Getting Acquainted with a Developing Robot

Kerstin Fischer & Joe Saunders

Abstract

Two factors that been suggested to influence the ways in which people interact with robots, namely users' initial expectations on the one hand and their increasing acquaintance with their robotic partner due to repeated interaction over time on the other. In the current study, eight participants interacted with a humanoid robot in five different sessions. Between the sessions, the robot was trained on the linguistic material presented to it by its human tutor in the preceding session, and thus the robot exhibits increasingly more knowledge of the domain. The results uncover the interaction between users' preconceptions and feedback-driven interactional effects that shape human-robot interactions. While considerable differences between users can be observed, all users respond to the robot's feedback and increasing linguistic capabilities in comparable ways.

Introduction and Previous Work

In this paper, we investigate how people interact with a developing robot. In order to study the role of increasing acquaintance, we analyze users' linguistic strategies by means of which they teach the robot over time. This will show us in how far the robot's behavior and increasing capabilities influence the way people interact with it and thus which impact social communication over time and, in particular, acquaintance with the robot may have. Studies in cognitive psychology have shown that acquaintance plays a crucial role in the way in which people make use of common ground (see Clark [2]). Acquaintance has also been found

adfa, p. 1, 2011.

to be a factor in studies of human-computer interaction; for instance, Amalberti et al [1] compare participants' linguistic behaviors when they believe that their communication partner is either another human or a computer; they find that the considerable linguistic differences between speech directed at a computer and speech directed at another human, which can initially be observed in participants' speech, disappear gradually over several sessions. Thus, there is evidence that acquaintance plays a crucial role in interaction. However, it is so far unclear how such interactional effects are related to the preconceptions and expectations people bring into the interaction; several studies have shown that users' expectations also play a crucial role in the ways in which they interact with a communication partner (e.g. Fischer [6], Turkle [19]). This holds for interactions with communication partners with slightly different capabilities than one's own, such as foreigners (e.g. Zuengler [20]), as well as for interactions between younger and elderly people [11], but has also been shown for interactions with robots. For example, Turkle [19] argues that people's personal needs shape the ways they interact with relational artifacts, such as social robots. Fischer [6] shows that people's preconceptions about the degree of socialness of the human-robot interaction situation are an important factor in determining the way these people talk to a robot. Paepke and Takayama [13] manipulated users' expectations about the robot 'Pleo' by means of different introductory leaflets and find significantly different evaluations of the same robot after the interaction. Thus, preconceptions and users' expectations may have a considerable impact on HRI, yet it is unclear in how far these preconceptions are related to, and influenced by, what is happening in the course of the interactions between humans and robots.

The current study therefore aims to identify the effects of repeated interaction while taking people's initial expectations into account. We address this problem by investigating interactions between humans and a humanoid robot over time. In the current study, eight participants interacted with a humanoid robot in five different sessions. Between the sessions, the robot was trained on the linguistic material presented to it by its human tutor in the

preceding session, and thus the robot exhibits increasingly more knowledge of the domain.

Data Elicitation

Eight adult participants took part in the study. Participants were between 27 and 58 years old (five female and three male). The backgrounds of the participants were either administrative (6) or research related (2), the latter not connected with robotic language research. Each of the eight participants took part in five interaction sessions of approximately two minutes with the robot (in total 40 robotic interaction sessions), and all of the sessions were videotaped for later analysis. The experiment was carried out over a three month period between March and June 2009 based on the availability of the participants. Participants were paid a small stipend of £20 if they completed all sessions (which all participants did).

Figure 1: A participant teaching Kaspar about shapes

In the experiment we asked the participants to teach the humanoid robot Kaspar (Dautenhahn et al. [4]) a series of shapes pasted on boxes. The robot was pre-programmed to track and habituate for a given period on these shapes. There was no constraint on participants' language. How to talk to the robot and what teaching strategies to use, was thus entirely up to the respective participant.

