Is Talking to a Simulated Robot like Talking to a Child?

Kerstin Fischer
University of Southern Denmark
IFKI, Alsion2
DK-6400 Sonderborg
kerstin@sitkom.sdu.dk

Kilian Foth
University of Hamburg
Department of Informatics
Vogt-Koelln-Str. 30
D-22527 Hamburg
foth@informatik.uni-hamburg.de

Katharina Rohlfing and Britta Wrede
Bielefeld University
Universitätsstr. 23-25
D-33615 Bielefeld
{kjr, bwrede}@uni-bielefeld.de

***Abstract*—Previous research has found people to transfer behaviors from social interaction among humans to interactions with computers or robots. These findings suggest that people will talk to a robot which looks like a child in a similar way as people talking to a child. However, in a previous study in which we compared speech to a simulated robot with speech to preverbal, 10 months old infants, we did not find the expected similarities. One possibility is that people were targeting an older child than a 10 months old. In the current study, we address the similarities and differences between speech to four different age groups of children and a simulated robot. The results shed light on how people talk to robots in general.**

***Index Terms*—child-directed speech, human-robot interaction, mindless transfer, fine-tuning**

I. Introduction

In this study, we investigate in how far people talk to an infanoid (simulated) robot like to a real infant. Previous proposals suggest that people will make use of their knowledge of speaking to other humans for speaking to robots. In a previous study (Fischer et al. 2011), we could not support this conclusion: participants chose more complex, less interactive and more abstract language when explaining the same tasks to a simulated robot as to young children. However, results from a questionnaire which we handed out at the end of the human-robot interactions suggest that even though the robot was designed to resemble a baby (actually it was called 'babyface' (see Figure 1; Nagai & Rohlfing 2009) because of its similarity with a young child), participants estimated its age much higher (see Figure 2). Thus, the lack of similarities between infant-directed and robot-directed speech may be due to the fact that participants were targeting an older child. In the current study, we address whether the transfer hypothesis is correct after all, which predicts a close connection between speech to humans and to robots, and whether speech to older children serves as a more adequate point of comparison. Therefore, for the current study, we elicited speech to older children, with which we compare the speech to our simulated robot observed.

Figure 1: Instructing 'Babyface'

II. Mindless transfer

Much recent work suggests that people talk to robots and other artificial communication partners like they talk to other people. Reasons for this come from two different lines of argumentation:

First, people may have stored particular ways of speaking in certain situations, which they may make secondary use of when in a similar situation. 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 1986). In other words, if speakers attempt to simplify their speech, for instance, when talking to foreigners, to elderly people or to pet animals, they are expected to draw on the speech forms they employ for children and which they got acquainted with in early childhood themselves.

Second, people may 'transfer mindlessly' from social

Paper submitted to the IEEE ICDL-Epirob 2011 joint conference.

situations to interactions with artificial communication partners. 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 fill 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 out the questionnaire at the same computer. The study reported in Nass (2004) thus shows that the same effect that characterizes human social interaction can be found with computers. In similar studies, Nass and colleagues have investigated a broad range of such social behaviors, with the result that people were found to react similarly to the computer's behavior (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 (Nass & Brave 2005). For instance, Nass and Brave (2005) have shown that people draw on all kinds of clues about artificial communication partners and transfer characteristics from these situations. For example, a synthesized female voice will trigger the attribution of female characteristics to the computer persona. Nass proposes that the reasons 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) find people to present themselves in interviews similarly to human and to computer interviewers. Similarly, Lee et al. (2008) report transfer of cultural background depending on where the robot was reported to have been built.

On the other hand, some studies have found that the effects described by Nass and colleagues can only be observed for selected groups of people or for selected phenomena only; Johnson, Gardner & Wiles (2004) show that the flattery effect described by Fogg & Nass (1997) holds only for some participants; Lee (2008) finds mediating effects of gender and presentation, and Shechtman & Horowitz (2006) report that especially the social aspects of language are lost when people talk to computers instead of to other humans. To sum up, while some research suggests that people will talk to computers and robots like to other people, other research suggests a less deterministic connection.

III. HYPOTHESES

In the current study, we aim to investigate whether our previous failure to show mindless transfer in interactions with the simulated robot 'Babyface' is due to an unsuitable comparison: while 'Babyface' is a preverbal robot designed to look like an infant, participants rated the robot's age higher than that of a child eleven months or younger (see Figure 2). The current study thus addresses whether people talk to the 'Babyface' robot like to an older child. In order to do that, we elicited corpora with children between 12 and 30 months of age using the same tasks as in the infant-directed and robot-directed data sets. If mindless transfer does indeed occur and an older addressee has been targeted in speech to 'Babyface' as suggested by the questionnaire data, speech to older children should become increasingly more similar to speech to the infanoid robot. The hypothesis tested in this study is thus:

H1: Speech to the infanoid robot 'Babyface' is similar to speech to children older than 8-11 months.

