Guest Editorial
Special Issue on Multidisciplinary Perspectives on Mechanisms of Language Learning

I. Scope of This Special Issue

HUMANS excel at learning from other humans [item 1) in the Appendix). Language facilitates such learning and plays a crucial role. On the one hand, it coordinates our interactions and cooperative behavior [item 2) in the Appendix). On the other hand, language and communication allow to directly incorporate novel knowledge gathered from social interaction or from reading [item 3) in the Appendix). It has been a long-standing goal of artificial intelligence to leverage such communicative abilities [item 4) in the Appendix) for robots and smart software agents which would, at first, simplify our interactions with machines through a more human-like way of coordinating between humans and robots. But furthermore, this would allow us to easily teach these machines, increasing their abilities and skills further which would allow them to become real partners and companions.

As we turn towards communication in humans for inspiration [item 5) in the Appendix), we should consider as well how these capabilities develop [item 6) in the Appendix) and are learned [item 7) in the Appendix). Children acquire language by interacting with their caregivers and others in their social environment. When children start to talk, their sensory-motor intelligence (visual perception, body movement, navigation, object manipulation, auditory perception, and articulatory control) is already reaching a high level of competence [items 8) and 9) in the Appendix). Importantly, communication is based on representations and skills that have started to develop much earlier and that are shaped already in first (social) interactions [items 10) and 11) in the Appendix). These interactions are multimodal in nature and vary across contexts. The contexts vary not only across developmental time and situations within individuals but also between individuals, socioeconomic groups, and cultures. Continuously, representations become further enriched in ongoing interactions and across different contexts [item 12) in the Appendix). A developmental perspective allows putting a focus on different mechanisms or influences that contribute differently over time and analyze these as well as their interactions more closely.

There is now an increasing number of approaches in developmental robotics [item 7) in the Appendix) that start to model communication. Such approaches have already lead to key insights. For example, the continuous acquisition of knowledge in different contexts and being able to further enrich the underlying representations provides a potentially powerful mechanism (cross-situational learning) that is already well recognized in learning in children. Still, we need to know more about how children recognize contexts and how their language learning benefits from different language use varying across contexts as the emergence of symbolic communication is still an open problem. We are still lacking unifying theories [item 13) in the Appendix) and implementations that show how cooperation and interaction skills could emerge in long-term experiments with populations of robotic agents or how these skills develop in children.

To this end, the special issue surveys the state of the art of the emergence of communication in humans and applied on artificial systems. The contributions include, first, developmental studies that range from experiments with a focus on specific contributions of mechanisms over longitudinal studies on language learning in children to cross-linguistic comparisons. Second, it consists of language learning models and representations for language learning. Importantly, in many of the presented research, both perspectives are addressed as clearly defined phenomena are identified, often further investigated through experimentation and addressed in a modeling approach that helps to understand hypothetical underlying mechanisms and allows driving further analysis and research questions. Overall, the special issue brings together quite diverse disciplines, as are developmental psychology, robotics, artificial language evolution, complex systems science, computational linguistics, and machine learning.

II. Contributions to the Special Issue

This special issue includes 13 articles. These articles contribute toward understanding language learning from different perspectives, including experimental and developmental studies, as well as modeling approaches. In particular, there are two complementary and overlapping overarching perspectives. On the one hand, there is a specific focus on the act of communication and interaction, as well as underlying mechanisms [item 14) in the Appendix). On the other hand, a focus is on representation and the organization of the underlying conceptual systems [item 15) in the Appendix).

In "An epigenetic approach to semantic categories," Gärdenfors provides an overview of the organization of long-term memory and in particular the structure of conceptual spaces. He characterizes five primary knowledge structures as fundamental dimensions of conceptual spaces: 1) space;

Digital Object Identifier 10.1109/TCDS.2020.2991470

2) objects; 3) actions; 4) numbers; and 5) events. This article reviews psychological work highlighting the emergence of these aspects in child development. This can provide a foundation for conceptual systems in robots, and it is used to argue for such an epigenetic model of language acquisition that employs these dimensions.

