The Need for Interactional Scenarios in Grounded Language Learning

Kerstin Fischer
University of Southern Denmark
kerstin@sitkom.sdu.dk

Abstract

This study addresses the experiential basis for grounded language learning experiments and the role interaction plays in such experiments; while recent studies concern the grounding of linguistic structure in sensorimotor experience and through dialogue games, no attention so far has been paid to interaction as a domain of experience, which constitutes the basis for constructions such as grammatical mood. The grounding of these constructions can however only take place in interactional learning scenarios.

1. Introduction

Recent years have seen parallel developments in cognitive linguistics, psychology and robotics concerning the embodied nature of natural language. According to this trend, language is grounded in experience, grammatical structure is taken to reflect cognitive categorization, and grammar and cognition are taken to co-develop in the child during language acquisition (e.g. Barsalou 1999; Langacker, 2008; Tomasello, 2003). These notions underlie numerous recent studies on language emergence and automatic language learning (e.g. Steels, 2004, 2009; Dominey, 2006). The notion of interaction considered, however, is usually restricted to sensorimotor experience and to the playing of well-defined, pre-specified language games.

2. Previous Work

The grounding of linguistic meanings in experience has been investigated in language emergence and in language learning experiments. For instance, Steels (e.g. 2004, 2009) demonstrated in various language emergence experiments in which robots negotiated linguistic descriptions of perceived scenes that sets of intersubjectively shared conventions and linguistic structure may emerge from such interactions. That is, linguistic constructions such as equivalents of the English ditransitive construction, may be developed by interactants as a result of communication pressure only, rendering the assumption of innate linguistic categories superfluous. Extending such peer-to-peer description games, also other areas of language, such pronouns or spatial and temporal concepts (e.g. van Trijp 2010, Steels et al, submitted, Schulz et al. 2008) have been shown to emerge.

In the language learning paradigm, Sugita and Tani (2005, 2008) have shown that the system can learn to break down holophrases into parts and recombine them to understand new sentences if presented with pairs of actions and holophrases. In addition, recent experiments illustrate how the learning of linguistic labels and perception may co-determine each other (Cangelosi et al., 2007; Marocco et al., in press). That is, robots learn about objects by means of their perceptual, manipulative and linguistic properties.

Previous work on the grounding of language in experience has thus focused on a) interaction with the physical environment and b) interaction with other agents in well-defined, pre-specified dialogue games, focusing on peer-to-peer description, even if these involve question-answer pairs (cf. Micelli and van Trijp, submitted).

3. The Role of Interaction

In spite of its central role in the theoretical premises underlying grounded language learning, so far interaction thus plays a rather limited role: it concerns either interaction with the physical environment or the process in which linguistic conventions are negotiated and interactively established. Yet interaction as the domain of experience that gives rise to grammatical meaning itself is not taken into account.

To understand the impact of this more narrow view taken in epigenetic robotics, let us consider the scenarios in which language learning experiments may take place. For instance, the learning scenario used by Sugita and Tani (2005, 2008) consists in remote-controlling the robot's behaviour, so that the robot can build up the meaning of an action like *push*, *point* or *hit* from generalising over the sensory input it receives from its own (remote-controlled) actions. This paradigm is called learning by demonstration.

In current experiments in the framework of the ITALK project (www.italkproject.org), we are extending this framework to increasing linguistic complexity, aiming at demonstrating the learning of several different argument structure constructions and of more naturalistic verb-construction pairings (Zeschel and Fischer 2009). The learning stimuli consist of holophrastic imperative clauses, such as *push-the-block*. The robot learns to analyse the linguistic parts of the holophrases and to carry out actions corresponding to novel combinations of these component parts. Thus,

asked to "push-the-block-to-the-left", the robot will push the block to the left, even if it has not seen this utterance before but has only been familiarized with utterances such as "push-the-cup-to-the-left" and "pull-the-block". It has its own, grounded, representations of cups and blocks, as well as of pushing and pulling. However, although the robot possesses grounded representations of the action, the object and the caused motion construction, and the choice of imperative mood provides plausible results in the scenario chosen, its meaning is not understood by the robot in a learning by demonstration scenario. An understanding of the imperative construction implies an understanding of its role in interaction, which would need to be grounded in interaction itself.

