Mindful Tutors: Linguistic Choice and Action Demonstration in Speech to Infants and a Simulated Robot

Kerstin Fischer, University of Southern Denmark

Kilian Foth, University of Hamburg

Katharina Rohlfing, Bielefeld University

Britta Wrede, Bielefeld University

Abstract

It has been proposed that the design of robots might benefit from interactions that are similar to caregiver-child interactions, which is tailored to children's respective capacities to a high degree. However, so far little is known about how people adapt their tutoring behaviour to robots and whether robots can evoke input that is similar to child-directed interaction.

The paper presents detailed analyses of speakers' linguistic and non-linguistic behaviour, such as action demonstration, in two comparable situations: In one experiment, parents described and explained to their nonverbal infants the use of certain everyday objects; in the other experiment, participants tutored a simulated robot on the same objects. The results, which show considerable differences between the two situations on almost all measures, are discussed in the light of the computer-as-social-actor paradigm and the register hypothesis.

Keywords:

child-directed speech (CDS), motherese, robotese, motionese, register theory, social communication, human-robot interaction (HRI), computers-as-social-actors, mindless transfer

1. Introduction and Motivation

In this paper, we compare tutoring sequences in parent-child and human-robot interactions regarding a broad spectrum of verbal behaviours and action demonstration. In particular, we investigate linguistic choice, especially with respect to linguistic complexity, interactivity and verbosity, as well as demonstrating hand movements.

Comparing child-directed and robot-directed speech is both theoretically and practically interesting. While much is known about speech directed at infants and young children, little is known about how people talk to artificial agents. Speech to infants and children has generally been described as

highly adapted to the respective child's linguistic and cognitive development (e.g. Brown 1977) and as highly contingent to the child's own actions, such as gesture and gaze (e.g. Filipi 2009). Thus, fine-tuning and contingent reactions are understood as the main mechanisms underlying the process in which adults adjust to children (Snow 1987). In contrast, very different mechanisms have been proposed to characterize speech to computers or robots. In particular, for the communication with artificial communication partners, it has been suggested that speakers make use of their knowledge on how to speak to other humans. These proposals suggest a close relationship between parent-child and human-robot interactions: In particular, Nass and colleagues hold that people treat computers as social actors just like other humans (Reeves & Nass 1996, Nass & Moon 2000) and transfer mindlessly from knowledge on domains of social interaction between humans to artificial communication partners (Nass & Brave 2005). Thus, if the simulated robot resembles a child, people should treat it similar to a human child.

Similarly, the register hypothesis predicts a close relationship between child-directed speech and other simplified registers, an example of which is speech to computers. Register theory assumes that speakers have stored certain situation-specific speaking styles (e.g. Crystal 2001, Biber 2001), which are partly conventionalized and partly functionally motivated. While some authors have suggested speech to artificial communication partners to constitute a register by itself (Krause & Hitzenberger 1992), others propose that there is a set of simplified registers among which child-directed speech constitutes the prototype (Ferguson 1977, 1982; DePaulo & Coleman 1986). The prototype hypothesis predicts that people will take child-directed speech as the core example for simplified speech, from which speech in other situations, such as speaking to a cognitively and linguistically limited robot, is derived.

Both theoretical perspectives predict a high degree of similarity between child-directed and robot-directed speech, which should be particularly strong in our scenario in which the tasks the children and the simulated robot had to fulfil were adapted to the cognitive development of preverbal infants and in which the robot was designed to resemble a baby (Nagai & Rohlfing 2009, see Figure 1). However, especially with respect to linguistic behaviour and action demonstration very little concrete work has been carried out to back up the two hypotheses, and the current study fills this gap by presenting an in-depth analysis of comparable corpora on a broad scale.

Figure 1: Human-Robot Tutoring Scenario

On the practical side, several recent studies have shown that an understanding of how naïve users interact with artificial communication partners can improve automatic learning considerably. Since users have been found to provide a robot or interactive program with useful simplification and chunking (Thomaz & Cakmak 2009), some researchers have proposed to carry out machine learning tasks in interactive scenarios in general, which has lead to the socially guided machine learning paradigm (Thomaz & Breazeal 2008; Wrede et al. 2009).

Yet also in learning scenarios that are not interactional, input design for action and language learning can profit from the concrete analyses of the features available in child-directed, and possibly robot-directed, speech, in which case we may be able to exploit helpful cues from parent-child interactions for human-robot interactions (Rohlfing et al. 2006). For speech processing and dialog modelling at human-computer and human-robot interaction interfaces, it may be beneficial to be able to predict what a potential user is going to say in a given situation, allowing the designer to tailor the system specifically to the most probable kinds of utterances. Thus, the development of optimally adapted interface systems may profit from understanding how people talk to artificial agents.

Moreover, another practical benefit of a thorough understanding of the differences between speech to robots and speech to infants and young children concerns the potentially facilitative effects of child-directed speech for language acquisition that can be exploited in automatic grammar learning. In particular, there is both experimental evidence and evidence from simulation studies that the peculiar properties of child-directed speech may facilitate child language acquisition (e.g. Cartwright & Brent 1997; Theakston et al. 2005; Goldberg 2006; Onnis et al. 2008). Understanding the ways in which speakers adapt to the capabilities of children in comparison with robots could therefore be exploited in automatic language learning.

Finally, the present study combines analyses of two different modalities, allowing an integrated

perspective on linguistic and non-verbal cues in the two situations under consideration.

2. Previous Work

Numerous studies have shown that mothers, fathers, older siblings and other adults speak in similarly peculiar ways to infants and young children (e.g. Brown 1977, Snow 1994, Pine 1994, Brodsky et al. 2008, Roy et al. 2009). The adjustments found generally include higher pitch, more variable intonation contours, highly repetitive and reformulative speech, short utterances with simplified grammar, restricted lexis, and preference for basic level categories. This particular and quite homogeneous speaking style has been described as a register (Ferguson 1977, 1982; DePaulo & Coleman 1986), often also called *baby talk* or *motherese*. The speaking style has been found to change corresponding to the child's age (e.g. Snow 1972; Grimm 1999; Veneziano 2001). For instance, while prolonged acoustic patterns characterize the speech to younger infants, when children begin to speak, the input they receive is often scaffolded with respect to vocabulary and syntax (Snow 1977).

Also with respect to other modalities, adjustments have been detected. Adaptations concerning gesture have been called *multimodal motherese* (Zukow-Goldring 1996; Gogate et al., 2000). In their cross-cultural investigations, Gogate and her colleagues found that when introducing a new label for an object or one of its properties, mothers move the object in temporal synchrony with the new word. In addition, gestural behaviour was found to convey information about the referent that is redundant with the information provided in the speech addressed to infants (Iverson et al., 1999). Adaptations concerning action performance have also been observed and termed *motionese* (Brandt et al., 2002). When showing infants a new function of objects, adults seem to perform less round motion with more pauses within single actions (Rohlfing et al., 2006). Thus, adapted behaviour towards infants can be observed across different modalities.

With regard to speech directed to robots, much less homogeneity has been found; in fact, the comparatively few studies describing speech or gesture directed at computers or robots produce rather inconclusive and sometimes even contradictory results (cf. Fischer 2006; Herberg et al., 2008).

