Human Tutors Intuitively Reduce Complexity for Socially Guided Embodied Grammar Learning

Kerstin Fischer

Abstract—**The current investigation addresses whether the socially guided machine learning paradigm can be extended to a new domain, embodied grammar learning. Experimental results show that naive users indeed reduce the complexity of linguistic utterances in tutoring sessions for a simulated robot, even though their own knowledge of the subject area is only tacit. These findings have implications for the usability of robots as 'teachable agents', as well as for automatic language learning from interaction.**

I. Introduction

THIS paper addresses the question to what extent socially guided machine learning can be usefully extended to other domains, such as embodied language learning. In previous studies, Thomaz and Breazeal (2008), Thomaz and Cakmak (2009) and Cakmak et al. (2010) have shown that naive participants provide robots with useful information that facilitates robots' familiarization with objects and shapes. For instance, Thomaz and Cakmak (2009) find participants to provide robots with 1) a useful balance between positive and negative examples, 2) numbers of examples that proportionally match the complexity of the problem, 3) a progression from simple to complex, 4) structuring and chunking, 5) useful information on objects' affordances and main characteristics, and 6) parsing of action goals. Cakmak & Thomaz (2010) find users' intuitive strategies to facilitate speed and quality of learning and demonstrate that these strategies can even be more effective by simple instructions to the human tutors.

The question addressed in this investigation is thus whether the behavior observed by Thomaz, Cakmak, Breazeal, and colleagues can be generalized to other domains, especially to grounded language learning by a robotic learner. In Cakmak et al. (2010), the authors find different social learning strategies to be useful in different situations; it thus needs to be considered what the particular challenges of embodied grammar learning consist in and how these challenges are addressed intuitively by naïve users.

This work was supported by the European Union in the framework of the ITALK project under grant number 214668.

Kerstin Fischer is associate professor at the University of Southern Denmark, IFKI, Alsion 2, DK-6400 Sonderborg. phone: +45-6550-1220; e-mail: kerstin@sitkom.sdu.dk.

II. Embodied Language Learning

The domain investigated is embodied language learning. In this approach, language learning is carried out by embodied systems, i.e. robots, since meaning is taken to be grounded in speakers' own sensorimotor experiences (Barsalou 1999, Cangelosi et al. 2007). The theoretical framework to which embodied grammar learning is anchored is cognitive linguistics, in which language knowledge is taken to consist in constructions, form-meaning pairs (Goldberg 1995, Tomasello 2003, Steels 2004, Dominey 2006). This view of language as an inventory of symbolic structures of varying degrees of schematicity does not distinguish categorically between lexicon and grammar, i.e. between words and the structures in which they occur, which can account for central observations on children's language acquisition processes (Tomasello 2003).

In the construction grammar view, a sentence instantiates, for instance, an abstract argument structure construction, corresponding roughly to who does what to whom. For example, the ditransitive, i.e. a construction with three participants, such as *John gives her the ball*, is taken to have the meaning *an agent transfers an object to a recipient*. It is noteworthy that although the prototypical verb in this construction is *give* (Goldberg 2006), any verb in this construction will assume a transfer reading, as in *He sneezed her a napkin* (Goldberg 1995). Thus, it is the abstract pattern *Noun Phrase – Verb – Noun Phrase – Noun Phrase* that carries the transfer meaning.

In the grounded learning approach taken here, robots learn these constructional meanings as generalizations over perceived scenes (Steels & Loetsch 2009, van Trijp 2008). Language learning input in this approach thus consists of form-scene pairings. In the ITALK project, constructions are being learned using recurrent neural network models (Sugita and Tani 2005, 2008), matching sensorimotor data with linguistic forms (cf. Marocco et al. 2010).