Following each interaction, the speech stream of the human was converted into phoneme strings marked with word boundaries. These phoneme strings were subsequently aligned with the sensorimotor modalities experienced by the robot during the interaction session. The aligned speech and sensory modalities were then processed to highlight words of long duration and words that appeared at the end of utterances. This processed modality stream became the basis for the robot's learnt experiences for the next interaction session with the human. In other words, the robot learned to associate the stressed words in a particular participant's speech stream with its visual perception of the shape presented to it during the sessions.
In subsequent sessions (from session 2 onwards) the robot then matched its current sensorimotor input (that it was experiencing during the interaction) against that learnt in the previous session(s) with the particular tutor. This allowed the robot to react to the human by expressing (via its own speech) what it had learnt during the previous session(s). Thus, the robot produced feedback to the respective teacher by repeating words it had previously learned from associations of sounds to sensorimotor data. Full details of the experimental procedure can be found in Saunders et al. [16, 17].

Method

The method for analysis makes use of the principle of recipient design [15], which holds that people choose the linguistic features of their utterances to be suited best for their particular communication partners; for instance, people design their speech differently when speaking to children than when speaking to other adults (e.g. Snow [18]). In the current investigation, we make use of this principle by analyzing the participants' speech to the robot in order to identify who the participants think they are talking to. Thus, in the same way as we can identify speech to children by the shorter utterances, lower type-token ratio, lower complexity, more interactivity and more attention getting devices, we can study the properties of speech to a robot as a window into participants' concepts about their artificial communication partner and their ideas about what it will be good at and what it will have problems

with. Thus, participants' linguistic choices reveal their concepts of their communication partner. The procedure thus consists in analyzing those linguistic features that may be revealing regarding participants' concepts of the robot and to identify which of these features are affected by the variables investigated, here: the acquaintance with the robot. In some sense, this method is exploratory, as the main aim of the statistical analysis is to identify the nature of the adjustments participants make, rather than testing specific hypotheses. On the other hand, the linguistic features analyzed have certain functions, and thus certain predictions can be made with respect to the areas in which changes take place.

Data Encoding

The data were orthographically transcribed and analyzed semi-automatically using shell scripts, whose results were manually controlled for correctness. The features investigated concern different linguistic features that may safely be assumed to be indicators of certain communicative functions and of people's conceptualizations and understandings of the robot and of the human-robot interaction situation. In particular, unambiguous linguistic features were automatically extracted from the transcripts if these are revealing with respect to participants' preconceptions and expectations about the robot, the task and the human-robot interaction. Since the linguistic features were extracted automatically, human contribution to this step is minimal, so that there is no manual encoding that would need to be checked by a second encoder. The only qualitative judgments made concern the selection of linguistic features investigated, which are therefore explained in detail below.

First, we looked for indicators that provide useful measures for the level of competence ascribed to the robot. These comprise structuring functions, for instance, items like *now, next*, but also *another*. These structuring cues presuppose that the interaction partner keeps track of the interaction and builds up a coherent representation of what he/she/it encounters. Another indicator of ascribed competence in the current scenario are ascriptions of

memory and learning. For instance, if the robot is asked whether it memorizes something it had previously been told, this shows that participants expect that the robot learns and remembers what they teach it. Uses of past tense that refer to previous teaching sessions are indicators of such beliefs.

Second, in order to determine the social effects of the interaction, we investigated in how far users involve the robot directly. For instance, we counted instances of the personal pronoun *you*, instances of feedback signals, such as *good, well, excellent*, as well as instances of *yes* and *no*. Furthermore, we analyzed how often participants ask the robot questions, such as probing questions like *what's this?* and tag questions like *isn't it?*. Moreover, we looked at how often users call for the robot's attention by means of *look* or the robot's name.

We furthermore calculated the number of different words and, on the basis of the total number of words, the type-token ratio. The number of turns and the number of words are used to inform us on the one hand on how much effort the user put into the interaction, on the other, these numbers are used to calculate normalized numbers of the other features investigated, so that the numbers presented are always relative to the total number of turns or words used. The total numbers of turns and of the words used, as well as the type-token ratio, provide good indicators for how easy or difficult users make their utterances for their robotic partner. In speech to children, for instance, the number of different words and the type-token ratio are usually much lower than in speech to other adults (e.g. Snow [18]). Especially the diversity measure, i.e. the type-token ratio, thus tells us whether users simplify their speech for the robot. These features thus function as indicators of suspected competence. They are common measures in readability tests, and speech adjusted to linguistically somewhat limited communication partners, such as children, is generally simplified in these terms. The same holds for the mean length of utterance (MLU), which is reliably reduced in speech to children (cf. Snow [18]; Roy et al. [14]).