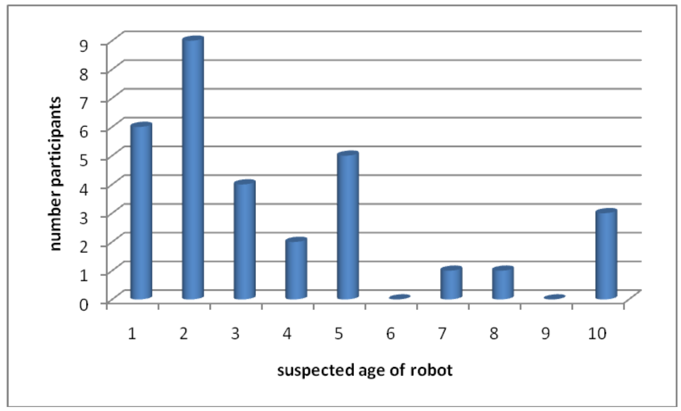

Figure 2: Suspected Age for the 'Babyface' Robot

We investigate this hypothesis by comparing linguistic features of the speech directed at 'Babyface' with the linguistic features of child-directed speech at different age levels by means of univariate analyses of variance.

IV. METHOD AND DATA

A. Participants

Robot: The robot used is a simulated robot; this means that it was displayed on a screen, yet it was interacting with its environment in the form of eye gazing behavior based on the visual saliency of objects in the room. The simulated robot gazes at salient points within objects which are derived from movements, colors, 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. So unbeknownst to the users, the feedback given in form of gaze behavior is based on a purely data-driven reactive saliency-based visual attention model which takes neither prior (e.g. semantic) knowledge nor the participants' linguistic utterances into account.

Human subjects (HRI): Participants were 30 native speakers of German (14 female and 16 male), seven of whom were parents as well. Participants' ages range from 18 to 63 years, and they come from various different fields; however, seven of the 30 participants have a computer science background. Participants were recruited on a word-of-mouth basis and compensated for their participation with a big bar of Swiss chocolate.

Human subjects (CDS): The data sets of child-directed speech (henceforth CDS) consist of parent-child interactions in which parents explain the functioning of certain objects and actions to their children. 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.

CDS I: The first corpus consists of interactions between 28 German speaking parents, i.e. mothers and fathers, and their (pre-verbal) infants, whose age ranged from 8 to 11 months.

CDS II: The second corpus consists of interactions between 30 parents and their 12-17 months old children.

CDS III: The third corpus of child-directed speech consists of interactions between 32 parents and their 18-23 months old children.

CDS IV: The fourth corpus consists of interactions between 36 parents and their 24-30 months old children.

Certainly it would have been desirable to compare the speech to 'Babyface' also with speech to much older children since some of the participants suspected the robot to be as old as 10 years; however, one cannot plausibly make parents teach their 10 year olds how to stack cups or switch on a lamp. Thus, different tasks for the tutoring scenario would have been necessary, rendering the dialogs uncomparable. Furthermore, older children already engage in fluent dialog, whereas 'Babyface' does not produce any linguistic utterances at all. Therefore, we decided to concentrate on speech to children up to 30months of age in this study.

B. *Procedure and Stimuli*

CDS: Parent and child were seated across a table. Then the experimenter put a tray with a set of objects (e.g. four colored 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.

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 a particular age. The same objects as in the experiment with the children were used.

C. *Data Preparation and Encoding*

All five 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.[1]

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 for further analyses 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; since child-directed speech is the prototype of a simplified register, low complexity of utterances is a crucial characteristic of the way of talking to children. Several dependent measures are used 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

[1] Child-directed speech is also characterized by prosodic peculiarities, such as high F0 and high pitch variability (e.g. Snow 1977). Thus, it would be useful to compare child-directed and robot-directed speech also in this respect. In fact, acoustic analyses are in progress, yet results are not available at this point.

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. Previous research has identified interactivity to be a defining characteristic of child-directed speech (e.g. Snow 1977). 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 are however often toned down by means of **modal particles**, sentence medial particles that serve politeness and grounding functions (cf. Fischer 2007). In the current data sets, the down-toned imperative occurs frequently in expressions with 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 five different data sets and divided by the number of utterances used by this person. The numbers underlying the statistical comparison are thus the numbers occurring per number of speakers' utterances.