One proposal addressing the nature of HCI and HRI is the computer-as-social-actor hypothesis, which has been developed by Nass and colleagues (e.g. Reeves & Nass 1996, Nass & Moon 2000, Nass & Brave 2005). This hypothesis holds that people transfer the way they interact with other humans to interactions with artificial communication partners. The reason for treating computers just like people lies in, according to Nass (2004: 37), evolutionary psychology, since identifying

other humans constitutes a "significant evolutionary advantage". Nass & Moon (2000) refer to this process as *mindless transfer*.

The discovery procedure of the studies in this framework consists in taking a stable finding of human-human interaction and applying it to human-computer scenarios (cf. Reeves and Nass 1996). For instance, Nass (2004) describes an experiment in which participants first receive a tutoring session from a computer plus testing and evaluation. After that, one third of the participants fills out a questionnaire about the computer's performance at the same computer they have worked with, one third at another computer and one third on paper. The ratings of the computer's performance are significantly better if participants fill the questionnaire out at the same computer. In human communication, it is impolite to bluntly tell other persons that one does not approve of their performance since disapproval generally constitutes a threat to the communication partner's face (Brown & Levinson 1987); people will therefore make their judgements less face-threatening and thus more positive if they are evaluating the person directly than when they report their evaluation to another person. The study reported in Nass (2004) shows that the same effect can be found with computers. In similar studies, Nass and colleagues have investigated a broad range of such social behaviours, with the result that people were found to react similarly to the computer's behaviour (Reeves & Nass 1996; Fogg & Nass 1997), or that they transfer human characteristics to the agents, such as intentionality (Ju & Takayama 2008), ethnicity (Pratt et al. 2007) or gender, where, for example, a synthesized female voice will trigger the attribution of female characteristics to the computer persona (Nass & Brave 2005). For instance, concerning flattery, Reeves & Nass (1996:57-71) have found that people react to flattery from a computer in the same way as they react to flattery from humans; in particular, they like the computer better if it flatters the user irrespective of how justified the flattery is, and they like a computer that flatters another computer better than a computer that criticises other computers, but find the criticising computer more intelligent. Nass proposes that the reason for this transfer, which has also been called 'media equation' (Reeves and Nass 1996), is mindlessness, an error, albeit a sympathetic one: "polite responses to computers represent the best impulse of people, the impulse to err on the side of kindness and humanity" (2004:37).

Several studies support the predictions made by Nass and colleagues. For instance, Aharoni & Fridlund (2007) investigate how interviewees present themselves in job interviews carried out by a computer or a human (in fact it were the same ten prerecorded and disguised questions played by a wizard) non-verbally and how they are emotionally involved (pre- and post test questionnaire). Of all the non-verbal measures they used (smiles, frowns, yuck-faces, silence-fillers, self-manipulations and signs of embarrassment and politeness (all ranked on an absence – presence

basis, not based on frequency, p. 2177)), only smiles and silence fillers were different depending on whether people believed to be talking to a human (by long-distance call) or a computer. The most important effect on the judgement of their interviewer was whether participants had been accepted or rejected for the job. Regarding media equation, it seems that there were no differences between self-presentations for a computer or a human listener.

However, other authors report differences between human-to-human and human-computer interaction that should not occur if people transferred natural interactions mindlessly to interactions with computers. Amalberti et al. (1993) found considerable linguistic differences between human-to-human and human-to-computer interaction, even though the 'wizard', i.e. the human operator introduced as a computer to some participants and as a human to others, behaved identically in both situations. While the conceptualisation of the communication partner thus seems to play an important role, Amalberti et al. found that speech in the two corpora became increasingly similar over time, when speakers adapted to the features of their communication partners' linguistic behaviour. That is, those participants who believed to be talking to a computer used a speaking style that was initially significantly different from the style employed by those informed that they were talking to a human operator (who did not know in which group participants were), yet the differences disappeared over time since participants adapted their speech increasingly to the speech of their interlocutor.

Similarly, Okita et al. (2008) find that the mere conceptualisation of the communication partner as a human versus a machine influences the participants' learning behaviour considerably, and also Kanda et al's (2008) results on human-robot interaction call for a more diversified perspective on the phenomenon; they compare two humanoid robots and a human interactant with respect to verbal and non-verbal behaviours. They do not find significant differences in the amount of information presented and the amount of politeness formulas used, yet significant differences with respect to non-verbal behaviours: speakers bow deeper for their human interlocutor in the greeting, they respond more slowly to the robots' than to the human communication partner's greeting, they gesture less for the robots than for the human, they respond more slowly to one of the robots' pointing gestures, and they approach one of the robots more than the other and the human in terms of proximity. The results point to differences particularly on the social level.

That mindless transfer may provide only a partial explanation is also supported by studies extending the exploration of the computers-as-social-agents hypothesis. Johnson, Gardner & Wiles (2004), for instance, find that the flattery effect described by Reeves & Nass (1996) and Fogg & Nass (1997) only holds for some speakers and also under certain conditions. If transfer is mediated, however, it cannot be claimed to be mindless any more. Ju (2008) also investigates previous findings by Fogg

& Nass (1997) on flattery and finds that whether flattery is effective when coming from a computer depends on whether the information is presented verbally or textually (no effect for verbal presentation), whether it is presented to men or women (men are not affected by flattery if the computer persona is female, and even the opposite effect has been observed), and whether it is presented by a male agent image (the female image has no effect). Thus, if transfer is involved in these situations, it is mediated by a set of factors.

Further problems regarding mindless transfer have been encountered specifically with respect to linguistic interaction. In particular, Shechtman & Horowitz (2006) found that human-to-human and human-to-computer communication differ particularly along various social dimensions.

That interaction by means of language should be less viable to transfer than other social behaviours is particularly strange if we consider that language is generally taken to suggest anthropomorphism and to encourage mindless transfer; thus, Nass holds that natural language is among these cues that enforce mindless transfer, suggesting that certain computer interfaces 'trigger mindlessness'. In Nass (2004), the following set of 'triggers' for 'etiquette responses' are presented:

- language use
- voice
- face
- emotion manifestation
- interactivity
- engagement with and attention to user
- autonomy/unpredictability
- filling of traditional roles

The assumption that certain cues trigger transfer is supported by Gong's (2008) study on the role of anthropomorphism, in which she finds that the more anthropomorphic the artificial agent is, the higher the degree of transfer. FMRI-studies have furthermore shown that the more human-like a robot looks the more will the human interaction partner activate brain regions that are associated with reasoning about the interaction partner's intentions (Krach et al. 2008).

While it is plausible that situations that resemble human-to-human communication to a greater extent invite treatment of artificial communication partners as a social interactant more, it should be strange, since language use triggers mindless transfer, if speakers did not transfer mindlessly in verbal interactions with the artificial communication partner. Yet the results by Shechman & Horowitz (2006) as well as by older studies, such as Amalberti et al. (1993) or Johnstone et al. (1994), suggest that particularly in linguistic interfaces there is less transfer, especially with respect to social information.

Another problem is constituted by the fact that interactions between human adults are not homogeneous at all; factors like the task, the social status of the participants, the communication channel or the partner's native or non-native-like command of the language, among many others, influence the ways speakers make choices for their communication partners (e.g. Halliday 1985, Finegan & Biber 2001). Thus, if speakers transfer linguistic behaviours from human communication to human-computer interaction, one of the questions arising is which linguistic variety would constitute the source for the transfer. The variability of communications among human interlocutors constitutes a problem for the mindless transfer hypothesis, since it is unclear which variety speakers mindlessly transfer to human-computer interaction.

