WHY ROBOTS?

MARTIN LOETZSCH
AI-Lab, Vrije Universiteit Brussel, Pleinlaan 2, 1050 Brussels, Belgium
Martin.Loetzsch@vub.ac.be

MICHAEL SPRANGER
SONY Computer Science Laboratory Paris, 6, Rue Amyot, Paris, 75005, France

In this paper we offer arguments for why modeling in the field of artificial language evolution can benefit from the use of real robots. We will propose that robotic experimental setups lead to more realistic and robust models, that real-word perception can provide the basis for richer semantics and that embodiment itself can be a driving force in language evolution. We will discuss these proposals by reviewing a variety of robotic experiments that have been carried out in our group and try to argue for the relevance of the approach.

1. Introduction

Computational modeling has become an invaluable tool for studying the origins and evolution of human languages. Mathematical investigations and computer simulations help us to test whether assumptions of a particular theory are explicit, detailed and consistent enough so that the operationalization of that theory can generate phenomena found in reality. Furthermore, computational models allow researchers to manipulate both external conditions as well as details of the assumed mechanisms in a controlled way that is not possible with human subjects. Standards for scientific experimentation using computer models are starting to be established within the evolution of language community (e.g. Cangelosi & Parisi, 2002; Steels, 2006) and work that follows these methodologies gets more and more accepted.

From the very beginning on, robots have been used to carry out experiments on the origins and evolution of language. And a question that is often asked by reviewers of such work is: *What is the added value of using robots for that particular study?* This is a very valid question and in fact a lot of the computational work on the evolution of language has been done successfully without robots. A common strategy is to scaffold all aspects related to perception (and often also conceptualization) by presenting agents with stimuli generated by world simulators of varying detail. Such generated stimuli can range from simple "meaning vectors" or readily pre-conceptualized semantic structures to more complex de-

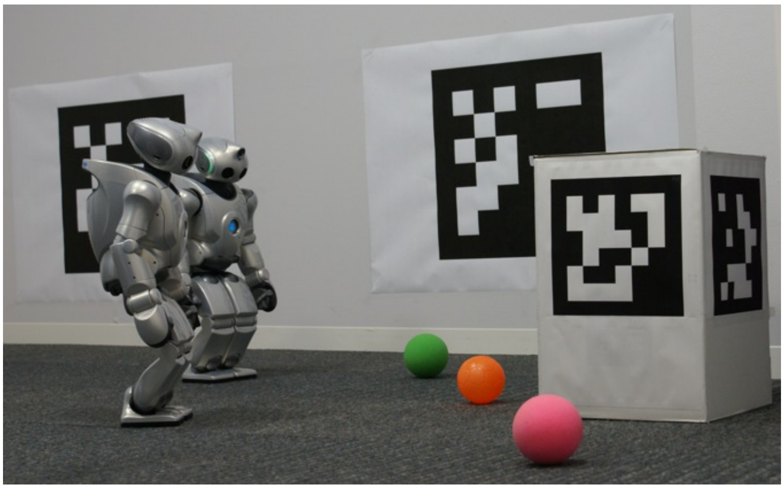

Figure 1. Robotic setup. Two robots are placed in an office environment consisting of colored objects, carton boxes and markers on the walls. Both robots engage in communicative interactions about objects in their environment.

scriptions of perceptual input that require further conceptualization. The advantage of this methodology is that researchers can focus on the "core" cognitive functions and representations that are involved in the processing and learning of language without having to deal with seemingly "remote" problems.

Why one would want to remove this scaffold can be easily answered when the topic of research is directly related to symbol grounding or human-robot interaction: such work simply requires to use robots. But for research on how and why a particular feature of language evolved, the value of using robots is much less obvious. And whereas other disciplines, such as the relatively new field of biorobotics, have already well-established criteria for judging whether a robotic implementation is a good model for a (biological) phenomenon (e.g. Webb, 2001), such a consensus does not exist for the field of language evolution. Nevertheless, as we will argue in this paper, there are clear and undoubtable advantages to using robots. In fact, we will claim that in order to start an investigation into language from the right point of view and to avoid getting trapped in solutions and theories of particular phenomena that are misleading or, even worse, false, robotic experiments are indispensable.

We will propose some answers to the question of "why robots?" by discussing a variety of language game experiments with Sony humanoid robots (Fujita et al., 2003, see Figure 1) that have been carried out in our group over the past few years. Compared to similar experiments based solely on computer simulations, setting up such experiments immediately becomes much more complex and difficult. In addition to finding the appropriate cognitive functions and representations for using and learning language, the robots have to be endowed with mechanisms for visual perception, joint attention and social interaction (pointing, non-linguistic feedback, etc.). In the next section, we will try to demonstrate that this additional effort is justified.