Other grounded learning scenarios are based on learning by observation (cf. Steels, 2009). Here, the robot's utterances correspond to declarative sentences describing the scene perceived. For instance, one robot will suggest to the other how to describe a certain scene, such as Jill pushing a block to Jack. The other robot will either accept the description or propose another one, until the robots have jointly negotiated the linguistic representation of the perceived scene.

However, since a robot in these experiments has no choice but to produce structures corresponding to declarative sentences which are the indirect result of the in-built language game the robot is designed to play, it does not understand the meaning of the declarative mood either, i.e. that it is describing a scene to someone, communicating a certain state of affairs. Thus, in both scenarios the robots ground actions, objects and argument structure constructions in their perception, yet they do not possess a grounded understanding of the pragmatics of the utterances as a whole.

4. Conclusion

As Langacker (2008: 470) argues, "the speaker-hearer interaction is part of an expression's meaning, whether or not it is put onstage and profiled." Thus, we cannot produce utterances without making a choice for the one or the other grammatical mood. Understanding the meaning of an imperative, interrogative or declarative clause however presupposes an understanding of its role in interaction. The only way for a robot to learn natural language utterances in a grounded manner is therefore from interaction. Interaction thus has an impact on the symbolic structures of language themselves, which needs to be accounted for in future grounded language learning experiments.

References

Barsalou, L. W. (1999). Perceptual symbol systems. *Behavioral and Brain Sciences*, 22(4), 577-660.

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

Dominey, P.F. (2006). From holophrases to abstract grammatical constructions: Insights from simulation studies. In: E. Clark & B. Kelly (Eds.), *Constructions in Acquisition* (137-162). Stanford: CSLI Publications.

Langacker, R. W. (1998). Conceptualization, symbolization, and grammar. In M. Tomasello (Ed.), *The New Psychology of Language.* Englewood Cliffs, NJ: Erlbaum.

Langacker, R. W. (2008). *Cognitive Grammar: A Basic Introduction*. New York: Oxford University Press.

Marocco, D., Cangelosi, A., Fischer, K. and Belpaeme, T. (in press). Grounding action words in the sensory-motor interaction with the world. *Journal of Neurorobotics.*

Micelli, V., van Trijp, R. (submitted). Information Structure in Grounded Embodied Agents. http://www.fcg-net.org/information-structure/demos/langro.html

Schulz, R., Prasser, D., Stockwell, P., Wyeth, G., and Wiles, J. (2008). The formation, generative power, and evolution of toponyms: Grounding a spatial vocabulary in a cognitive map. *Proceedings of EVOLANG7,* 267-274.

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

Steels, L. (2009). Cognition and Social Dynamics Play a Major Role in the Formation of Grammar. In D. Bickerton and E. Szathmáry (eds.), Biological Foundations and Origin of Syntax, Cambridge, MA: MIT Press

Steels, L., Loetzsch, M., Pauw, S., Spranger, M. (submitted). The Comprehension and Production of Spatial Language by a Humanoid Robot. http://www.fcg-net.org/space/langro-web-demo/index.html

Sugita, Y. and Tani, J. (2005). Learning semantic combinatoriality from the interaction between linguistic and behavioral processes. *Adaptive Behavior*, 13.3.211-225.

Sugita, Y. and Tani, J. (2008). 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).

Tomasello, M. (2003). *Constructing a Language*. Cambridge, Mass.: Harvard University Press.

van Trijp, Remi (2010). Strategy Competition in the Evolution of Pronouns: A Case-Study of Spanish Leísmo, Laísmo and Loísmo. In: *Proceedings of EVOLANG* 8.

Zeschel, A. and Fischer, K. (2009). Constructivist grammar classifications for grounded language learning. ITALK D3.1, www.itakproject.org.

Acknowledgement

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