However, linguistic variation in human communication has been addressed in register theory; in order to account for linguistic variability between different communicative situations, the notion of register was developed, which describes linguistic variation according to situation. The register hypothesis assumes that speakers will use a particular speaking style specific for a given situation. While some aspects of this speaking style may be conventional (cf. Crystal 2001: 7, for instance), mainly the markers of this variety (Finegan & Biber 2001), there are also functionally determined statistical probabilities characteristic of language use in a given situation. One particularly well-studied linguistic register is child-directed speech, which has been found to differ from adult-directed speech in reliable ways (e.g. Pine 1994, Snow 1994, Söderstrom 2007).

Because of its pervasiveness and homogeneity and due to the fact that all speakers have encountered this speaking style during their own childhood, child-directed speech has been proposed to constitute the prototype of simplified registers, from which speech to other possibly restricted addressees is derived (Ferguson 1977, DePaulo & Coleman 1990). In other words, if speakers attempt to simplify their speech, for instance, when talking to foreigners, mentally challenged people or to pet animals, they are expected to draw on the speech forms they employ for children, which they got acquainted with in early childhood themselves.

So is the register hypothesis more suited to account for the findings? Regarding the register hypothesis, some problems arise, too. In particular, it predicts a) homogeneity, b) the existence of conventionalized markers, and c) functionally determined probabilities of choice based on the task. As we have seen above, and the results of the current study will support this, child-directed speech (CDS) is rather homogeneous and marked by peculiar linguistic features, but human-computer interaction is not; Fischer (2006) shows in great depth that human-computer interaction is very heterogeneous, and in many respects the choices made for artificial communication partners are by no means simpler than speech directed at human adults; for instance, speakers use technical vocabulary, explicit quantitative measures and external reference systems, none of which people use

to simplify their speech for children.

However, the register hypothesis may still be relevant concerning the prediction that speakers may use the simplifications of child-directed speech to simplify their speech for other communication partners. DePaulo and Coleman (1990) have addressed this problem in detail, distinguishing between several functions of CDS: expressing affect, securing the child's attention, simplifying speech, and clarifying speech, which are expected to play different roles in other situations in which only some of these components may be relevant. For instance, in speech to foreigners, simplification and clarification are expected, but not attention securing and affective functions. However, in their study, in which the authors compare CDS with speech to foreigners, mentally challenged and normal adults, DePaulo & Coleman (1990) do not establish a clear picture with respect to particular functions, nor do they provide evidence for the prototype hypothesis. In contrast, their results suggest that each linguistic feature may be used based on its particular function in each situation. To conclude, also the register hypothesis exhibits some problems and does not account for the findings so far.

To sum up, while CDS is relatively homogenous and characterized by fine-tuned responses to the child's cognitive and linguistic capabilities and by contingent interactional behaviour, computer- and robot-directed speech have been suggested to be based on mindless transfer or based on conventionalized speaking styles that are adapted to novel situations, such as talking to a computer or robot. Both proposals have been criticized especially with respect to their inadequacies to account for linguistic interactions with artificial communication partners. In this study, we address this relationship by comparing tutoring interactions in two scenarios that differ only with respect to the communication partner, who is the speakers' own child in the first corpus and a simulated robot in the other. In accordance with the discussion presented above, the focus of the comparison concerns the degree of similarity between particular behaviours in the two situations, particularly regarding the dimensions of verbosity, complexity and interactivity as well as action parameters.

3. Data and Methods

The main procedure used in this study is the quasi-experimental investigation of how people demonstrate and explain a set of objects to two different kinds of communication partners: their own children or a simulated robot.

3.1. Data
We investigate two comparable corpora taking a corpus-linguistic approach to the analysis of the

linguistic and non-verbal behaviours exhibited by two groups of participants in two sets of interactions. The first corpus (henceforth CDS) consists of parent-child interactions in which parents explain the functioning of certain objects and some actions to their pre-lexical children. The second corpus (henceforth HRI) consists of interactions in which participants explain the same objects and actions to a simulated robot (Nagai & Rohlfing, 2009).

Subjects

CDS: 28 German speaking parents, i.e. mothers and fathers, and their pre-lexical children, whose age ranged from 8 to 11 ($M = 10.25$) months, participated in this study. They were recruited through advertisement in local newspapers. Parents were compensated for their travel expenses (EUR 5), and children were given a small toy for their participation.

HRI: 30 German speaking participants (14 female and 16 male) were recruited for this study. In this sample, seven of the participants in this experiment were parents as well. Participants' age ranges from 18 to 63, and they come from various different fields; however, seven of the 30 participants have a computer science background. They were recruited on a word-of-mouth basis and compensated for their participation with a big bar of Swiss chocolate.

Procedure and Stimuli

CDS: The parent and the child were seated across a table. Then, the experimenter put a tray with a set of objects (e.g. three coloured cups of different sizes) in front of the parent. The arrangements of objects was fixed across subjects and allowed for a comparable starting position. In the next step, the experimenter instructed the subjects, for example, "Zeigen Sie, wie die Becher ineinander gestapelt werden können [Please show how to stack the cups into each other]". Other tasks were to explain, for instance, how to switch on the light by pulling a string on a lamp, how to ring a bell, how to sprinkle salt from a salt carrier, as well as how to stack blocks onto each other. Parents were asked to show the function of the objects first before they were allowed to move the tray more closely to the child and to let him or her play with it. The data reported on here encompasses the first part, the demonstration in which the children were observing the actions performed.

HRI: The procedure of the experiment with the simulated robot was similar to the procedure in the experiment with children. The participants sat across a table and faced the simulated robot 'Babyface'. In the next step, participants were instructed to 'explain' the objects to the simulated robot, which was introduced by its Japanese name, *Akachang*, in order not to bias the participants according to its suggested age. The same objects as in the experiment with the infants were used. The simulated robot provides feedback in the form of eye gazing behaviour based on visual saliency. This means that the simulated robot gazes at salient points within objects which are

derived from movements, colours, orientation, flicker and intensity of objects or persons in the visual field (Nagai & Rohlfing, 2009). In addition, its eyelids blink randomly, and its mouth opens and closes randomly. Note that, unknown to the users, this feedback model in form of the gaze behaviour is driven by a purely data-driven reactive saliency-based visual attention model that does not take any prior (e.g. semantic) knowledge nor the participant's linguistic utterances into account.

Concerning the behaviours Nass (2004) suggests as prompting anthropomorphism, the simulated robot used here suggests implicitly an understanding of natural language but no capacity to produce it, and it has no voice. In contrast, it has a face and is meant to express some emotion, i.e. friendliness, which however does not change. It is interactive within the boundaries of its limited capabilities, to follow participants' movements with its eyes. It engages with and attends to the user, and it is autonomous with respect to its limited behaviours. It fulfils the role of an instructee, although it is unclear, for instance, what the simulated robot needs to learn how to operate a lamp for. Thus, the simulated robot fulfils several of the criteria mentioned by Nass (2004) that are suspected to trigger mindless transfer.