2. Why robots?

We will make four arguments for why the use of robots can lead to better models of language evolution: it increases the *realism* of models, it leads to more *robust* models, it provides *richer semantics* and, finally, real embodiment can serve as a *driving force* for language evolution.

2.1. *Increased realism*

One of the best-studied models on the origins of language is the naming game (Steels, 1995). In such a scenario, simulated agents of a population engage in local communicative interactions and over the course of many such interactions create and align a shared lexicon of proper names for individual objects. It is the simplest lexicon formation model that can be imagined and therefore proved to be an "E. coli paradigm" for investigating alignment strategies, mathematical proofs of convergence, impact of network structure and so on. Since it is such a simple model, it has also led to views that proper names are semantically simpler than words for kinds of objects (e.g. "red" or "block") and that they might be precursors of compositional communication systems (e.g. in Steels, 2005).

However, when this model is brought to real robots, as done by Steels, Loetzsch, and Spranger (2010), then it turns out that the dynamics of the grounded naming game differ drastically from the non-grounded version and that the underlying semantics of proper names are much more complex. Whereas in the non-grounded version word meanings refer directly to pre-given shared symbolic representations of individuals, in the grounded variant these representations need to be constructed from the continuous flow of visual perceptions. Since each individual physical object can be viewed from different angles and thus may look very different each time it is encountered by a robot, the agents can not know a priori whether a perception of an object belongs to the same individual or not. Additional heuristics, such as temporal-spatial continuity need to be employed to successively construct mental representations of individual objects. Consequently, tackling the emergence of proper names by using real robots introduces an additional level of realism because it forces researchers to incorporate processes for object individuation in their model instead of assuming them.

2.2. *Robust models*

Robots provide a tough testing ground for computational models in terms of their robustness. Agents can view a scene from different angles, lighting conditions may vary and thus the perceptions that two different robots have of the same physical object will never be the same. Even a single robot will perceive an object differently over the course of time due to camera noise, robot motion and general uncertainty in computer vision systems. Nevertheless, human concepts, such as, for example, the color red, are robust to such influences – we will recognize an object as red under very different lighting conditions and even subjects with color deficiencies are often able to communicate about colors.

One challenge stemming from real-world perception is *perceptual deviation*, i.e. that specific continuous features (e.g. position, shape, width and height, color information, etc.) computed by the vision system for an object differ drastically between the perception of speaker and hearer. For example one robot might perceive the height of an object as being `0.72` and the other one as `0.56`. This will inevitably cause each agent to have a different notion of a word such as "high". Additionally, perceptual deviation makes the task of guessing what a novel word refers to (whether it is about an object as a whole, one particular sensory channel or a combination of features) harder than when simulated shared contexts are used.

The problem of inferring the meaning of an unknown word has been extensively studied in the field of artificial language evolution (e.g. Smith et al., 2006). However, many of these models work only when tested in shared simulated contexts and could not be successfully transferred to robotic scenarios. Trying to overcome this lack of robustness, Wellens, Loetzsch, and Steels (2008) proposed a lexicon formation model that challenged the way previous approaches represented word meanings and tackled the task of word learning. Perceptual uncertainty was put at the core of word meaning representations and thus the agents learned to rely less on sensory channels with higher perceptual deviation. As a result, the model was not only able to cope well with perceptual deviation but also turned out to scale better with increasing population size and larger meaning spaces in simulated environments (Wellens, 2008). Hence, using robots and not relying on simple world simulations can lead to qualitatively new models that exhibit properties closer to human language.

Furthermore, real-world perception can provide more structured stimuli than randomly created artificial contexts. For the domain of color, Bleys et al. (2009) systematically analyzed the impact of using robotic vision compared to artificial contexts (as used in anthropologic color research) on the performance of color naming games. They found their model to be robust with respect to perceptual deviation because perceived colors of objects are usually around prototypical centers of color categories.

2.3. *Rich semantics*

Rich conceptual structures are a requirement for the emergence of grammar. When one follows a functional perspective on language, then grammatical structure arises in large part due to the multiplicity and ambiguity of conceptualizing the world. More precisely, grammar gets shaped and adapted to solve problems emerging from ambiguity in interpretation or explosion of search in parsing and production (Steels & Wellens, 2006).

Prime examples for such phenomena can be found in the domain of space, where different ways of conceptualizing the same spatial scene compete. For example, the English utterance "in front of the TV" is ambiguous, because English allows speakers to conceptualize the world using *relative* or *intrinsic* frames of references (Levinson, 2003). The phrase can mean before the screen of the TV (intrinsic frame of reference, using the front of the TV) or before the screen from the viewpoint of the speaker. In many contexts there might be no difference between these two interpretations as they might happen to discriminate the same referent. But in certain cases English also requires speakers to disambiguate the meaning, by marking the particular conceptualization strategy used.

