lexical Alignment

(25) R: Was kann ich für Sie tun? (*what can I do for you?*)
A004: (1) geh zur linken Tasse (...) links nach links (1) (*go to the left cup (...) left to the left*)
R: Soll ich mich zu dem Objekt begeben, das vorne liegt? (*do you want me to move to the object which is in front?*)

A004: – nein ähm zu dem Objekt was links liegt (*no um to the object which is to your left*)

In the example, user A004 initially employs a basic level term to describe the object, namely *cup*. After the robot uses the more abstract term *object*, the user aligns with that term. She furthermore aligns with the syntactic constrction the robot employs, namely the relative clause. In the following example, the speaker also aligns with the robot's construction by expanding it in the reply:

Constructional Alignment

(26) R: Welches Objekt soll ich ansteuern? (*which object should I aim at?*)
A003: (2) (at=breathing)hm, (/a) (3) (...) (at=quiet)ähm, (/a) – ja (2) das (3) zweite. – (*um, well the second one.*)

In the example below, the user employs the extrinsic reference system that the robot had introduced turns before:

Alignment of Reference System

(27) A003: (2) zu der Tasse, nord-östlich. (2) (*to the cup, north-east*)
R: Soll ich mich zu dem Objekt begeben, das vorne liegt? (*do you want me to go to the object that is in front?*)
A003: (4) nord-west. (laughter) mein Fehler. (laughter) (*north-west. my mistake.*)

Alignment of Instructional Strategies

(28) A058: gehe vorwärts. (*go straight.*)
Robot: Soll ich mich zu dem Glas begeben? (*do you want me to go to the jar?*)
A058: nein, geh zu dem Plastikbehälter in der Mitte vor Dir. (*no, go to the plastic container in the middle in front of you.*)

In the previous example, the speaker changes from his previous path-based instructional strategy to a goal-based strategy, aligning with the robot's orientation towards objects. In the follwong example, the user picks up the robot's formal form of address:

Alignment of Form of Address

(29) R: Welches Objekt meinten Sie? (*which object did you mean? (formal form of address)*)

A029: - bitte, fahren Sie erstmal geradeaus. (1) (*please, first drive straight (formal form of address)*)

Thus, speakers may take up the linguistic structures presented to them by the system. Moreover, a comparison between human-Aibo interaction with and without linguistic output shows that after the robot's initial utterance *hello, how do you do* many linguistic differences can be found, some of which can be attributed to alignment, some of which must however be due to changes in the conceptualization of the robot due to the fact that the robot produces verbal output. Thus, it seems that the concept of a language-generating robot is more sophisticated than that of an only language-understanding robot. In [4], I have shown the effect on spatial language; for example, the amount of consistent use of higher level spatial instructions increased from 15.4% to 41.2% in the comparison of the two conditions without and with verbal robot output. Similarly, the number of relative clauses rises significantly as well as the complexity of the syntactic structures used. Thus, the robot's linguistic behaviour contributes to speakers' conceptualization about it.

However, the speakers' preconcepts may also define the limits to this kind of adaptation, as can be seen in the following example:

(30) R: Ja guten Tag, wie geht es Ihnen? (*yes hello, how do you do?*)
A008: (2) geh vorwärts. – (*go straight*)
R: Was kann ich für Sie tun? (*what can I do for you?*)
A008: - gehe vorwärts. (7) (*go straight*)
R: Soll ich mich zu dem Objekt begeben, das vorne liegt? (*do you want me to go to the object that is in front?*)
A008: (1) nein. gehe vorwärts. (10) (*no. go straight.*)

The speaker does not adapt to the robot's utterances from the start. A bit later in the dialogue, the effect persists such that the speaker takes up neither the linguistic constructions nor the object-naming strategy presented by the robot. In the last utterance of the excerpt, he minimally aligns with the first part of the adjacency pair produced by the robot by providing the answer 'the box', but immediately after that he switches back to path-based instructions:

(31) R: Soll ich mich zum Glas begeben? (*do you want me to move to the jar?*)
A008: (3) gehe vorwärts. - (*go forward*)

R: Entschuldigung, welches der Objekte wurde von Ihnen benannt? (*excuse me, which of the objects did you name?*)
A008: (1) die Dose. (5) gehe links. (5) gehe links. (2) (*the box. go left. go left.*)

We can thus conclude that alignment, though a natural mechanism in HRI as much as in human-to-human communication, crucially depends on the users' concepts of their communication partner. That is, the less they regard the computer or robot as a social actor, the less they align. This is generally in line with the reasoning in Branigan and Pearson's article (this volume), who also argue that alignment is affected by speakers' prior beliefs. However, they hold users to align with computers only because they consider them to be lingistically limited in their linguistic capabilities, not because they would treat computers as social actors. In contrast, the findings presented here show that users do not constitute a homogeneous group, since speakers' beliefs about their artificial communication partners may vary considerably; those who regard computers as social actors will indeed align with them.

5 General Conclusions

To sum up, the users' concepts of their communication partner turned out to be a powerful factor in the explanation of inter- and intrapersonal variation with respect to linguistic features at all linguistic levels. In particular, two prototypical preconceptions could be identified, one of the artificial communication partner as a tool, one as another social actor. These prototypes can be reliably identified on the basis of the speakers' first utterances which display their orientation towards a social communication or a tool-using situation. These preconceptions have significant correlations with linguistic behaviour on all linguistic levels. Thus, speech directed to artificial communication partners is not constitute a homogeneous variety, and should thus not be referred to as a *register* [14], unless it is captured in terms of microregisters as suggested by Bateman (this volume). Moreover, depending on their attention to social aspects even in the human-computer or human-robot situation, speakers are inclined to align to their artificial communication partners' utterances. Thus, the users' preconceptions constrain the occurrence of, and define the limits for, alignment in human-computer and human-robot interaction.