3.2.1. Data Analysis: Linguistic Coding

Both corpora were manually transcribed and syntactically analysed. The linguistic analysis was carried out using the constraint-based parser described in Foth et al. (2000). This system performs morphological classification and syntactic and referential dependency analysis on the word level and assigns every dependency to one of 35 syntactic classes; it also computes a measure of how well each utterance adheres to the norm of the standard grammar. The output format allows the quick computation of basic frequency counts such as mean length of utterance (MLU) or category distribution, but also supports searches among inflected words for their stems, or for the syntactic roles of words. The label set employed allows distinctions such as those between subjects, direct objects, and indirect objects, or between active and passive voice, to be retrieved easily. To rule out distortions of our results due to any systematic imperfections of the parsing accuracy, all analyses were fully verified for correctness manually, i.e. the automatic analysis served only to speed up the annotation process. The linguistic analysis concerns three different factors: verbosity, complexity and interactivity.

Linguistic verbosity
The first general property investigated concerns the amount of speech presented to a communication partner, i.e. linguistic verbosity. The verbosity measures tell us about how much

effort speakers spend on each task and how much information they consider suitable or necessary for their communication partner to understand, thus providing indirect information about speakers' recipient design for their respective communication partners. Moreover, the number of different words tells us about the suspected competence level of the communication partner. Thus, to begin with, for each corpus we counted the **total number of words** and the **number of different words** per speaker in each of the six tasks as well as **number of utterances** per task.

Complexity of utterances

The second measure concerns the complexity of utterances; as we have seen, CDS has been shown to constitute a simplified register, and human-robot interaction has been suggested to be derived from that. Thus, the degree of complexity in the tutors' utterances plays a crucial role with respect to the theoretical proposals. Complexity is also of practical interest since the more simplified robot-directed speech is, the easier it can be made use of in language processing or language learning tasks. In the following, several dependent measures are presented and proposed as operationalizations of the complexity of utterances.

A very common measure (Snow 1977) of sentence complexity is the **MLU**, the mean length of utterance. To calculate the MLU, we simply divided the number of words per speaker by the number of utterances by the same speaker. By utterance we understand all turn-constructional units, that is, units consisting of clause complexes, of single clauses, but also smaller units, such as noun, verb or prepositional phrases, answer particles and feedback signals (Sacks et al. 1974; Ford, Fox & Thompson 1996).

Another measure of complexity, and at the same time a feature revealing the suspected competence of the communication partner, concerns the **concreteness** versus **abstractness** of terms used. Whereas parents have been found to often use concrete, basic level terms (Rosch & Mervis 1973), such as *cup*, *bowl*, or *block*, when communicating with their children, people interacting with computers have been suggested to use more abstract terms, such as *object, container* or *obstacle* (cf. Fischer 2006).

Furthermore, some structures are more complex than others. The **passive**, for instance, is a structure that is acquired quite late in the development (cf. Abbot-Smith & Behrens 2006). It introduces a perspective in which the patient or undergoer of an action is foregrounded and the agent is backgrounded. The construction is also formally quite complex and thus a useful indicator for assumed competence.

Sentence complexity is also reflected in the number and type of objects used. In particular, we distinguish between **direct objects, indirect objects** and **object complement clauses**, for instance, *she hit it, she gave him the ball*, and *she said that it is sad*, respectively. As, for instance, Hawkins

(1994) shows, these three types of objects exhibit increasing degrees of complexity. Relative clauses, such as *the man who walks on the other side of the street is my uncle*, have been found to be good indicators of suspected partner competence and linguistic proficiency; thus, in human-robot interaction speakers only use **relative clauses** if they are certain to be understood or if their partner uses them as well (Fischer 2006). We therefore take it as an indicator for complexity here.

Embedding is a composite feature, combining all structures that can be embedded in the main sentence structure, such as relative clauses, object complement clauses, dependent main clauses, subclauses, appositions, infinitival complements, and subject clauses. In particular, we use the following definitions: Subclauses are subordinate clauses like *whenever he goes to school, he feels sick* which in German exhibit a characteristic verb-last word order, for example, *wenn er in die Schule geht* (V)*, wird ihm ganz schlecht*. Appositions are added elements, such as *see the button, the red one*. An example for an infinitival complement is *she wants to go* and for a subject clause *what she really wants is love*.

Pragmatic function

A third property concerns the amount of social information used and the degree with which speakers involve their communication partner. Here, we distinguish between attention-getting and response-eliciting pragmatic functions and basic grammatical repercussions of interpersonal relationships. Thus, one feature concerns the sentence type, in particular, imperative, declarative, interrogative or infinitive mood. The declarative is generally used to make assertions. Furthermore, instructions by means of **declarative** sentences are very common, thus avoiding that the speaker directly imposes his or her wishes onto the communication partner, as it is the case with a simple imperative, such as, for instance, *move*! In German, imperatives, which directly involve the addressee, on the other hand, are often toned down by means of **modal particles**, sentence medial particles that serve downtoning and grounding functions (cf. Fischer 2007). In the current data sets, the down-toned imperative occurs frequently in attention getting functions, such as *guck mal* (*look*).

In situations without a concrete addressee, such as on public signs (cf. Deppermann 2007), or with a highly unfamiliar addressee, such as a computer or robot (cf. Fischer 2006), instructions and explications using the **infinitive** are very common, for instance: *den blauen nehmen*; this corresponds roughly to the English use of the gerund, as in, for example, *no smoking*. Moreover, speakers can ask **questions** to involve their addressees, or they can use understanding **checks**, such as tag questions like *doesn't it* or *don't you* in English and *ne?* in German.

Also, personal pronouns are useful indicators of the relationship between speakers in a communicative situation. For instance, speakers may avoid addressing the partner, using the impersonal form *man* (*one*). Alternatively, speakers can address their partner using *du* (*you*), or they can refer to themselves with or without including the partner, using either *ich* (*I*) or *wir* (*we*). Similarly revealing regarding the degree with which the communication partner is involved is the use of the **vocative**, for instance, the partner's first name.

The absolute occurrences of these features, besides the verbosity features, were counted in the six comparable tasks per person in the two conditions and divided by the number of utterances used by this person. The numbers reported are thus the numbers occurring per speakers' utterances.

3.2.2. Data Analysis: Action demonstration

Both corpora were analyzed for modifications in manual demonstration behaviour. The dependent variable in this analysis was the motion of the hand captured as a 2 D hand trajectory during a task performance. This semi-automatic tracking method was applied to the video only and did not require the subject to wear markers (which might have been obtrusive). The system, described in Vollmer et al. (2009), allows for tracking both hands with an Optical Flow based algorithm. The capturing of the hand trajectory was supervised by a human coder. In the case of tracking deviation, manual adjustments were made. For the results presented below, we chose a task in which demonstration allows high comparability across subjects: a block world task (s. Figure 2). The demonstration of the block world task was evaluated using movement parameters in both the robot- directed and the child-directed scenario.

*Figure 2: Action Demonstration in
the Blockworld Task*

For the analysis of the action needed for the blocks, the presentation was broken down into three sub-actions, each consisting of grasping one block until releasing it into the end position. We use the term *action* in reference to the whole process of transporting all objects to their goal positions;

with the term *movement*, we refer to phases in which the velocity of the hand is above a certain threshold, and all other phases are defined as pauses.