In order to have conceptual structures that leave room for ambiguities, the perceptual space underlying conceptualization needs to be complex enough so that the reality can be construed in different ways. Robotic models are a way to appreciate this: there are numerous ways of how interlocutors can be positioned relative to each other and hence how they view a scene, there is non-linear perceptual noise in position estimation and there are numerous ways to choose objects as reference points. Thus, rich perceptions of spatial setups have proven to be crucial for investigating the emergence of spatial perspective reversal (Steels & Loetzsch, 2009) and the alignment frames of reference choice (Spranger et al., 2009).

2.4. *A driving force for language evolution*

Many theoretical proposals of how language conveys meaning have focussed on the grounding of language in the body (Johnson, 1987), on how specific systems of language interact with sensorimotor processes and how semantic and syntactic structures become recruited from one domain to another, for example, from the bodily domain to the domain of space (MacLaury, 1989) or from space to time (Kuteva, 1999). Linguists, especially those in the cognitive linguistics tradition, have hypothesized that the adaptation and exaptation processes that guide such conceptual transfer are deeply rooted in the concrete embodiment of humans and our particular interaction with the environment.

For instance, Lemmens (2002) demonstrated how posture verbs such as *sit, stand, lie* become metaphorically extended to the domain of space in Germanic languages. Clear examples can be found in the Dutch language, where posture verbs have been extended from their original bodily meaning (their anthropocentric prototypical semantic structure denoting human postures) to animals, things and even abstract spaces and entities. In Dutch butter *lies* in the fridge and one even *sits* in an economical crisis.

To model such transfer processes, embodiment needs to be taken seriously because the interaction with the environment plays a crucial role both in constructing conceptual structures and linking them across different perceptual domains. For example Spranger and Loetzsch (2009) and Steels and Spranger (2009) used humanoid robots to show how the semantics of posture verbs can emerge from sensorimotor interaction, how these conceptual structures can be linked to language and, finally, how bodily representations can be metaphorically extended. It is hard to imagine how such processes could have been studied without using real robots because 1) the meaning of posture verbs is directly grounded in behavior and 2) perceptual capabilities and representations are at the heart of the exaptation process.

2.5. *Discussion*

It could be argued that all of the work cited above can be done in (sometimes complex) simulations, which would make setting up the experiments less difficult and time consuming, without decreasing the realism of models. This article in no way refutes the idea of simulating phenomena as an invaluable source of knowledge, inspiration and scientific progress. In fact, we ourselves have used simulations of different types ranging from full-blown world simulators including a simulated physical environment with agents, to simulators that reproduce the output of particular cognitive systems such as perception or conceptualization. In the process of building artificial systems to study certain aspects of language evolution one often anyway finds oneself building simulators for the purpose of testing and studying subsystems.

Setting up simulators seems easier on first sight, but in fact it is not. First, developing realistic simulations of complex interactions with a virtual world often turns out to be as difficult (if not even more) than dealing with actual robots. And second (more importantly), it entails a lot of decisions and assumptions that actually require justification.

When building a simulation one inevitably needs to make certain choices as to how the particular simulated entity behaves, what properties it has, for instance which noise and timing properties are assumed, but also which representations are the interface between the simulation and the system studied. These choices are governed by a set of explicit or implicit assumptions that restrain the test space of the system in question. This is true for physical simulations, where the choice amounts to which physical properties one includes in the simulation and how they are interacting with the studied system, as well as for more abstract simulations that for instance simulate the output of certain cognitive systems. These assumptions can of course be discussed and ideally researchers make an effort to find all hidden and implicit assumptions in their simulated models, but it seems hard to prove the realism of a particular model without showing its operation in the real world. After all "…physical robots cannot violate the laws of physics, even if those laws are unspecified by the investigator. The performance of a physical robot is immediately informative about what works and what does not" (Long, 2007, p.1193, see also Webb, 2000, p.552, for a discussion of simulated robots vs. real robots).

One point of computational modeling in this field is to operationalize and therefore test theories of language evolution in concrete experiments. And since language is a way of interaction of real agents in the real world, showing the adequacy of the proposed solution cannot be avoided by resorting to the success of a system in simulation. Hence, building a fully autonomous system interacting with the real world entails making certain choices on all levels of cognitive systems including perception and action, but building a simulation involves important choices and specific assumptions about how certain subsystems or the physical world works that need to be justified. While these choices are always explicit in a real world computational systems, they might be hidden and unnoticed in simulation based approaches.

