Situated Language Understanding for an Anthropomorphic Robot based on Linguistic Constructions and Embodied Simulation

Matthias A. Priesters, Malte Schilling, Stefan Kopp
Bielefeld University, Cognitive Interaction Technology (CITEC)

How do actors gain an understanding of their environment? On the one hand, they familiarize themselves with physical objects through bodily interaction and exploration; on the other hand, they gain understanding through linguistic interaction with other knowledgeable agents. We report work in progress from a project investigating such familiarization processes through modeling them in an anthropomorphic robot. The robot consists of a torso equipped with two arms with touch-sensitive hands enabling familiarization with novel objects and their affordances through bimanual action. This process is to be guided through situated dialogue with a human instructor, possibly including for example deictic expressions, anaphors or underspecified references to actions, such as in the following: "Open this box." – "*Okay.*" – "No, you have to grasp it at the top." – "*Like this?*" – "Yes, good, keep going". Our approach aims to model the needed linguistic and dialogue abilities.

We follow an embodied cognition approach, proposing three layers of representation, which ground higher-level knowledge in lower-level experiences and sensorimotor behaviors: The lowest (*sensorimotor*) level is constituted by sensor input and motor primitives, i.e. control structures of the robot. The highest (*linguistic*) level deals with speech input and generating answers. Between these two layers, a *schematic* level of representation mediates between symbolic language representations and sensorimotor information. For our embodied approach, action and the execution of actions are central. Therefore we focus on executable action schemas (implemented as X-Schemas; Narayanan 1999), which are selected and parameterized based on linguistic input and which directly provide a level of mental simulation and prediction. Crucially, the schematic layer is active during language understanding as well as during action execution: During language understanding it can invoke appropriate action schemas and test different hypotheses to find the most likely meaning of the language input. During execution it keeps the system aware of the current action state and facilitates situated language interpretation as well as precise error reporting or requesting missing information.

In its current state, the system parses linguistic input using the Embodied Construction Grammar (ECG) formalism (Bergen & Chang 2013, Bryant 2008). ECG in general works by mapping linguistic form to meaning in terms of embodied schemas (based on image schemas). We aim to enhance embodied language processing by (1) tying sensorimotor information into the schemas by means of the intermediate layer mentioned above and (2) incorporating situational knowledge (i.e. the configuration of the environment and the current action states) into language processing by coupling the parser to a knowledge base and the schematic representation layer. Thus, we aim to improve comprehension of dialogue acts pertaining to currently ongoing actions by enabling the schematic layer to influence parsing decisions.

We will present the current state of our project, discuss theoretical and practical requirements for the implementation of the formalism and demonstrate first results regarding situated simulation-based natural language understanding.

References

Bergen, Benjamin K.; Chang, Nancy (2013): Embodied Construction Grammar. In Thomas Hoffmann, Graeme Trousdale (Eds.): The Oxford handbook of construction grammar. Oxford University Press, pp. 168–190.

Bryant, John (2008): Best-fit constructional analysis. Doctoral dissertation, University of California, Berkeley. Available at http://icsi.berkeley.edu/~jbryant

Narayanan, Srinivas (1999): Reasoning about actions in narrative understanding. In Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI '99), pp. 350–358.