The data of the hand trajectory provided the following movement parameters (cf. Vollmer et al., 2009; Rohlfing et al., 2006):

- **movement velocity** was computed using the derivative of the 2-dimensional hand coordinates of the hand which performed the action per frame; in Rohlfing and colleagues (2006), a statistically significant trend was observed suggesting that hand movements in adult-adult interaction are faster than hand movements in adult-child interactions;

- **acceleration** was defined as the second derivative of the hand trajectory;

- **pace** was defined for each movement by dividing the duration of the movement (in ms) by the duration of the preceding pause (in ms); in Rohlfing et al. (2006), pace was lower in child-adult than in adult-adult interaction meaning that more pauses between the single movements were identified;

- **roundness** was defined for each movement by the motion path covered (in meters) divided by the distance between motion on- and offset (in meters) meaning that the rounder the motion, the higher the roundness value;

- **motion pauses** are characterized by their frequency defined as the number of motion pauses per minute, the average length (in frames) and the total length computed as the percentage of time of the action without movement.

3.3 Statistical Analysis

The corpora described above were analyzed according to the features outlined, and statistical analyses were carried out to determine the differences between the two corpora using a univariate ANOVA with the different addressee (robot vs. child) as an independent variable.[1] Furthermore, a questionnaire, which had been given to each of the participants in the human-robot condition at the end of the interaction, was used to interpret the results of the quantitative comparison.

The questionnaire comprised questions for participants' age, occupation, computing background and gender, as well as questions concerning participants evaluation of the interaction and the robot and their attention to particular robot behaviours.

4. Results

---

1 Since univariate analyses of variance provide identical results to t-testing, we decided for the former.

In the following, we present the results with respect to all dependent variables.

4.1. Linguistic Choice

The linguistic analyses show that there are significant differences with respect to almost all measures concerning complexity and interactivity of the data. Only with respect to the amount of speaking there are fewer differences.

4.1.1. Verbosity

The first measure investigated is verbosity. As Table 1 shows, participants do not differ significantly in terms of the numbers of turns used. However, speakers in the human-robot condition used significantly more different words than the parents.

Table 1: Linguistic Characteristics: Verbosity

	CDS M	HRI M	CDS SD	HRI SD	$F(1,57)$
number of turns	54,52	49,27	31,13	24,25	0,510
diversity	279,44	363,40	166,50	133,22	4,46*

$+p<0.1$, $*p<0.05$, $**p<0.01$, $***p<0.001$

Thus, parents tend to talk to their children more, yet they use the same words over and over again whereas in the interactions with the simulated robot, participants use a more diverse vocabulary.

4.1.2. Complexity

Regarding the complexity measures, as can be viewed in Table 2, all except one show that speech to the robot is significantly more complex than speech to children.

Table 2: Linguistic Characteristics: Complexity

	CDS M	HRI M	CDS SD	HRI SD	F
MLU	5.20	8.40	1.07	2.71	32.97***
embedding	0.07	0.19	0.07	0.09	26.76***
abstract nouns	0.00	0.05	0.01	0.06	16.26***
subclauses	0.01	0.11	0.02	0.07	43.41***
relative clauses	0.01	0. 02	0.01	0.03	8.54***
passive	0.00	0.04	0.01	0.08	6.05**
object subclauses	0.00	0.02	0.00	0.03	5.52*

$+p<0.1$, $*p<0.05$, $**p<0.01$, $***p<0.001$

- The mean length of utterances (MLU) is much higher in HRI than in CDS, and the difference is statistically highly significant.
- Furthermore, utterances in HRI are much more complex than those in the CDS corpus in that they exhibit many more embedded structures.

- The number of abstract nouns is significantly higher in HRI than in CDS.
- The number of subclauses differs significantly between the two corpora; in the HRI corpus, speakers use ten times more subclauses than in the CDS corpus.
- The number of relative clauses is, accordingly, also much higher in HRI than in CDS.
- Speakers use significantly more passive constructions in HRI than in CDS.
- There are furthermore significantly more object subclauses in the HRI corpus.

The results thus suggest that on the whole, talking to the 'Babyface' robot is characterized by more complex language than talking to preverbal children.

4.1.3. Interactivity

As can be seen in Table 3, in contrast to the complexity features, which are consistently higher in HRI than in CDS, the HRI corpus exhibits many fewer instances of attention-getting and other interactive devices than the CDS corpus. These differences are also highly significant.

Table 3: Linguistic Characteristics: Interactivity

	CDS M	HRI M	CDS SD	HRI SD	F
Mp	0. 31	0. 10	0. 15	0. 07	46.61***
Imperative	0. 23	0.02	0. 13	0. 03	85.40***
Question	0.16	0. 04	0.10	0. 05	39.79***
Check	0. 04	0. 00	0. 04	0. 00	43.21***
Vocative	0. 08	0. 01	0. 06	0. 03	27.52***
'you'	0.09	0. 03	0.07	0. 09	7.90*

$+p<0.1$, $*p<0.05$, $**p<0.01$, $***p<0.001$

In particular, the following features can be found:

- Many more speakers use the imperative in the CDS than in the HRI corpus. The imperative directly addresses the communication partner and thus implies a direct involvement of the interlocutor.
- Speakers also use more modal particles (mps) in CDS than in HRI, which corresponds to the increased use of imperative mood in CDS.
- Speakers also address their communication partner more often directly with the second person personal pronoun 'you' in CDS than in HRI.
- Similarly, they use the vocative more often in CDS than in HRI.
- They also involve the communication partner more often with checking signals in CDS than in HRI.
- Speakers also ask their children more questions than the simulated robot.

Considering the distribution of the sentence types, we can find that speakers use significantly more imperative, vocative and interrogative mood in CDS than in HRI and that speakers use more declarative sentences in HRI than in CDS instead. That means that speakers use sentence structures that are designed to describe a state of affairs rather than sentence structures whose core function is to elicit behaviour or information from the communication partner (cf., for instance, Halliday 1985). Only the infinitive, which is neutral to these uses, shows non-significant distributions.

Table 4: Linguistic Characteristics: Grammatical Mood

	CDS *M*	HRI *M*	CDS *SD*	HRI *SD*	*F*(1,57)
Declarative	0. 44	0. 91	0. 17	0. 26	68.21***
Imperative	0. 23	0.02	0. 13	0. 03	85.40***
Infinitive	0. 04	0. 06	0. 04	0. 05	2.21
Vocative	0. 08	0. 01	0. 06	0. 03	27.50***
Interrogative	0. 16	0. 04	0. 10	0. 05	39.79***

+p <0 . 1 , *p <0 . 0 5 , **p <0 . 0 1 , ***p <0 . 0 0 1

The results from the linguistic analyses thus suggest that the speech to the simulated robot is characterized by fewer attempts at involving the addressee on the one hand and by more complex sentence structures on the other.

4.3. Motion Analysis

Grouped into the different movement parameters (s. the first column of Table 5), the data from the hand trajectory during action demonstration was submitted to a one-way ANOVA. This allows a comparison between child-directed and robot-directed demonstrations; Table 5 summarizes the results.