Multiple studies address how language influences the learning of representations. Beckage *et al.* in "Quantifying the role of vocabulary knowledge in predicting future word learning" present a neural network-based approach that models learning of vocabulary. While past research on children's language learning was mainly based on child information—such as the number of words a child knows—they focus on the influence of detailed vocabulary knowledge for future language learning. A comparison of the results of different feature models demonstrated that both the vocabulary knowledge and the way the words are represented affect the prediction of future language learning. Such a model was able to successfully predict longitudinal data on vocabulary learning in children (aged 15–36 months).

Importantly, representational systems are multimodal and a conceptual system must be able to bind together information from different modalities. Multiple contributions address this aspect and in particular, how language supports learning such bindings and how language is deeply integrated into such representational structures. As a first example, the article "Neurocomputational models capture the effect of learned labels on infants' object and category representations" by Capelier-Mourguy *et al.* addresses the question of how labels are integrated into representational structures. Starting from the experimental finding that children look longer toward recently named objects compared to nonlabeled objects, the authors present an autoencoder neural network model that differentiates between two possible theoretical accounts. As a result, their simulation data support an account in which labels are considered as features on the same level as haptic or visual information.

Xie *et al.* in "Integrating image-based and knowledge-based representation learning" provide supporting brain evidence that highlights the emergence of representations used in language from regularities and co-occurrence of features among different modalities. Representations are assumed as structures that bind together information from different sources and become accessible through language. They propose an image-based knowledge representation learning model that combines features extracted through a deep convolutional neural network and structured information representations.

A recurrent neural network model is presented by Hinaut and Twiefel in "Teach your robot your language! Trainable neural parser for modeling human sentence processing: Examples for 15 languages" that learns to map an input sentence toward a thematic role represented as a meaning structure. While the model had already been applied successfully before, here the authors show that the underlying learned representation transfers to different languages. The model is realized as an echo state network and is applied in 15 different languages which demonstrates that the underlying structure of the core of the network is not reflecting the particular sentence structure of an individual language, but could be assumed to capture the emerging conceptual structure. The authors propose to apply such a model in a real robotic architecture.

In "The subject–object asymmetry revisited: Experimental and computational approaches to the role of information structure in children's argument omissions," Graf *et al.* challenge the subject–object hypothesis which assumes that in childrens' language use subjects are more often omitted than objects. First, they present a study that did not show this effect for German children using constructions that are matched for discourse pragmatics but differ with respect to the ordering of arguments (subject–verb–object and in contrast object–verb–subject). In their findings, ordering of arguments appears as the decisive factor as initial arguments—independent of being subject or object—are significantly more often omitted compared to final arguments. Second, they propose a computational model with recency bias and show that this can account for the described effect.

As learning of representation occurs mostly in social contexts, there are further studies that broaden their scope toward such a more social perspective. As a first example, Acevedo-Valle *et al.* focus on a lower level of learning of vocalization in their article titled "Social reinforcement in artificial prelinguistic development: A study using intrinsically motivated exploration architectures." In their sensorimotor model, the authors combine two influences that drive learning. On the one hand, intrinsic motivation autonomously drives explorative behavior. On the other hand, the authors found that the emergence of complex structures required social reinforcement in their simulation. They argue that this stresses the importance of social reinforcement signals in interactions between mothers and children for prelinguistic development.

In "Beyond the self: Using grounded affordances to interpret and describe others' actions" by Saponaro *et al.*, a learning architecture is proposed that, initially, learns object affordances as associations of object properties, actions, and the resulting effects through explorative manipulation of objects. This representation is afterward connected to language through the learning of labels. In the next step, a mapping from own action representations toward similar actions performed by a human is acquired. Knowledge of details of the own actions is used for prediction and to infer the current task. The integration of the different representations and binding of complementary information is seen as a prerequisite for social collaboration between humans and robots.

In "When object color is a red herring: Extraneous perceptual information hinders word learning via referent selection," Horst *et al.* consider how bindings of such a representation generalize depending on the context of the communicative situation. First, an experiment is presented in which two and a half year old children learned novel labels for objects. These objects were presented together with distractor objects that either were all of the same color or differed in color from the target object. During testing, children showed problems in the generalization of novel concepts when learning that these simply could be distinguished based on color. In contrast, when learning that all objects were of the same color, children showed better generalization. It appears that other forms of representation and features were learned additionally and bound together during the learning phase. The experiment was reproduced in an epigenetic multimodal architecture that models binding different information sources, such as color and shape, together. Overall, the experiment shows how less information can guide learning and in this way allow for better generalization.