The constructions the robot is supposed to learn here are the caused motion (CM) construction, the ditransitive (DTr) and the passive, which, in German, comes in two different constructions, corresponding to the stative and the event reading of the English passive. The learning task consist in 1) segmenting a holophrase, i.e. a string of sounds, into its components, 2) identifying the components, 3) identifying the constructional meaning, and 4) associating the constituents of the sentence with the semantic roles specified by the construction. Concretely, this means for the four constructions investigated that the robot has to learn an association between the following formal patterns and the corresponding meanings (Goldberg 1995, 2006):

Caused Motion

Noun Phrase	*Verb*	*Noun Phrase*	*Prepositional Phrase*
Agent	cause move	Theme	target location/state
The girl	*rolls*	*the ball*	*into the goal.*

Ditransitive

Noun Phrase	*Verb*	*Noun Phrase*	*Noun Phrase*
Agent	Transfer	Recipient	Theme
The lion	*gives*	*her*	*the ball.*

Stative Passive (*sein*)

Noun Phrase	*Copula*	*Participle*
Undergoer		target state
The paper	*is*	*folded.*

Event Passive (*werden*)

Noun Phrase	*Copula*	*Participle*
Undergoer		process
The paper	*is (being)*	*folded.*

The grammar learning task is thus not trivial. The current paper explores whether a socially guided machine learning paradigm (Thomaz and Breazeal 2008, Thomaz and Cakmak 2009) may be useful for embodied grammar learning as well. This is particularly interesting since people are usually not aware of the associations between form and meaning as they are encoded in the argument structure constructions. Thus, in contrast to previous studies on socially guided machine learning, the tasks investigated here concern merely tacit knowledge. Moreover, the embodied learning paradigm requires the provision of learning input in the form of scene-utterance pairs, and part of the motivation for the current investigation is to see whether naïve users devise such learning input for a robot.

III. Methods and Data

A. Robot

The robot simulation used in these experiments interacts with its environment via eye gaze based on visual saliency. It consists of a face and upper body modeled after a young child. The simulated robot gazes at salient points within objects or persons, based on a purely data-driven, reactive visual attention model, in addition to exhibiting random movements of mouth and eyelids (Nagai and Rohlfing 2009). While the robot is thus not physically embodied, it possesses a central trait of embodiment, namely structural coupling with the environment by means of a contingent input-output relationship (cf. Dautenhahn et al. 2002).

Unbeknownst to the users, the robot does not learn and takes neither semantic knowledge nor the participants' linguistic utterances into account; nevertheless, it reacts contingently to tutors' action demonstrations. Tutoring behaviors observable can therefore be concluded to rest a) on the belief of learning to take place (see also Okita et al. 2007), or b) on the robot's contingent responses by means of eye gaze to the tutor's actions. In a previous study, Fischer et al. (2011) have analyzed interactions with the Babyface robot and found that while people do not speak to the robot like parents speak to their children, they adjust their gestures even more to the robot's eye gaze than parents do for their children. The robot's contingent eye movements thus serve as a signal for continuous attention that participants take into account.

B. Participants

30 native speakers of German participated in this experiment, 14 female and 16 male. Participants' age range is from 18 to 63 years. Participants were recruited on a word-of-mouth basis and received a large bar of chocolate for their efforts.

Figure 1: Instructing Babyface

C. Procedure

Each participant was asked to 'explain the meanings of the following sentences to Akachan'. No further directions or explanations were given. The Japanese translation of the robot's name 'Babyface' was used in order not to guide participants into particular representations of the task. Participants were provided with certain props by means of which they could illustrate the meanings of the target sentences. Thus, one possibility for users is to demonstrate the meaning of the sentences given by a representative scene. What other strategies they took in the task was left to them, yet the term 'explain' in the instruction did suggest verbal explanations. Participants were then handed a card with the sentence to be illustrated and props relevant to the illustration of the sentence's meaning. Example sentences are *the girl is rolling the ball into the goal* (CM), *the lion is giving the frog the ball* (DTr), and *the paper is folded* (stative passive).

D. Materials

Participants received certain props relevant to the sentences provided. These were: crackers, sheets of paper, a ball, a yoyo, a dollhouse chair, a dollhouse table, a toy car and a toy soccer goal (for the ball). In addition, participants received a large puppet girl and two smaller puppets, a frog and a lion. These props did not only function as props for the illustration of sentence meanings, but participants also concentrated on the acting task, which many of them seemed to enjoy very much, thus diverting their attention from the explanation/speaking task.