We finally encoded whether participants greeted the robot at the beginning of each session. Whether a user greets a robot or not has been found to be a reliable indicator of the degree of socialness

attributed to the robot, and as a useful predictor of the way this user will interact with the robot throughout the dialogs (Fischer [5, 6]; Lee et al. [12]).

Results

In order to assess the amount by means of which participants adjust their speech to the robot's behaviors over time, we compared the different sessions with each other, thus determining the likelihood that the interactions all stem from the same session. The results show that participants adjust their speech to the robot over time such that general tendencies in users' behaviors over time can be observed (see Table 1).

Table 1: Changes over time

	F(4,35)	p
turns	3.759235	0.012026
hello	0.261682	0.900508
words	1.150642	0.349159
diff_words	0.639448	0.637883
robot	1.000000	0.420651
now	0.770968	0.551458
another	1.607017	0.194327
interest	0.761221	0.557603
past	1.566045	0.204993
robot's name	0.764929	0.555259
look	3.204979	0.024169
lets	0.233275	0.917758
tag question	1.942775	0.125081
probing	1.822449	0.146530
expository	1.434195	0.243253
you	0.888735	0.480868
we	0.449707	0.771871
I	0.363485	0.832902
feedback	3.269179	0.022272
yes	1.406362	0.252151
no	0.891855	0.479093

MLU	5.429102	0.001651
typetoken	0.713872	0.588075

The analysis of the linguistic features shows that some significant changes occur over the five sessions. In particular, participants adjust the amounts of speaking such that the initial interactions are significantly shorter than especially the second interactions, and then interactions stabilize at a relatively high level. Thus, users spend different amounts of effort in the teaching sessions. Second, in the initial sessions, participants use significantly more devices by means of which they try to get the robot's attention; the number of instances of *look* is two-to-four times higher in the first session than in later sessions. In contrast, the number of feedback signals increases significantly over time, and most likely in correspondence to the robot's increasing linguistic capabilities. Finally, the mean length of utterance changes significantly after the first session and is adapted to the robot's linguistic capabilities in the later sessions.

As Table 1 shows, there are however no statistically significant differences in the amounts of structuring cues and references to the past, the use of the robot's name and other indicators of social relationship, tag questions and probing questions, pronouns, teaching strategies and linguistic diversity. Table 2 presents the means and standard deviations for the four features that change significantly during the five sessions:

Table 2': The four features 'number of turns', 'look', 'feedback' and 'MLU' across the five sessions

sessions	**turns**		**look**		**feedback**		**MLU**	
1	33.125	(5.16)	0.138	(0.14)	0.008	(0.016)	7.261	(1.528)
2	46.250	(8.28)	0.033	(0.03)	0.013	(0.020)	4.786	(1.137)
3	42.375	(8.57)	0.076	(0.08)	0.042	(0.063)	5.235	(1.393)
4	44.750	(5.39)	0.031	(0.03)	0.058	(0.051)	4.773	(1.409)
5	43.750	(9.41)	0.021	(0.04)	0.082	(0.066)	4.295	(1.546)

So people adjust their speech according to the developing capabilities of the robot, in particular with respect to the amount of effort put into the interaction (number of turns), their perception of the need to keep the robot's attention, the amount of feedback

given, and a central complexity measure, namely the mean length of utterances. At the same time, other linguistic features, which are generally subject to adjustments in child-directed speech, for instance, are not affected by the robot's increasing linguistic capabilities. Thus, participants do not structure the task more, do not reduce the number of different words, do not conceptualize themselves and the robot more as a team (as indicated by uses of 'let's' and 'we'), nor do they show differences in interpersonal relationships, such as by calling the robot's name, greeting it more, or referring less to themselves (by means of 'I') and more to the robot (by means of 'you'). While these features have been found to be affected by other aspects of robot behavior and embodiment, such as contingency of feedback and degrees of freedom (cf. Fischer, Lohan and Foth [9]; Fischer and Lohan [10]), they are obviously not affected by the robot's word learning.