Learning across situations in adults was analyzed using experimental data on adults' gaze behavior. In "Adults use cross-situational statistics for word learning in a conservative way" by Aussems and Vogt, experimental data are captured on gaze behavior in a cross-situational word learning task in which the correct word-referent mapping is to be established from a series of ambiguous scenes. The study compares different word-learning strategies and finds that participants mostly rely on a conservative cross-situational learning strategy for consecutive learning, rather than on a strategy that is more driven by a guess-and-test approach. The authors argue that many of the guess-and-test accounts for word learning are consistent with experiments where participants are forced to select a potential meaning of a novel word. However, in the real world, learners are not forced to choose, but instead can be more conservative and wait until they have seen enough instances to actually disambiguate the different scenes.

Bortfeld and Oghalai turn toward the importance of joint attention in a learning setting in "Joint attention in hearing parent–deaf child and hearing parent–hearing child dyads." In particular, they contrast how joint attention is established differently when there is no shared formal communication system between interaction partners. This is the case in deaf children who go through early cochlear implant surgery and the parents do not know a visual language. In their experimental study, the authors compares hearing parent–hearing child dyads with hearing parent–deaf child dyads. As a group difference, they find that during the free play sessions, hearing parent–hearing child dyads engage in successful interactions in which parents initiate joint attention to a significantly greater proportion of time than in hearing parent–deaf child dyads. Interestingly, regardless of the child's hearing ability, hearing parents tend to use the auditory modality to engage their children, which behavior can serve to explain the larger proportion of unsuccessful parent-initiated joint attention in hearing parent–deaf child dyads. Overall, the authors emphasize the child's role in establishing joint attention and different modalities that need to be investigated more.

Besides joint attention, in larger groups, it becomes important to establish who is currently speaking. Stefanov *et al.* propose a self-supervised machine learning model that can detect the active speaker in "Self-supervised vision-based detection of the active speaker as support for socially aware language acquisition." The architecture integrates multimodal input from multispeaker interaction scenarios and binds together auditory and visual input. The neural network model is trained in an unsupervised fashion and reaches a high accuracy. Such a model can be seen as a prerequisite for artificial systems that should communicate in groups of persons.

The final contribution addresses interaction on a level of turn-taking. In "Multimodal turn-taking: Motivations, methodological challenges, and novel approaches," Rohlfing *et al.* deal with interaction partners' exchange. They emphasize the sequential nature of interaction and that this should be reflected in studies as well as models of turn taking. Through their analysis of corpus data in early interactions, they show that the structure of this process emerges from sources of different modalities, and the partners respond to the multimodal patterns of interaction. As a conclusion, they further highlight that interaction requires such a dynamic and multimodal perspective.

III. Conclusion

Language—as a symbolic system—allows us to conceptualize reality [item 16) in the Appendix) and transmit these conceptualizations to others. But in order to serve communication, such conceptualization has to be shared between agents. This requires that, on the one hand, language is a social mechanism. It emerges as a population of agents negotiate and coordinate which symbols to use in order to talk about the world. On the other hand, language as a conceptualization of the world has to be grounded [item 17) in the Appendix) and embodied [item 16) in the Appendix): it has to be strongly connected and integrated with multimodal perceptions as well as actions [items 18) and 19) in the Appendix). Importantly, learning of language is, therefore, deeply intertwined with the development of representations and embedded as well as guided in social interactions. This special issue contributes here to a broad perspective on this dynamic learning process and puts a focus on different mechanisms that highlight the social nature of language learning or the development of embodied conceptual representation.