The sentences chosen instantiate different constructions, and the verbs in these sentences were selected to be compatible in German with most constructions investigated. The verbs chosen were *roll, topple, break, slide, fold*, and *give*. The constructions selected are the five major argument structure constructions (Goldberg 1995, 2006), plus the two passives in German, passive with *sein* and passive with *werden*. These two passive constructions express stative and event readings respectively; these two readings are both covered by the English passive. In contrast to the five major argument structure constructions, both passives are relatively rare, especially in input to language learners (Abbot-Smith & Behrens 2006); Cakmak et al. (2010) suggest that socially guided machine learning may be particularly useful for rare cases, and thus participants' tutoring behavior for these constructions is particularly interesting. All constructions exhibit relatively stable semantics (Goldberg 1995, 2006) and are sufficiently complex. In the data analysis, we focus here on the most complex of these constructions, the caused motion (or resultative) construction (CM, Goldberg and Jackendoff 2006), the ditransitive (DTr, Goldberg 1995) and the two passive constructions (Abbot-Smith and Behrens 2006, Langacker 2008). Not presented here are participants' tutoring behaviors for the intransitive (ITr, e.g. *the ball is rolling*), the transitive (Tr, *the frog is rolling the ball*), and the intransitive resultative (ITrR, *the ball is rolling into the goal*). However, these constructions do play a role in the participants' explanations of the more complex constructions.

E. Data Encoding

The data were analyzed concerning the contents of participants' utterances. Each utterance was encoded for the kind of contents expressed. In particular, the following categories were distinguished:

role	**description**	**example**
agent	participant doing something	*this is the girl*
possession	description of possession	*the girl has a dress*
state	description of start or end state	*and now: broken*
theme/ undergoer	participant undergoing something	*this is the ball*
mental	description of an intention serving as explanation	*the girl wants the ball*
goal	description of the goal	*the ball has to go into the goal*
result	description of target state	*and now the frog has the ball*
action	description of action: - intransitive - transitive - in-/transitive resultative	*the ball is rolling*

Note that only if the whole utterance concerns the description of the semantic role filling entity, the utterance was encoded as expressing the semantic role in question.

IV. Results

All participants used demonstration of a relevant scene as a strategy of tutoring the robot on the relevant sentence meaning. This result is expected since participants were provided with the props for such demonstrations and thus most likely felt obliged to use them.

In addition, participants exhibit considerable tutoring behavior; in many tasks, participants decompose scenes into

sub-actions, represented by more elementary actions, such as balls rolling or chairs toppling, which are expressed by means of less complex argument structure constructions, such as the intransitive or the transitive construction. Furthermore, participants initially introduce actors, themes/ undergoers, goals, etc. figuring in the scene presented. Consider, for instance, the following example instruction for the caused motion construction:

vp009_cm_s3168: das Pendel, [*the yoyo* - THEME]
vp009_cm_s3169: das Auto [*the car* - GOAL]
vp009_cm_s3170: das Mädchen [*the girl* - AGENT]
vp009_cm_s3171: das Mädchen schwingt das Pendel [*the girl is swaying the yoyo* -TR]
vp009_cm_s3172: das Mädchen schwingt das Pendel gegen das Auto. [*the girl is swaying the yoyo against the car* - CM]

Figure 2 shows the percentage of participants elaborating on particular semantic roles in their instructions for the caused motion construction:

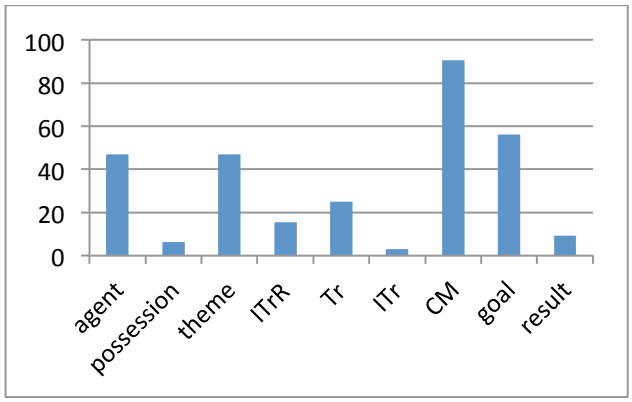

Figure 2: Caused Motion Construction Analyses: Occurrences of semantic roles in percent

Regarding the caused motion construction, such as *the girl is rolling the ball into the goal* (this scene is displayed in Figure 1), participants exhibit considerable tutoring behavior: 46.9% introduce the agent and the theme separately before illustrating the whole scene. In addition, 56.3% of the participants devote a whole utterance to describing the goal. 9.4% mention the result state, for instance, the ball being in the goal, and 43.7% use either intransitive, transitive or intransitive resultative constructions to describe the action in easier terms.