IV. Results

Table 1 shows the comparison by means of univariate ANOVAs of each of the four CDS corpora with the corpus of robot-directed speech. As can be seen, the numbers of features in which speech to the simulated robot is not significantly different from speech to children of increasing age is very small, and it does not increase with children's increasing age:

Table 1: F-values Describing Differences between Speech to Babyface and the Four Corpora of CDS; t = p<.10; *=p<.05; **=p<.01; ***=p<.001

	CDS I	CDS II	CDS III	CDS IV
diff. words	4,46 *	13,23 ***	7,09 **	7,22 **

utterances	0,51	1,84	3,52 t	2,25
MLU	32,98 ***	14,97 ***	8,97	6,68 *
abstract	16,26***	9,28 **	4,38 *	0,08
concrete	60,52 ***	24,57 ***	18,94 ***	27,51 ***
passive	6,05 *	2,85 t	3,88 t	5,20 *
perfect	0,23	0,13	1,93	0,13
dir-object	45,10***	43,70 ***	67,10 ***	111,49 ***
ind-object	2,28	0,79	0,05	0,71
obj-main clause	77,82 ***	52,54 ***	92,45 ***	139,59 ***
relative clause	8,54**	2,20	5,89 *	2,80 t
embedding	26,76***	12,22 **	19,65 ***	4,18 *
copula	12,60 ***	5,50 *	8,10 **	43,96 ***
subclauses	43,41 ***	18,73 ***	22,56 ***	10,70 ***
declarative	68,21***	20,14 ***	12,52 ***	0,06
imperative	85,40 ***	48,26 ***	86,20 ***	152,20 ***
infinitive	2,21	5,67 *	0,17	6,39 *
modal particles	46,61 ***	28,59 ***	88,80 ***	91,82 ***
questions	39,79 ***	97,10 ***	53,71 ***	105,21 ***
check	43,22***	27,52 ***	40,91 ***	0,004
vocative	27,52 ***	13,12 ***	32,77 ***	7,75 **
'du'	7,90 **	29,83 ***	29,90 ***	49,11 ***
'ich'	14,95 ***	3,26	0,86	5,13 *
'man'	17,23 ***	12,41 ***	10,27 **	8,95 **
'wir'	0,0002	0,90	1,11	1,26

In particular, the results show that with respect to most features the speech directed to the simulated robot differs significantly from the speech directed at children between the age of 8 and 30 months. Exceptions are, for instance, the number of utterances, the use of the perfect tense and the number of uses of 'wir' (*we*). Crucially, however, these numbers are similar to speech to infants as much as to speech to older children, so that age of the addressee does not seem to play a significant role here, either.

Only for the three features abstract language, checking signals, such as tag questions, and declarative sentences, the differences between speech to robot and speech to older children disappear. Some other features do exhibit significant differences at some age levels and non-significant differences at other levels, but the pattern is never systematic. So while the use of 'ich' (*I*) in robot-directed speech is significantly different from infant-directed speech (CDS I), it is more similar to speech for children of 12 to 23 months of age; however, it is again dissimilar to speech directed at children age 24-30 months. A different pattern emerges for relative clauses, whose numbers do not differ significantly for speech to children age 12-17 months, and there is a tendency for speech to 24-30 months olds again to be statistically different from robot-directed speech. Thus, no clear pattern emerges with respect to a progression related to age.

Finally, most of the linguistic features investigated differ significantly from robot-directed speech in all four corpora of child-directed speech alike. Thus, the age of the child does not seem to play a role with respect to these features. The results thus imply that people do not talk to the infanoid robot like they talk to children, irrespective of the children's age.

V. Discussion

The results obtained in this study suggest that people do not transfer mindlessly from speaking to children to speaking to artificial communication partners or make use of patterns of speaking to children for speaking to the simulated robot. Although parents adapt their speech depending on their children's age (Snow 1972, Veneziano 2001, Fischer et al. submitted), the expected age of the respective communication partner had not much impact on how people chose their utterances for the simulated robot. That is, in Fischer et al. (submitted), we have shown for the current data set that parents adapt their speech systematically in accordance with children's age such that there is a linear progression towards increasing complexity. Yet when speaking to the simulated robot, participants do not employ linguistic forms similar to those they employ when talking to children of a particular age range and thus they do not transfer mindlessly from speech to children between 8 and 30 months of age. Even though some features may be similar to the speech to children at some age level, this applies only to selected features, which does not suggest mindless transfer of a way of speaking but rather the selective use of certain features. It is of course possible that speech to the 'Babyface' robot is similar to speech to much older children than pre-verbal infants (i.e. older than 30 months), and that the point of comparison is rather pre-school or even schoolchildren. However, the data considered do not suggest any progression regarding addressee's age; furthermore, the older the comparison group becomes, the less likely is mindless transfer, given the robot's functionality (lack of speech) and infanoid appearance. To conclude, hypothesis H1, that participants talk to the simulated robot like they talk to children (at least between 8 and 30 months of age), needs to be rejected. This leads us to looking for alternative explanations.