Malte Schilling, *Guest Editor*
Neuroinformatics Group
Faculty of Technology
Bielefeld University
D-33501 Bielefeld, Germany

Sony Computer Science Laboratories Inc.
Tokyo 141-0022, Japan
E-mail: mschilli@techfak.uni-bielefeld.de

Katharina J. Rohlfing, *Guest Editor*
Department of German Studies
and Comparative Literature Studies
Paderborn University
D-33098 Paderborn, Germany
E-mail: katharina.rohlfing@uni-paderborn.de

Paul Vogt, *Guest Editor*
Department of Cognitive Science
and Artificial Intelligence
Tilburg University
5037 AB Tilburg, The Netherlands
E-mail: p.a.vogt@uvt.nl

Chen Yu, *Guest Editor*
Department of Psychological
and Brain Sciences
Indiana University
Bloomington, IN 47405 USA
E-mail: chenyu@indiana.edu

MICHAEL SPRANGER, *Guest Editor*
Sony Computer Science Laboratories Inc.,
Tokyo 141-0022, Japan
E-mail: michael.spranger@gmail.com

APPENDIX
RELATED WORK

1) M. Tomasello, *The Origins of Human Communication*. Cambridge, MA, USA: MIT Press, 2008.
2) S. Nolfi, "Emergence of communication in embodied agents: Co-adapting communicative and noncommunicative behaviours," *Connect. Sci.*, vol. 17, nos. 3–4, pp. 231–248, 2005.
3) A. Zadbood, J. Chen, Y. C. Leong, K. A. Norman, and U. Hasson, "How we transmit memories to other brains: Constructing shared neural representations via Communication," *Cerebral Cortex*, vol. 27, no. 10, pp. 4988–5000, 2017.
4) S. Harnad, "Symbol grounding and the origin of language," in *Computationalism: New Directions*, M. Scheutz, Eds., Cambridge, MA, USA: MIT Press, 2002, pp. 143–158.
5) B. M. Lake, T. D. Ullman, J. B. Tenenbaum, and S. J. Gershman, "Building machines that learn and think like people" *Behav, Brain Sci.*, vol. 40, pp. 1–101, Nov. 2016.
6) L. Steels and E. Szathmáry, "The evolutionary dynamics of language," *Biosystems*, vol. 164, pp. 128–137, Feb. 2018. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0303264717302885
7) A. Cangelosi and M. Schlesinger, *Developmental Robotics: From Babies to Robots*. Cambridge, MA, USA: MIT Press, 2015.
8) E. Thelen, "Motor aspects of emergent speech: A dynamic approach," in *Biological and Behavioral Determinants of Language Development*. Hillsdale, NJ, USA: Erlbaum, 1991, pp. 339–362.
9) J. M. Iverson, "Developing language in a developing body: The relationship between motor development and language development," *J. Child Lang.*, vol. 37, no. 2, pp. 229–261, 2010.
10) M. Spranger and L. Steels, "Discovering communication through ontogenetic ritualisation," in *Proc. Joint IEEE Int. Conf. Develop. Learn. Epigenetic Robot. (ICDL-Epirob)*, 2014, pp. 14–19.
11) I. Nomikou, M. Schilling, V. Heller, and K. J. Rohlfing, "Language-at all times: Action and interaction as contexts for enriching representations," *Interact. Stud.*, vol. 17, no. 1, pp. 120–145, 2016. [Online]. Available: https://www.jbe-platform.com/content/journals/10.1075/is.17.1.06nom
12) V. Gallese and G. Lakoff, "The Brain's concepts: The role of the sensory-motor system in conceptual knowledge," *Cogn. Neuropsychol.*, vol. 22, nos. 3–4, pp. 455–479, Jun. 2005.
13) K. J. Rohlfing, B. Wrede, A.-L. Vollmer, and P.-Y. Oudeyer, "An alternative to mapping a word onto a concept in language acquisition: Pragmatic frames," *Front. Psychol.*, vol. 7, p. 470, 2016. [Online]. Available: https://www.frontiersin.org/article/10.3389/fpsyg.2016.00470
14) G. Pezzulo, L. W. Barsalou, A. Cangelosi, M. H. Fischer, K. McRae, and M. J. Spivey, "Computational grounded cognition: a new alliance between grounded cognition and computational modeling," *Front. Psychol.*, vol. 3, p. 612, Jan. 2012.
15) M. Kiefer and F. Pulvermüller, "Conceptual representations in mind and brain: Theoretical developments, current evidence and future directions," *Cortex*, vol. 48, no. 7, pp. 805–825, 2012. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0010945211001018
16) L. W. Barsalou, M. Spivey, K. McRae, and M. Joanisse, "The human conceptual system," *The Cambridge handbook of psycholinguistics*. New York, NY, USA: Cambridge Univ. Press, 2012, pp. 239–258.
17) L. Barsalou, "Grounded cognition," *Ann. Rev. Psychol.*, vol. 59, pp. 617–645, Jan. 2008.
18) J. R. Binder and R. H. Desai, "The neurobiology of semantic memory," *Trends Cogn. Sci.*, vol. 15, no. 11, pp. 527–536, 2011.
19) R. H. Desai, M. Reilly, and W. van Dam, "The multifaceted abstract brain," *Philosoph. Trans. Roy. Soc. B Biol. Sci.*, vol. 373, no. 1752, 2018, Art. no. 20170122. [Online]. Available: https://royalsocietypublishing.org/doi/abs/10.1098/rstb.2017.0122