However, besides for functional reasons, the failure to find more statistically significant differences between sessions may be due to high interpersonal variation. In a next step, we therefore investigated interpersonal differences in the interactions.

In order to assess the interpersonal differences between the eight different participants, we compared their linguistic behaviors in the five sessions with each other.

The investigation of differences in the linguistic features between participants shows that there are considerable differences between users throughout. In fact, only tag questions, number of turns, instances of 'look' and instances of *yes* are not significantly different between participants.

Table 3: Interpersonal Differences

	F(7,32)	p
turns	0.70765	0.665664
hello	2.92517	0.017482
words	4.47183	0.001450
diff_words	12.30139	0.000000
now	4.76811	0.000929
another	4.20360	0.002190
interest	2.56840	0.032148
past	3.97814	0.003117

robot's name	3.12951	0.012395
look	1.95766	0.092588
lets	4.37695	0.001676
tag questions	1.00065	0.448921
checking	2.54773	0.033312
expository	3.03480	0.014529
you	3.92834	0.003372
we	16.42579	0.000000
I	4.53277	0.001322
feedback	2.83286	0.020446
yes	1.56765	0.180879
no	6.10001	0.000142
MLU	2.62818	0.029008
typetoken	11.17388	0.000000

Thus, the analysis shows extreme interpersonal differences between speakers, basically concerning all linguistic choices. This suggests that participants differ considerably in their understanding of the situation (cf. Fischer [6]). However, while people differ in almost all linguistic behaviors, with respect to two of the four features that were found to be adjusted to the robot over time people converge in their linguistic choices; in fact, we can also understand the lack of differences in the use of 'yes' from the same perspective since the most important function of 'yes' is to provide feedback. The robot's developing capabilities can consequently be taken to guide people subtly into similar behaviors.

Discussion

The linguistic analyses presented show that the human tutors adjust their instructions to the robot's linguistic behavior over time. The linguistic features changed are functionally related to the different communicative tasks that users encountered in the five sessions. In particular, in the first session, users' communicative efforts largely concerned getting the robot's attention, which corresponds to the fact that the robot's only means of feedback was to display its attention nonverbally. So users' communicative focus in the first session is consistent with users' orientation at the

robot's behavior (Fischer et al. [8]). These communicative efforts change already in the second session when the robot starts producing verbal output.
The other changes made by the participants over the course of the sessions concern the mean length of utterance, the amount of speaking and the amount of linguistic feedback. These changes can be related to different tutoring behaviors on the one hand and the robot's increasing linguistic capabilities on the other. The changes observed are thus in accordance with a model of human-robot interaction that assumes high amounts of cooperation from the side of the users (cf. Fischer [5]) and considerable attention to the robot's capabilities (Fischer [7]; Fischer et al. [8]).
The results concerning interpersonal variation have shown that users' expectations and preconceptions play a considerable role in interaction. However, irrespective of their different preconceptions, all users converge on the same behaviors in response to the robot's behavior.

Conclusion and Future Work

We can conclude that both users' preconceptions and feedback-driven interactional effects shape human-robot interactions. While the initial differences between users persist over time, all users respond to the robot's feedback and increasing linguistic capabilities in comparable ways. Thus, the good news for robot developers is that the kinds of behaviors the robot produces subtly guide users into similar kinds of responses, irrespective of their initial expectations. Future work will have to identify the factors that lead to the high interpersonal variation identified – what makes participants understand the same human-robot interaction situation so differently that they make significantly different linguistic choices for their partners that persist over time? Furthermore, besides understanding interpersonal variation, it will also be useful if people's differing behaviors can be predicted; on the other hand, the current results suggest that the robot's behavior can guide people into particular behaviors; future work should thus explore in more depth how participants' ideas of the HRI situation and the robot's capabilities can be shaped.