Table 5: Child-directed demonstration (CDD) versus robot-directed demonstration (RDD)

	CDD*M*	RDD *M*	CDD*SD*	RDD *SD*	*F*
Velocity	0.13	0.09	0.03	0.02	14.57***
acceleration	0.04	0.02	0.01	0.00	17.37***
Pace	4.33	2.95	1.91	1.33	4.37*
roundness	2.67	1.80	1.20	0.72	4.83*
Pauses: frequency	41.57	43.37	5.75	5.91	0.62
Pauses: average length	9.31	12.77	3.68	4.67	4.47*
Pauses: total length	25.68	35.97	10.10	10.89	6.24*

+p <0 . 1 , *p <0 . 0 5 , **p <0 . 0 1 , ***p <0 . 0 0 1

The analysis of the movements performed during the demonstration of the blocks reveals significant differences in all movement parameters comparing child-directed to robot-directed action. In

contrast to the results obtained by Vollmer and her colleagues (2009) on another task (stacking cups), our results show significance for the parameter roundness as well. According to our results, when a tutor demonstrates a function of an object towards the simulated robot, the movement of the demonstrating hand is slower, less accelerated and less round when compared to a demonstration towards a child. In addition, a significantly lower value of the pace parameter could be observed in robot-directed demonstration suggesting that more pauses between the single movements were performed. This indicates that complex actions are similarly decomposed into shorter movements, in which the objects, their function or properties can be highlighted (Rohlfing et al., 2006). Concerning the motion pauses, the data show that even though pauses were more frequent when the robot was addressed, this difference was statistically not significant.

Taken together, our results suggest that an action is demonstrated differently when the recipient is a simulated robot in comparison to child-directed interaction, and the movement parameters are even more accentuated. Interestingly, in contrast to child-directed movement – which is very variable across subjects – we observe less variability in robot-directed movement.

4.4 Questionnaire Results

The last set of results concerns participants' answers in the questionnaire. First, we asked participants to estimate the robot's age; the average age suggested for the simulated robot, although it was intended to resemble a 'baby' (Nagai & Rohlfing 2009), was 4.7 years, ranging from 0.5 to 10 years. Thus, the suspected age of the robot does not correspond to the age of the children addressed in the CDS corpus. However, the suspected age of the robot correlates linearly only with few of the linguistic features analysed, in particular with checking signals ($r= .47$, $p< .05$) and the use of abstract nouns ($r= .35$, $p< .05$); that is, for most features concerning linguistic complexity, there is no linear relationship between suspected age and the use of complex language such that those speakers who suspect the robot to be older use more complex features. However, although speakers' linguistic choices can thus not be regarded as suitable adaptations to the expected age and mental capabilities of the simulated robot, the high mean and the broad range in the suspected age indicate that participants were expecting higher cognitive capabilities than those of a preverbal infant.

Second, we asked participants whether they witnessed any differences in the robot's reaction to the bell compared to the lamp; while the robot is sensitive to differences in brightness, it does not react to sound at all. In the questionnaire, 16.6% of the participants replied that they had not watched the robot closely at all. Some of them had not even realized that the robot's eye gaze followed their movements. Thus, participants' expectations of the robot's capabilities were so low that they did not

pay enough attention to its behaviour to notice that it was indeed interacting with its environment.

In contrast to those participants who were not aware of the robot's capability regarding eye-gaze, several other participants reported behaviours of the robot that it did not have, such as signalling understanding by means of a smile, a twinkle or a clicking sound (16.6%) or reaction to sound (10%). One participant remarked that she found the robot to be "alert and interested." Two participants moreover blamed themselves explicitly for the possibility that the robot may have failed to understand their instructions.

5. Discussion

The comparison of child-directed versus robot-directed speech presented is intended to shed light on the question whether the 'mindless transfer hypothesis' or the 'register hypothesis' accounts for the robot-directed speech; both hypotheses predict considerable similarities between the two scenarios.

Our analyses however show that the two corpora, in spite of the similar tasks, are significantly different with respect to almost all parameters investigated. Thus, we can conclude that people do not talk to the simulated robot as they talk to a human child. Furthermore, the differences observed are highly systematic: while those linguistic features that occur more frequently in HRI than in CDS are exclusively indicators of a much more elaborate partner model since they are all related to higher complexity and linguistic diversity, those features that occurred less frequently in HRI than in CDS concern, without exception, a higher amount of addressee orientation. Similarly, the parameters in speakers' action demonstrations uniformly point into the same direction, namely that actions are demonstrated more slowly when the recipient is a robot instead of a child and that the movement parameters are even more accentuated. Thus, there are significant differences between the corpora, and these differences are not random at all. Instead, the linguistic utterances directed at the simulated robot are linguistically significantly more complex and significantly less interactive; at the same time, gestures for the robot are significantly less complex than those addressed to the infants.

Now, the questionnaire data may suggest that participants were targeting an older communication partner than the parents; thus, one might be inclined to suggest that participants do treat the robot just like another human, just not like a pre-verbal child; however, the results on action demonstration show that participants in the HRI sessions slowed the demonstrations down even more and made the trajectories even more pronounced for the robot than the parents for the much

younger infants. Furthermore, as we have seen above, only few of the linguistic properties investigated correlate with the suspected robot's age. Thus, the results of this study cannot be explained away with the argument that the preverbal infant as an addressee does not provide an appropriate point of comparison for the HRI data. That is, while participants in the interactions with the simulated robot judged the robot's age very differently, the average higher age suspected for the robot does not correspond to speaking to older children since participants' action demonstrations were clearly designed for an even less mature communication partner than the 9-11 months old infants. Consequently, participants in HRI did not simply transfer from interactions with older children. Furthermore, the fact that the age range suggested for the simulated robot is so varied suggests that participants' conceptualizations of the robot were not very homogeneous; yet instead this being a backdrop in the results, we firstly have to acknowledge that participants do not make the same sense out of the robot's behaviour in a unified way; secondly, the considerable differences rather point to differences in conceptualization of the robot and thus to individual sense making procedures.

Given the spectrum of changes observed, we can thus conclude that participants do not transfer mindlessly from caregiver-child interactions to adult-robot interactions, nor do they employ a babytalk register to interact with the simulated robot. Not only do their behaviours differ consistently on all levels investigated, they also do so systematically.

The systematicity in participants' adaptations in the interactions with 'Babyface' corresponds not to the robot's actual capabilities, but to participants' perceptions of the robot's capabilities as indicated by its feedback. In particular, participants did not receive feedback from the robot on the appropriateness of their linguistic instructions; we thus find a high variability in this area as well as utterances that are much more complex than utterances directed at a child who contingently displays either understanding or lack of understanding. Furthermore, attempts at linguistic involvement of the simulated robot would have been futile since it does not have any linguistic production capabilities, whereas attention-getting cues are highly effective for children. However, participants did receive feedback on their motion trajectories and action demonstrations from the robot's eye movements. That this domain should thus be the only one exaggerated for the robot compared to the child is consistent not with the robot's true capabilities (in which case they should not have spoken at all) but with the feedback from the robot that they received. This indicates that participants were not at all transferring mindlessly, but choosing deliberately on the basis of the perceived requirements of the current interaction. This does not imply that these choices are being made consciously; rather, they are functionally appropriate and rest on participants' analyses of the affordances of the current communicative situation. Participants' choices are neither based on

transfer, nor on conventions since they are perfectly adjusted to the respective communication partner's capabilities as inferable from its feedback. Thus, the interaction with the simulated robot has to be understood as guided by similar mechanisms as child-directed speech (cf. Snow 1987), such that participants also fine-tune their linguistic choices to their artificial communication partner's suspected capabilities and provide contingent responses to the robot's non-verbal behaviour.