Malte Schilling received the Diploma degree in computer science from Bielefeld University, Bielefeld, Germany, in 2003, with his thesis on knowledge-based systems for virtual environments, and the Ph.D. degree in biology from Bielefeld University in 2010, working on decentralized biologically inspired minimal cognitive systems.

He is a Responsible Investigator with the Center for Cognitive Interaction Technology, Bielefeld. He was a Postdoctoral Fellow with ICSI, Berkeley, CA, USA, and did research on the connection of linguistic to sensorimotor representation. His work concentrates on learning of adaptive behavior, internal models, their grounding in behavior and application in higher-level cognitive function like planning ahead or communication.

Katharina J. Rohlfing received the Ph.D. degree in linguistics from Bielefeld University, Bielefeld, Germany, in 2002, within an interdisciplinary graduate program task-oriented communication.

Since 2015, she has been a Full Professor of psycholinguistics and the Head of SprachSpielLabor, Paderborn University, Paderborn, Germany. Her research interest is children's acquisition of language, and how it can be scaffolded in social interaction.

Paul Vogt received the M.Sc. degree in cognitive science and engineering from Rijksuniversiteit Groningen, Groningen, The Netherlands, in 1997, and the Ph.D. degree from the Artificial Intelligence Laboratory, Vrije Universiteit Brussel, Brussels, Belgium, in 2000.

He is an Associate Professor with the Department of Cognitive Science and Artificial Intelligence, Tilburg University, Tilburg, The Netherlands. He is interested in investigating how humans and machines can ground the meaning of linguistic utterances in the real world, and how they learn language from each other through social interactions. To study this, he uses a variety of techniques, ranging from agent-based modeling, child–robot interaction, and psycholinguistic experiments to ethnographic research of childrens language acquisition in different cultures. His most recent line of research concerns the application of social robots in (second language) education.

Chen Yu received the M.Sc.Eng. degree in automation from Beijing University of Technology, Beijing, China, in 1998, and the Ph.D. degree in computer science from the University of Rochester, Rochester, NY, USA, in 2004.

He is a Professor with the Department of Psychological and Brain Sciences, Indiana University, Bloomington, IN, USA, where he is also affiliated with the Cognitive Science Program and the School of Informatics, Computing and Engineering. His research focuses on using head-mounted eye tracking, computational modeling and data mining techniques to understand mechanisms of early language learning, statistical learning, visual and social attention, and multimodal human–human and human–robot interactions. He has published over 150 papers in the fields of developmental psychology, cognitive science, computer vision, and artificial intelligence. His research has been funded by NIH and HSF. He has also served as a panelist for multiple NIH study sessions and NSF panels.

Prof. Yu received the Early Career Award from the International Society on Infant Studies in 2008 and the Robert L. Fantz Memorial Award from the American Psychological Foundation in 2013. He is currently a Fellow of the Association for Psychological Science.

Michael Spranger received the Ph.D. degree in computer science from Vrije Universiteit Brussel, Brussels, Belgium, in 2011.

He was a Researcher with Sony CSL, Paris, France. He then worked with the Research and Development Department, Sony Corporation, Tokyo, Japan, for two years. He is currently a Researcher with Sony Computer Science Laboratories Inc. Tokyo. He is a roboticist by training with extensive experience in research on and construction of autonomous systems, including research on robot perception, world modeling, and behavior control. After his undergraduate degree, he fell in love with the study of language and has since worked on different language domains from action language and posture verbs to time, tense, determination, and spatial language. His work focuses on artificial language evolution, machine learning for NLP (and applications), developmental language learning, computational cognitive semantics, and construction grammar.