Similarly, the ditransitive construction is also introduced using preliminary introductions of the agent, in this case even 76.4%, the theme (52.9%), and the recipient (17.6%). An example instruction for the ditransitive construction is the following:

vp001_dt_s2639: Akachan, dies ist ein Löwe. [*Akachan, this is a lion*. - AGENT]
vp001_dt_s2640: das ist ein Frosch. [*this is a frog*. - GOAL]
vp001_dt_s2641: und der Löwe hat einen Ball. [*and the lion has a ball*. - POSSESSION]
vp001_dt_s2642: der Löwe gibt dem Frosch den Ball. [*the lion is giving the frog the ball*. - DTR]
vp001_dt_s2643: bitteschön. [*here you are*.]

Furthermore, 52.9% mention the fact that the theme is initially in the agent's possession, and further 35.3% describe the result state, the theme being now in the possession of the goal:

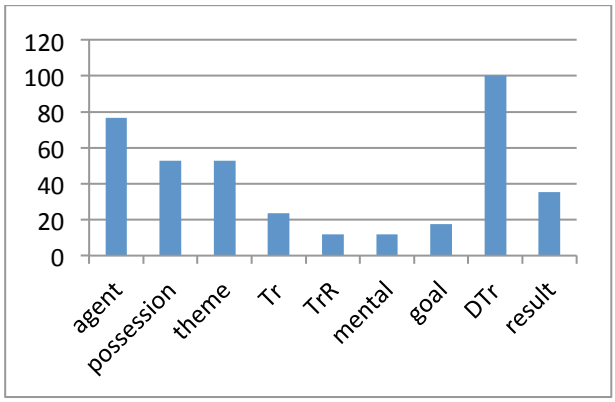

Figure 3: Ditransitive Construction Analyses: Occurrences of semantic roles in percent

Similar results can be observed for the two passive constructions; most striking, however, is the fact that the two constructions, the stative passive (with *sein*) and the event passive (with *werden*), are presented in considerably different ways that reflect the underlying constructional semantics. Thus, we can observe much more attention being paid to the action (54.7%) in the event reading than in the stative reading (29.2%). In contrast, in the stative passive sentences, start (state 1) and goal (state 2) states are highlighted (37.5% and 29.7% respectively). These presentations of the two passive constructions are quite remarkable since they correspond to the findings of highly sophisticated linguistic analyses (e.g. Langacker 2008). Thus, in this experiment, participants activated considerable tacit metacognitive knowledge in order to facilitate the learning task for the robot.

In addition, in both cases, the agent is hardly ever mentioned (2.3% and 0% respectively), which corresponds

to the work a passive construction does: it backgrounds the agent and foregrounds the undergoer of an action, which is mentioned by 66.6% and 75% of the participants respectively. An example instruction for the event passive, which actually makes of the stative passive to describe the end result, is the following:

vp029_p_s1905: das Papier wird geknickt.
[*the paper is being folded*. - WP]
vp029_p_s1906: ein Stück Papier. [*a piece of paper*. – UNDERGOER]
vp029_p_s1907: jetzt legen die eine Seite auf die andere darauf. [*now put one side onto the other* - ACTION]
vp029_p_s1908: und drücke hier zusammen und damit ist das Papier geknickt. [*and press here together and thus the paper is folded*. – ACTION – STATE2]

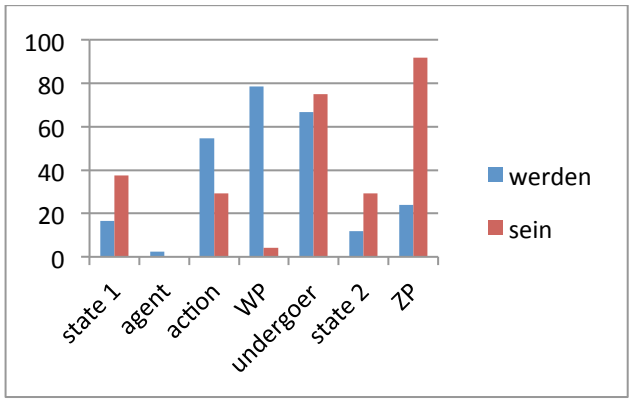

Figure 4: Passive Constructions Analyses: Occurrences of semantic roles in percent