More concrete support for the argument that participants are not transferring mindlessly or employing a conventional register but rather adapt their instructions on the basis of the robot's feedback comes from the analysis of the motion and pause parameters, which reveals that the movements in robot-directed demonstrations are even more accentuated than in the child-directed scenario and that actions were performed with longer pauses, suggesting that the tutors were awaiting feedback from the robot. This is a clear indication for participants' attention to the robot's capabilities, showing that participants are not transferring mindlessly behaviour from interactions with preverbal children, but are choosing their instruction strategies as they understand them to be appropriate for the current interlocutor. Thus, even though the behaviour parameters used in interactions with a robot are comparable with, and possibly inspired by, child-directed interaction, they are online customized for this particular interlocutor.

Finally, the questionnaire data show that participants in the human-robot interactions enter into a complex sense-making activity, as it has been suggested in recent literature on context and situation construal (e.g. Schegloff 1997; Fischer 2000, 2006; Fetzer 2004). This seems to be an active online process and thus also different both from mindless transfer and the register hypothesis.

6. Conclusion

We can conclude that mindless transfer seems to be suitable to account for some behaviours in interactions with artificial communication partners, yet that the situation may be more complicated especially with respect to linguistic and gestural choice.

Overall, the results do not rule out a combination of the mindless transfer hypothesis and the register hypothesis in that human subjects mindlessly transfer their behaviour from a known register which is most similar to the encountered situation (in this case child directed interaction). More plausible is, however, that speakers design their utterances on the basis of what they understand to

be at issue in the given situation. The heterogeneity of both the linguistic behaviour and participants' estimations of the robot's age indicate that the human-robot situation is not clearly defined for the participants. Yet, they all assume that slowed down, pronounced demonstrating actions are useful, which corresponds to the single type of feedback the robot produces: it follows moving objects and gestures with its eye gaze. At the same time, there is much greater uncertainty about its linguistic capabilities, which corresponds to the higher degree of complexity uniformly observed throughout the corpus and to the fewer instances designed to involve the addressee. Given that the robot does not provide any feedback on the linguistic instructions nor reacts to any attempts at involvement, participants' behaviour is situationally adequate. Thus, the linguistic and non-verbal behaviours observable in HRI are functionally appropriate given the characteristics of the current communication situation. This perspective casts considerable doubt on both the mindless transfer and the register hypotheses, since participants rather employ linguistic and non-linguistic features on the basis of their function only and based on what was available to them through the robot's feedback.

For the development of robotic systems[2] this means that if participants orient at the capabilities and affordances of their communication partner, the robot's feedback is crucial. That is, participants will adapt to a large degree to what they consider to be "at issue" in a given situation, which is influenced by what they perceive the robot to understand. This means for robot design that the robot's behaviour has to be congruent with the robot's real capabilities since users will use it to infer its capabilities and to build up a partner model that then determines their own behaviour in the interaction with their artificial communication partner. The closer the robot's feedback corresponds to its real processing states, the more users' behaviour will match the robot's current needs.

This interpretation is supported by studies on the lasting effect of users' preconceptions; Paepke & Takayama (2010), for instance, find that if users have low initial expectations, they rate the robot's behaviour consistently better than when their initial expectations are high. Subtly guiding human users into appropriate conceptions of the robot's capabilities seems therefore a suitable approach to

---

[2] A possible limitation concerning the generalisability of the results obtained could be that the artificial communication partner studied here is only a simulated robot on a computer screen and maybe an embodied robot would be tutored differently. Lohan and her colleagues (2010) addressed this question recently and found that the difference in tutoring behaviour is rather due to the difference in degree of freedom (i.e. how much feedback behaviour a robot actually exhibits) than simply to embodiment. Similarly, Powers et al. (2007) found rather few differences between simulated and actual robots.

predicting users' behaviour. In the long run, it would therefore also be desirable if the robotic system was able to build up an internal representation of the ongoing interaction and to give feedback that is understandable for human tutors such that they can appropriately adapt their teaching behaviour to the current cognitive needs of the system.

References

Abbot-Smith, K. & Behrens, H. (2006). How Known Constructions Influence the Acquisition of Other Constructions: The German Passive and Future Constructions. *Cognitive Science*, 30, 995-1026.

Aharoni, E. & Fridlund, A.J. (2007). Social Reactions toward People vs. Computers: How Mere Labels Shape Interactions. *Computer in Human Behavior*, 23, 2175-2189.

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.

Brand, R. J., Baldwin, D. A. & Ashburn, L. A. (2002). Evidence for 'motionese': modifications in mothers' infant-directed action. *Developmental Science*, 5, 72–83.

Brodsky, P., Waterfall, H. & Edelman, S. (2007).Characterizing Motherese: On the Computational Structure of Child-Directed Language. *Proc. Cognitive Science Society Conference.*

Brown, R. (1977). The Place of Baby Talk in the World of Language. In C. Snow & C.A. Ferguson (Eds.), *Talking to Children: Language Input and Acquisition*. Cambridge: Cambridge University Press.

Brown, P. & Levinson, S. (1987). *Politeness. Universals in Language Use*. Cambridge: Cambridge University Press.

Cameron-Faulkner, T., Lieven, E.V., & Tomasello, M. (2003). A construction based analysis of child directed speech. *Cognitive Science*, 27, 843-873.

Cartwright, T.A. & Brent, M.R. (1997). Syntactic Categorization in Early Language Acquisition: Formalizing the Role of Distributional Analysis. *Cognition*, 63, 121-170.

Crystal, D. (2001). Language and the Internet. Cambridge: Cambridge University Press.

DePaulo, B. M. & Coleman, L. (1986). Talking to children, foreigners, and retarded adults. *Journal of Personality and Social Psychology*, 51, 945-959.

Deppermann, A. (2007). Grammatik und Semantik aus gesprächsanalytischer Sicht. Berlin/New York: de Gruyter.

Ferguson, C.A. (1977). Baby Talk as a Simplified Register. In Snow, C.E. & Ferguson, C.A. (Eds.). *Talking to children: Language input and acquisition*. Cambridge: Cambridge University Press.

Ferguson, C.A. (1982). Simplified Registers and Linguistic Theory. In Obler, L. & Menn, L. (Eds.), *Exceptional Language and Linguistics*. New York: Academic Press.

Fetzer, A. (2004). *Recontextualizing Context: Grammaticality Meets Appropriateness.* Amsterdam: Benjamins.

Filipi, A. (2009). *Toddler and Parent Interaction*. Amsterdam/Philadelphia: John Benjamins.

Finegan, E. & Biber, D. (2001). Register variation and social dialect variation: The register axiom. In P. Eckert & J. R. Rickford (Eds.), *Style and Sociolinguistic Variation*, 235-67. Cambridge: Cambridge University Press.

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

Fischer, K. (2007). Grounding and common ground: Modal particles and their translation equivalents. In A. Fetzer & K. Fischer (Eds.), *Lexical Markers of Common Grounds*. Studies in Pragmatics 3. Amsterdam: Elsevier.

Fogg, B.J. & Nass, Clifford (1997). Silicon sycophants: the effects of computers that flatter. *International Journal of Human-Computer Studies*, 46, 5, 551 - 561.

Ford, C. E., B. A. Fox & S. A. Thompson (1996). Practices in the construction of turns: The TCU revisited. *Pragmatics*, 6, 427-454.

Foth, K., Menzel, W. & Schröder, I. (2000). A Transformation-based Parsing Technique with Anytime Properties. 4th Int. Workshop on Parsing Technologies, IWPT-2000, 89 - 100.

Goldberg, A.E. (2006). *Constructions at work: The nature of generalization in language*. Oxford: Oxford University Press.