References

[1] H. H. Clark. *Arenas of Language Use.* Chicago: University of Chicago Press, 1992.

[2] G. de la Flor. User modeling & adaptive user interfaces. Technical Report 1085, Institute for Learning and Research Technology, 2004.

[3] K. Fischer. Notes on analysing context. In P. Kühnlein, H. Rieser, and H. Zeevat, editors, *Perspectives on Dialogue in the New Millennium*, number 114 in Pragmatics & Beyond New Series, pages 193–214. Amsterdam: John Benjamins, 2003.

[4] K. Fischer. Discourse conditions for spatial perspective taking. In *Proceedings of WoSLaD Workshop on Spatial Language and Dialogue, Delmenhorst, October 2005*, 2005.

[5] K. Fischer. The role of users' concepts of the robot in human-robot spatial instruction. In *Proceedings of 'Spatial Cognition '06'*, 2006.

[6] K. Fischer. *What Computer Talk Is and Isn't: Human-Computer Conversation as Intercultural Communication.* Saarbrücken: AQ, 2006.

[7] K. Fischer and J. A. Bateman. Keeping the initiative: An empirically motivated approach to predicting user-initiated dialogue contributions in HCI. In *Proceedings of the EACL'06, April 2006, Trento, Italy*, 2006.

[8] M. Glockemann. Methoden aus dem Bereich des Information Retrieval bei der Erkennung und Behandlung von Kommunikationsstörungen in der natürlichsprachlichen Mensch-Maschine-Interaktion. Master's thesis, University of Hamburg, 2003.

[9] B. Horton and B. Keysar. When do speakers take into account common ground? *Cognition*, 59:91–117, 1996.

[10] W. Horton and R. Gerrig. Conversational common ground and memory processes in language production. *Discourse Processes*, 40:1–35, 2005.

[11] A. Johnstone, U. Berry, T. Ngyuen, and A. Asper. There was a long pause: Influencing turn-taking behaviour in human-human and human-computer spoken dialogues. *International Journal of Human-Computer Studies*, 41:383–411, 1994.

[12] A. Kobsa. User modeling and user-adapted interaction. In *Proceedings of CHI'94*, 1994.

[13] J. Krause. Fazit und Ausblick: Registermodell versus metaphorischer Gebrauch von Sprache in der Mensch-Computer- Interaktion. In J. Krause and L. Hitzenberger, editors, *Computertalk*, number 12 in Sprache und Computer, pages 157–170. Hildesheim: Olms, 1992.

[14] J. Krause and L. Hitzenberger, editors. *Computer Talk*. Hildesheim: Olms Verlag, 1992.

[15] R. Moratz, K. Fischer, and T. Tenbrink. Cognitive modelling of spatial reference for human-robot interaction. *International Journal on Artificial Intelligence Tools*, 10(4):589–611, 2001.

[16] M.-A. Morel. Computer-human communication. In M. Taylor, F. Neel, and D. Bouhuis, editors, *The Structure of Multimodal Communication*, pages 323–330. Amsterdam: North-Holland Elsevier, 1989.

[17] C. Nass and S. Brave. *Wired for Speech. How Voice Activates and Advances the Human-Computer Relationship*. Cambridge, MA., London: MIT Press, 2005.

[18] C. Nass and Y. Moon. Machines and mindlessness: Social responses to computers. *Journal of Social Issues*, 56(1):81–103, 2000.

[19] J. Pearson, J. Hu, H. Branigan, M. Pickering, and C. Nass. Adaptive language behavior in hci: how expectations and beliefs about a system affect users' word choice. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, Montréal, April 2006*, pages 1177–1180, 2006.

[20] M. J. Pickering and S. Garrod. Towards a mechanistic psychology of dialogue. *Behavioural and Brain Sciences*, 27:169–225, 2004.

[21] M. Richards and K. Underwood. Talking to machines: How are peoplpe naturally inclined to speak? In *Proceedings of the Ergonomics Society Annual Conference*, 1984.

[22] H. Sacks, E. A. Schegloff, and G. Jefferson. A simplest systematics for the organization of turn-taking for conversation. *Language*, 50(4):696–735, 1974.

[23] E. A. Schegloff. Notes on a conversational practise: Formulating place. In D. Sudnow, editor, *Studies in Social Interaction*, pages 75–119. New York: Free Press, 1972.

[24] M. F. Schober and S. E. Brennan. Processes of interactive spoken discourse: The role of the partner. In A. C. Graesser, M. A. Gernsbacher, and S. R. Goldman, editors, *Handbook of Discourse Porcesses*, pages 123–164. Hillsdale: Lawrence Erlbaum, 2003.

[25] D. Spenneberg and F. Kirchner. Scorpion: A biomimetic walking robot. *Robotik*, 1679:677–682, 2002.

[26] W. Wahlster, editor. *Verbmobil: Foundations of Speech-to-Speech Translation.* Berlin etc.: Springer, 2000.

[27] M. Zoeppritz. Computer talk? Technical Report TN 85.05, IBM Heidelberg Scientific Center, 1985.

[28] E. Zoltan-Ford. How to get people to say and type what computers can understand. *International Journal of Man-Machine Studies*, 34:527–647, 1991.