An example for the stative passive is the following:

vp008_zp_s273: also hier haben wir wieder unsere schöne Salzstange. [*so here we have again our nice cracker*. - UNDERGOER]
vp008_zp_s274: und [*and*]
vp008_zp_s275: wir brechen sie in der Mitte durch.
[*we break it in the middle*. - ACTION]
vp008_zp_s276: und damit ist die Salzstange zerbrochen.
[*and thus the cracker is broken*. – STATE2]

V. Discussion

The analysis of participants' tutoring strategies has shown that naïve users decompose complex constructional meanings for their robotic interaction partner. Subjects have been found to exhibit extensive tutoring behavior only on the basis of the belief that the robot will learn from the interaction and on the basis of the robot's contingent non-verbal responses. Subjects provide a systematic decomposition of complex scenes as they are represented in argument structure constructions into subevents, participants, states, and basic relationships. Crucially, these intuitive analyses correspond to the analyses cognitive linguists have provided for these constructions. This is all the more surprising since this kind of linguistic knowledge is usually tacit and not available to direct introspection. Besides facilitating learning for the robot, the results thus indicate that complex metacognitive activities were involved on the side of the users.

VI. Conclusions and Future Work

The results of the current investigation show that naïve participants decompose intuitively the complex constructional meanings of linguistic constructions for a robotic learner. The results therefore support and extend previous findings in the socially guided machine learning paradigm (Thomaz and Breazeal 2008, Thomaz and Cakmak 2009) by applying it to domains in which participants do not have explicit expert knowledge.

A possible limitation of the current findings is that the robot used for data elicitation was not physically embodied, while the robots learning language from interaction with human tutors and their environment will necessarily be physically embodied in order to create their own sensorimotor experiences corresponding to the linguistic structures they hear. In Fischer et al. (2012), we have addressed the impact of physical embodiment directly; we found that in general, physical embodiment makes a robot a more credible learner, and participants produce more tutoring behavior for the physically embodied than for the simulated robot. We can thus expect to find even more decomposing strategies in interaction with the embodied target robot, yet these hypotheses still need to be tested empirically.

Future work will also have to show how participants' behavior changes when the robot really learns from interaction; in Fischer & Saunders (submitted), we have seen that participants adjust very sensitively to the robot's increasing capabilities. So we have reason to speculate that participants adjust their decomposing behaviors also for the language learning robot.

VII. Design Implications

The fact that naïve users decompose complex constructional meanings for their robotic interaction partner is encouraging with respect to the goal to bootstrap language

in a robot from interactions with humans, attempted in the ITALK project (Cangelosi et al. 2010), similar to the language learning process of children (Tomasello 2003). At the same time, the data collected here serve as the starting point for the generation of input for perceptually grounded language learning experiments.

Besides facilitating learning for the robot, the results show that complex metacognitive activities are involved on the side of the users. These observations suggest that interactions with robots may evoke metacognitive strategies in users, such that speakers are required to re-structure and explicate their tacit knowledge for their artificial communication partner in the tutoring situation. Biwas et al. (2005) have shown that learning from metacognition involved in the tutoring of 'teachable agents' facilitates learning (cf. also Kinnebrew et al. 2011). The current findings on participants' metacognition regarding tacit linguistic knowledge opens up the possibility that robots may very well serve as 'teachable agents' also in language learning contexts. The fact that the mere belief that the robot will learn from the interaction, which is consistent with findings by Okita et al. (2007), initiates extensive tutoring behavior in the participants suggests that we do not need to wait for the perfect teachable agent before robots can be used as pedagogical tools. Especially in second language learning contexts, such metacognitive activations may prove useful.

ACKNOWLEDGMENT

I am much indebted to my ITALK partners at Bielefeld University, Katharina Rohlfing, Britta Wrede and Katrin Lohan, for their help with the data elicitation.

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, pp. 995-1026.

[2] L.W. Barsalou, " Perceptual symbol systems." *Behavioral and Brain Sciences*, 22,4, 577-660, 1999.

[3] G. Biswas, K. Leelawong, D. Schwartz and N. Vye, "Learning by teaching: A new agent paradigm for educational software". *Applied Artificial Intelligence*, 19, 363-392.

[4] M. Cakmak, N. DePalma, A.L. Thomaz and R. Arriaga, "Effects of Social Explortion Mechanisms on Robot Learning", in the 2009 *Proceedings of the 18th IEEE International Symposium on Robot and Human Interactive Communication* (Ro-MAN09).