Gogate, L. J., Bahrick, L. E. & Watson, J. (2000): A study of multimodal motherese: The role of temporal synchrony between verbal labels and gestures. *Child Development*, 71, 878–894.

Gong, L. (2008). How Social Is Social Responses to Computers? The Function of the Degree of Anthropomorphism in Computer Representations. *Computers in Human Behavior*, 24, 1494-1509.

Grimm, H. (1999). Störungen der Sprachentwicklung. Göttingen: Hogrefe.

Halliday, M.A.K. (1985). *An Introduction to Systemic Functional Grammar*. London: Arnold.

Herberg, J. S., Saylor, M. M., Ratanaswasd, P., Levin, D. T., Wilkes, M. D. (2008). Audience-Contingent Variation in Action Demonstrations for Humans and Computers. *Cognitive Science: A Multidisciplinary Journal*, 32, 1003-1020.

Iverson, J. M., Capirci, O., Longobardi, E. & Caselli, C., M. (1999). Gesturing in mother-child interactions. *Cognitive Development*, 14, 57-75.

Johnson, D., Gardner, J. & Wiles, J. (2004). Experience as a moderator of the media equation: the impact of flattery and praise. *International Journal of Human-Computer Studies*, 61, 3, 237 - 258.

Johnstone, A., Berry, U., Nguyen, T. & Asper, A. (1994). 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.

Ju, W. & Leifer, L. (2008). The Design of Implicit Interactions: Making Interactive Systems Less Obnoxious. *Design Issues*, 24, 3, 72-84.

Ju, W. & Takayama, L. (2008). Approachability: How People Interpret Automatic Door Movement as Gesture. In *Proceedings of Design & Emotion* 2008.

Krach, S., Hegel, F., Wrede, B., Sagerer, G., Binkofski, F. &Kircher, T. (2008), Can Machines Think? Direct Interaction and Perspective Taking with Robots Investigated via fMRI. PLoS ONE, vol. 3, 7, 07/2008.

Kanda, T., Miyashita, T., Osada, T., Haikawa, Y. & H. Ishiguro (2008). Analysis of Humanoid Appearances in Human-Robot Interaction. *IEEE Transactions on Robotics*, 24, 3, 725-735.

Lee, E.-J. (2008). Flattery May Get Computers Somewhere, Sometimes: The Moderating Role of Output Modality, Computer Gender, and User Gender. *International Journal of Human-Computer Studies*, 66, 789-800.

Lee, K. M. & Nass, C. (2003). Designing Social Presence of Social Actors in Human Computer Interaction. *CHI 2003*, April 5-10, 2003, Ft. Lauderdale, Florida 5, 1, 289-296.

Lohan, K., Gieselmann, S., Vollmer, A.-L., Rohlfing, K.J. & Wrede, B. (2009). Does embodiment affect tutoring behavior? In *Proceedings of the IEEE 8th International Conference on Development and Learning*, June (ICDL '09).

Nagai, Y. & Rohlfing, K. J. (2009). Computational analysis of Motionese toward scaffolding robot action learning. *IEEE Transactions on Autonomous Mental Development,* 1, 44-54.

Nass, Clifford (2004). Etiquette Equality: Exhibitions and Expectations of Computer Politeness. *Communications of the ACM*, 47, 4, 35-37.

Nass, C. & Brave, S. (2005). *Wired for Speech: How Voice Activates and Advances the Human-Computer Relationship.* MIT Press, Cambridge, MA.

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

Okita, S.Y., Bailenson, J. & Schwartz, D. L. (2008). Mere belief in social action improves complex

learning. In *Proceedings of the 8th International conference for the learning sciences - Volume 2*, Utrecht, The Netherlands, 132-139.

Onnis, L., Waterfall, H., & Edelman S. 2008. Learn locally, act globally: Learning language with variation set cues. *Cognition*, 109, 423-430.

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.

Pine, J. M. (1994). The Language of Primary Caregivers. In Gallaway, C. & Richards, B. J., (Eds.), *Input and Interaction in Language Acquisition*, 15-37. Cambridge: Cambridge University Press.

Powers, A., Kiesler, S., Fussell, S. & Torrey, C. (2007). Comparing a Computer Agent with a Humanoid Robot. *HRI '07*, 145-152.

Pratt, J. A., Hauser, K., Ugray, Z. & Patterson, O. (2007). Looking at human-computer interface design: Effects of ethnicity in computer agents. *Interacting with Computers*, 19, 4, 512-523.

Reeves, B. & Nass, C. (1996). *The Media Equation. How people treat computers, televisions, and new media like real people and places*. CSLI, Stanford & Cambridge University Press, Cambridge.

Rohlfing, K. J., Fritsch, J., Wrede, B. & Jungmann, T. (2006). How can multimodal cues from child-directed interaction reduce learning complexity in robots? *Advanced Robotics*, 20, 10, 1183-1199.

Sacks, H., Schegloff, E. & Jefferson, G. (1974). A simplest systematics for the organization of turn-taking for conversation. *Language,* 50, 697-735.

Shechtman, N. & Horowitz, L. M. (2006). Interpersonal and Noninterpersonal Interactions, Interpersonal Motives, and the Effect of Frustrated Motives. *Journal of Personality and Social Psychology,* 32, 8, 1126-1139.

Snow, C. E. (1972). Mothers' speech to children learning language. *Child Development,* 43, 549-565.

Snow, C. E. (1977). Mothers' speech research: From input to interaction. In Snow, C.E. & Ferguson, C.A. (Eds.), *Talking to children: Language input and acquisition* (pp. 31-49). London: Cambridge University Press.

Snow, C. E. (1987). Why Routines Are Different: Toward a Multiple-Factors Model of the Relation between Input and Language Acquisition. In: Nelson, K. E. & van Kleck, A. (Eds.), *Children's Language* (pp. 65-97). Hillsdale, N.J.: Erlbaum.

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* (pp. 3-12). Cambridge: Cambridge University Press.

Soderstrom, M. (2007). Beyond Babytalk: Re-Evaluating the Nature and Content of Speech Input to Preverbal Infants. *Developmental Review*, 27, 501-532.

Theakston, A.L., Lieven, E.V.M., Pine, J.M. & Rowland, C.F. (2005). The acquisition of auxiliary syntax: BE and HAVE. Cognitive Linguistics, 16, 281-311.

Thomaz, A.L. & Breazeal, C. (2008). Teachable robots: Understanding human teaching behaviour to build more effective robot. *Artificial Intelligence Journal*, 172, 716-737.

Thomaz, A. L. & Cakmak, M. (2009). Learning about objects with human teachers. *HRI '09*, 15-22.

Vollmer, A.-L., Lohan, K., Fischer, K., Nagai, Y., Pitsch, K., Fritsch, J., Rohlfing, K.J., Wrede, B. (2009). People Modify Their Tutoring Behavior in Robot-Directed Interaction for Action Learning. In *Proceedings of the IEEE 8th International Conference on Development and Learning,* June (ICDL '09).

Wrede, B., Rohlfing, K. J., Hanheide, M. & Sagerer, G. (2009). Towards Learning by Interacting. In Sendhoff, B. et al. (Eds.). *Creating Brain-Like Intelligence* (pp. 139-150). Berlin Heidelberg: Springer-Verlag.

Zukow-Goldring, P. (1996). Sensitive caregiving fosters the comprehension of speech: When

gestures speak lauder than words. In *Early Development and Parenting*, 5, 195-211.