References

[1] Amalberti, R., Carbonell, N. & Falzon, P. (1993): User Representations of Computer Systems in Human-Computer Speech Interaction. *International Journal of Man-Machine Studies* 38, 547-566.

[2] Clark, H.H. (1992): *Arenas of Language Use*. Cambridge University Press.

[3] Clark, H.H. (1996): *Using Language*. Cambridge University Press.

[4] Dautenhahn, K., Nehaniv, C. L., Walters, M. L., Robins, B., Kose-Bagci, H., Mirza, N. A., et al. (2009). Kaspar - a minimally expressive humanoid robot for human-robot interaction research. Applied Bionics and Biomechanics, Special Issue on 'Humanoid Robots', 6(3), 369–397.

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

[6] Fischer, K. (2011a). Interpersonal Variation in Understanding Robots as Social Actors. *HRI'11*, Lausanne, Switzerland.

[7] Fischer, K. (2011b). How people talk with robots – Designing Dialog to Reduce User Uncertainty. *AI Magazine,* 32(4):31-38.

[8] Fischer K., Foth K., Rohlfing K. and Wrede, B. (2011). Mindful tutors - linguistic choice and action demonstration in speech to infants and to a simulated robot. *Interaction Studies 12*(1), 134-161.

[9] Fischer, K., Lohan, K. and Foth, K. (2012). Levels of Embodiment: Linguistic Analyses of Factors Influencing HRI. *HRI'12*, Boston.

[10] Fischer, K. and Lohan, K. (submitted): How Robot Embodiment and Situatedness Influence Interaction. Cognitive Linguistics Yearbook, Berlin, New York: Mouton de Gruyter..

[11] Giles, H., Coupland, J. and Coupland, N. (1991), *Contexts of Accomodation. Developments in Applied Sociolinguistics*. Cambridge: Cambridge University Press.

[12] Lee, M.K., Kiesler, S. and Forlizzi, J. 2010. Receptionist or Information Kiosk: How Do People Talk with a Robot? *CSCW* 2010, Savannah, Georgia, February 6-10, 2010.

[13] Paepcke, S. & Takayama, L. (2010): Judging a Bot By Its Cover: An Experiment on Expectation Setting for Personal Robots. Proc. of Human Robot Interaction (HRI), Osaka, Japan.

[14] Roy, B., Frank,M. and Roy, D. (2009): Exploring word learning in a high-density longitudinal corpus. *Proceedings of the 31st Annual Meeting of the Cognitive Science Society*. Amsterdam, Netherlands.

[15] Saunders J., Lehman H., Sato Y. & Nehaniv C. (2011); Towards Using Prosody to Scaffold Lexical Meaning in Robots; Proceedings of ICDL-EpiRob 2011 : IEEE Conference on Development and Learning, and Epigenetic Robotics.

[16] Sacks, H., Schegloff, E.A. & Jefferson, G. (1974). A Simplest Systematics for the Organization of Turn Taking for Conversation. *Language* 50: 696–735.

[17] Saunders, J., Nehaniv, C. L. and Lyon, C. (2010) "Robot learning of lexical semantics from sensorimotor interaction and the unrestricted speech of human tutors," in Proc. Second International Symposium on New Frontiers in Human-Robot Interaction, AISB Convention, Leicester,UK, 2010.

[18] Snow, C.E. 1994. Beginning from baby talk: Twenty years of research on input and interaction. In: Gallaway, C. & Richards, B.J., eds. *Input and Interaction in Language Acquisition*, 3-12. Cambridge: Cambridge University Press.

[19] Turkle, S. (2006): A Nascent Robotics Culture: New Complicities for Companionship. AAAI Technical Report series, July 2006.

[20] Zuengler, J. (1991): Accomodation in Native-Nonnative Interactions: Going beyond the "What" to the "Why" in Second-Language Research. In Giles, Howard, Coupland, Justine and Coupland, Nikolas (eds.), *Contexts of Accomodation. Developments in Applied Sociolinguistics*. Cambridge: Cambridge University Press.