3. Conclusions

This paper discussed how exploiting the rich sensorimotor interaction of real robots with a physical world is a valuable methodology for investigating the evolution of language. Computational models using real robots benefit from higher realism, increased robustness and richer semantics. Thus, this paper compiled evidence as to why the study of language evolution should encompass real robots, in turn, arguing for a "whole systems" approach that aims to integrate processes of embodiment, sensorimotor intelligence, cognition, and social interaction.

Acknowledgements

Reported research was partly funded by the EU project ALEAR and carried out at the Sony CSL in Paris and at the AI lab of the University of Brussels. We are extremely grateful to Masahiro Fujita, Hideki Shimomura and their team at the Sony Intelligent Systems Research Lab in Tokyo for giving the opportunity and support to work with the Sony humanoid robot.

References

Bleys, J., Loetzsch, M., Spranger, M., & Steels, L. (2009). *The grounded colour naming game.* To appear in Proceedings Roman-09.

Cangelosi, A., & Parisi, D. (2002). Computer simulation: a new scientific approach to the study of language evolution. In *Simulating the evolution of language* (pp. 3–28). New York, NY, USA: Springer-Verlag New York, Inc.

Fujita, M., Kuroki, Y., Ishida, T., & Doi, T. T. (2003). Autonomous behavior control architecture of entertainment humanoid robot SDR-4X. In *Proceedings of the IEEE/RSJ IROS '03* (pp. 960–967, vol. 1). Las Vegas, Nevada.

Johnson, M. (1987). *The body in the mind: The bodily basis of meaning, imagination, and reason.* University of Chicago Press Chicago.

Kuteva, T. (1999). On'sit'/'stand'/'lie'auxiliation. *Linguistics*, *37*(2), 191–213.
Lemmens, M. (2002). The semantic network of dutch posture verbs. In J. Newman (Ed.), *The linguistics of sitting, standing and lying.* Amsterdam/Philadelphia: John Benjamins.
Levinson, S. (2003). *Space in language and cognition: Explorations in cognitive diversity.* Cambridge University Press.
Long, J. H. (2007). Biomimetic robotics: self-propelled physical models test hypotheses about the mechanics and evolution of swimming vertebrates. *Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science*, *221*(10).
MacLaury, R. (1989). Zapotec body-part locatives: Prototypes and metaphoric extensions. *International Journal of American Linguistics*, 119–154.
Smith, K., Smith, A. D. M., Blythe, R. A., & Vogt, P. (2006). Cross-situational learning: A mathematical approach. In *Symbol grounding and beyond: Proceedings of EELC 2006* (Vol. 4211, pp. 31–44). Springer Verlag.
Spranger, M., & Loetzsch, M. (2009). The semantics of SIT, STAND, and LIE embodied in robots. In *Proceedings of Cogsci'09.* Cognitive Science Soc.
Spranger, M., Pauw, S., & Loetzsch, M. (2009). *Open-ended semantics co-evolving with spatial language.* Submitted to Evolang 8.
Steels, L. (1995). A self-organizing spatial vocabulary. *Artificial Life*, *2*(3).
Steels, L. (2005). The emergence and evolution of linguistic structure: from lexical to grammatical communication systems. *Connection Science*, *17*(3/4).
Steels, L. (2006). How to do experiments in artificial language evolution and why. In *Proc. of Evolang 6.* World Scientific Publishing.
Steels, L., & Loetzsch, M. (2009). Perspective alignment in spatial language. In *Spatial language and dialogue* (pp. 70–89). Oxford University Press.
Steels, L., Loetzsch, M., & Spranger, M. (2010). A boy named sue. the semiotic dynamics of naming and identity. *submitted*.
Steels, L., & Spranger, M. (2009). How experience of the body shapes language about space. In *Proceedings of IJCAI 09.* Pasadena (CA).
Steels, L., & Wellens, P. (2006). How grammar emerges to dampen combinatorial search in parsing. In *Symbol grounding and beyond* (Vol. 4211, pp. 76–88). Rome, Italy: Springer Verlag.
Webb, B. (2000). What does robotics offer animal behaviour? *Animal Behaviour*, *60*(5), 545–558.
Webb, B. (2001). Can robots make good models of biological behaviour? *Behavioral and Brain Sciences*, *24*(6), 1033–1050.
Wellens, P. (2008). Coping with combinatorial uncertainty in word learning: a flexible usage-based model. In *Proceedings of Evolang 7.* World Scientific.
Wellens, P., Loetzsch, M., & Steels, L. (2008). Flexible word meaning in embodied agents. *Connection Science*, *20*(2 & 3), 173–191.