[5] M. Cakmak and A. Thomaz, "Optimality of Human Teachers for Robot Learners", in the 2010 *Proceedings of the 9th IEEE International Conference on Development and Learning,* pp. 64-69.

[6] A. Cangelosi, V. Tikhanoff, J. Fontanari and E. Hourdakis, "Integrating language and cognition: A cognitive robotics approach." IEEE *Computational Intelligence Magazine* 2.3.65-70, 2007.

[7] A.Cangelosi, G. Metta, G. Sagerer, S. Nolfi, C.L. Nehaniv, K. Fischer, J. Tani, T. Belpaeme, G. Sandini, L. Fadiga, B. Wrede, K. Rohlfing, E. Tuci, K. Dautenhahn, J. Saunders and A. Zeschel (2010): Integration of action and language knowledge: A roadmap for developmental robotics. *IEEE Transactions on Autonomous Mental Development*, 2(3):167-195.

[8] K. Dautenhahn, B. Ogden & T. Quick, "From Embodied to Socially Embedded Agents – Implications for Interaction-aware Robots." *Cognitive Systems Research* 3: 397-428, 2002.

[9] P.F. Dominey "From holophrases to abstract grammatical constructions: Insights from simulation studies". In: E. Clark and B. Kelly (Eds.), *Constructions in Acquisition* (137-162). Stanford: CSLI Publications, 2005.

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

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

[12] K. Fischer and J. Saunders. "Between initial expectations and acquaintance: Interacting with a developing robot", submitted.

[13] A.E. Goldberg, *Constructions: A construction grammar approach to argument structure.* Chicago: University of Chicago Press, 1995.

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

[15] A.E. Goldberg and R. Jackendoff, "The end result(ative)." *Language* 81 2, 2006, 474-477.

[16] J. Kinnebrew, G. Biswas, B. Sulcer and R. Taylor, "Investigating Self-Regulated Learning in Teachable Agent Environments". In R. Azevedo & V. Aleven (Eds.), *International Handbook of Metacognition and Learning Technologies*. Berlin, Germany: Springer, 2011.

[17] R.W. Langacker, *Cognitive Grammar: A Basic Introduction*. New York: Oxford University Press, 2008.

[18] D. Marocco, A. Cangelosi, K. Fischer and T. Belpaeme, "Grounding action words in the sensory-motor interaction with the world." *Frontiers in Neurorobotics 4, 7: 1-15,* 2010.

[19] Y. Nagai and K. Rohlfing, "Computational analysis of Motionese toward scaffolding robot action learning." IEEE Transactions on Autonomous Mental Development 1, 2009, 44-54.

[20] S. Okita, J. Bailenson and D.L. Schwartz, "The Mere Belief of Social Interaction Improves Learning." *Cognitive Science Conference*, 2007.

[21] L. Steels, "Constructivist development of grounded construction grammars." In D. Scott, W. Daelemans and M. Walker (Eds.), in 2004 *Proceedings of the ACL* (9-16). Barcelona: ACL.

[22] L. Steels and M. Loetzsch, "Perspective alignment in spatial language." In K. R. Coventry, T. Tenbrink, and J. A. Bateman (eds.), *Spatial Language and Dialogue*, pp. 70-89. Oxford University Press, 2009.

[23] Y. Sugita and J. Tani, "Learning semantic combinatoriality from the interaction between linguistic and behavioral processes." *Adaptive Behavior*, 13.3.2005, 211-225.

[24] Y. Sugita and J. Tani, "A sub-symbolic process underlying the usage-based acquisition of a compositional representation: Results of robotic learning experiments of goal-directed actions." *Proceedings of ICDL2008,* 127-132.

[25] A.L. Thomaz and C. Breazeal, "Teachable robots: Understanding human teaching behaviour to build more effective robot". *Artificial Intelligence* 2008.

[26] A.L.Thomaz and M. Cakmak, Learning about objects with human teachers. *HRI* 2009, pp. 15-22.

[27] M. Tomasello, *Constructing a Language*. Cambridge, Mass.: Harvard University Press, 2003.

[28] R. van Trijp, "The Emergence of Semantic Roles in Fluid Construction Grammar." In A. D. M. Smith, K. Smith and R. Ferrer-i-Cancho (eds.), *Proceedings of the 7th International Conference on the Evolution of Language* 2008, pp. 346-353.