In Fischer et al. (2011), the explanation we developed was based on findings with respect to participants' action demonstration towards the robot; while participants talked more abstractly and used more complex linguistic structures, they actually demonstrated actions more slowly and with more pronounced movements for the robot than for the preverbal, 8-11 months old infants. The reason, we suggested, is that the robot reacts contingently to participants' movements, but does not provide any feedback with respect to its linguistic processing of the participants' utterances. The interaction with the simulated robot is therefore shaped by the kind of feedback the robot provides.

Thus, participants make use of the only kind of feedback they get: the robot's eye gaze following the participants' movements. This is in line with findings by Cross et al. (1990) who have shown that in child-directed speech the receptive ability by the infant or child provides crucial information which caretakers take into account when formulating for their young communication partner. In the absence of such cues, people do not know how to adapt and thus may not adjust their speech at all or at least not in systematic ways. Thus, if the robot provides feedback on its receptive abilities regarding action demonstration, people respond to this, yet the robot used in these experiments did not provide any feedback with respect to its linguistic receptive capabilities, which thus did not lead participants to adjust their speech.

VI. Conclusion

This study has shown that people neither draw on their knowledge of speaking to infants of 8-11 months of age, nor on their knowledge of speaking to older children (12-30 months of age) when speaking to a simulated infanoid robot. The findings thus support the conclusion that participants do not transfer (mindlessly) from known situations to interactions with robots and that they do not apply speech to children to robots with infanoid appearance. Instead of transferring mindlessly or employing a particular linguistic variety, we have found people to interact with an infanoid robot in ways that respond to the kind of feedback the robot provides them with. This has considerable design implications, since we can expect people to interact with robots in response to the respective robot’s behavior; a robot's functionality is thus of crucial importance for the interaction. In future work we address concrete measures by means of which participants' behavior towards robots can be shaped implicitly by designing certain behaviors that serve as indicators for particular robot functionalities, especially its linguistic receptive abilities; that is, future work will concern the type of feedback from the robot that will guide users into appropriate models of their artificial communication partner.

Acknowledgment

This research has been funded by the European Union in the framework of the ITALK project under grant number 214668.

References

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

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

[3] T. Cross, J.E. Johnson-Morris and T.G. Nienhuys (1980): Linguistic Feedback and Maternal Speech: Comparisons of Mothers Addressing Hearing and Hearing-impaired Children. *First Language* 1, 3: 163-189.

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

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

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

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

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

[9] K. Fischer, K. Foth, K. Rohlfing and B. Wrede (2011): Mindful Tutors: Linguistic Choice and Action Demonstration in Speech to Infants and a Simulated Robot. *Interaction Studies* 12, 1: 134-161.

[10] K. Fischer, K. Foth, and K. Rohlfing, “Finetuning in Child-Directed Speech”, submitted.

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

[12] Ford, Cecilia E., Barbara A. Fox & Sandra A. Thompson (1996). Practices in the construction of turns: The TCU revisited. *Pragmatics* 6.427-454.

[13] K. Foth, W. Menzel and I. Schröder (2000): A Transformation-based Parsing Technique with Anytime Properties. *4th Int. Workshop on Parsing Technologies*, IWPT-2000, pp. 89 - 100.

[14] Johnson, Daniel, Gardner, John & Wiles, Janet (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.

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

[16] Lee, Eun-Ju (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.

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

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

[19] Nass, Clifford, Brave, Scott, 2005. *Wired for Speech: How Voice Activates and Advances the Human-Computer Relationship*. MIT Press, Cambridge, MA.

[20] Nass, Clifford, Moon, Youngme, 2000. Machines and mindlessness: Social responses to computers. *Journal of Social Issues* 56, 1, 81-103.

[21] Sacks, Harvey, Emanuel Schegloff & Gail Jefferson (1974). A simplest systematics for the organization of turn-taking for conversation. *Language*, 50.697-735.

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

[23] E. Rosch and C.B. Mervis (1975): Family Resemblances: Studies in the Internal Structure of Categories, *Cognitive Psychology* 7: 573–605.

[24] Shechtman, Nicole & Horowitz, Leonard M. (2006): Interpersonal and Noninterpersonal Interactions, Interpersonal Motives, and the Effect of Frustrated Motives. *Journal of Personality and Social Psychology* 32, 8: 1126-1139.

[25] Snow, Catherine E. (1972): Mothers' speech to children learning language. *Child Development* 43, 549-565.

[26] Snow, Catherine 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.

[27] Veneziano, Edy (2001): Displacement and Informativeness in Child-Directed Talk. *First Language* 21: 323-356.