From symbol grounding to socially shared embodied language knowledge

Arne Zeschel
IFKI
University of Southern Denmark
Alsion 2
6400 Sønderborg
Denmark
Email: zeschel@sitkom.sdu.dk

Elio Tuci
Computer Science Department
Llandinam Building
Aberystwyth University
Aberystwyth, Ceredigion
Wales SY23 3DB
Email: elt7@aber.ac.uk

I. Introduction

Much language-related research in cognitive robotics appeals to usage-based models of language as proposed in cognitive linguistics and developmental psychology [1, 2] that emphasise the significance of learning, embodiment and general cognitive development for human language acquisition. Over and above these issues, however, what takes centre stage in these theories are social-cognitive skills of "intention-reading" that are seen as "primary in the language acquisition process" [1] – and also as difficult to incorporate into computational models of language acquisition. The present paper addresses these concerns: we describe work in progress on a series of experiments that take steps towards closing the gap between 'solipsistic' symbol grounding in individual robotic agents and socially framed embodied language acquisition in learners that attend to *common ground* [3] with changing interlocutors.

A. Learning target and experimental design

The experiments focus on the acquisition and contextually appropriate interpretation of adnominal possessive constructions (e.g. *the dog's bone*). Adnominal possessives are a linguistic universal [4] and also among the first multiword constructions to emerge in early child speech [5]. The construction is heavily polyfunctional: among others, conventional meanings range from ownership (*John's car*), whole-part relations (*John's hand*) and kinship ties (*John's brother*) over mere disposal and context-specific association (*John's train*) to abstract participant-event (*John's arrival*) and even setting-event relationships (*last year's meeting*). In each case, the abstract core function of the construction is to narrow down the reference of the second nominal to the particular instance that is intended by the speaker. We restrict our attention to two readings that are amenable to a grounded robotic language learning approach, viz. the 'disposal' and 'ownership' interpretations of the construction.

Psycholinguistic research suggests that 'disposal' interpretations in which a human participant has privileged access to a spatially collocated concrete object are developmentally basic [6]. From here, learners' understanding of 'ownership' in the narrower sense develops from the realisation that the privileged access of a participant to an object may also be habitual, in which case it will supersede the privileges of other participants that have the object at their disposal merely temporarily. In addition, felicitous use of the construction involves social perspective taking by tailoring one's own production/comprehension of the construction to the assumed knowledge state of the present interlocutor: in a scenario where there are several instances of a particular object, the addressee will only be able to identify the actually intended referent of the construction if both interlocutors engage in a change of perspective and assess their common ground. The task for the speaker then becomes: given what I know about the knowledge state of the addressee, which formulation should I choose to guide the addressee's attention to the intended object? Conversely, the corresponding task of the addressee can be explicated as follows: given what I know about the knowledge state of the speaker, which object can I entrust the speaker to have intended by the chosen formulation? There are thus at least three important aspects to competent use of the construction in context: (i) the ability to narrow the reference of a class expression to a particular instance that is identified via an entity that is encoded as its possessor; (ii) the realisation that the entity functioning as possessor (i.e., the referent of the modifier nominal) may either be the one at which the target object is currently located (disposal) or another one where it is usually found (ownership); and (iii) the ability to relate one's own beliefs about the relation between the given possessee and a potential possessor to those of the partner (assessment of common ground).

As a preparation for the actual experiment, the robotic agent first acquires linguistic labels for a set of landmark entities X (the later possessors) and a set of trajectory entities Y (the later possessees) by learning to point to the appropriate entity in

response to the instruction *Point to X/Point to Y*. In the first step, landmark and trajectory entities are then presented together in the same scene, with the appearance of particular objects Y at particular landmarks X switched around between trials. There is always more than one instance of the given target category Y in the observed scene, and the instruction changes to *Point to X's Y*. In this step, the goal is to identify that instance of Y that is located at X (i.e., to learn the 'disposal' meaning of the construction). The second step then introduces the 'ownership' variant of the construction. Here, the learner keeps track of its categorisation history, and is rewarded for supplying responses compliant with either the 'disposal' or the 'ownership' interpretation (i.e., when pointing either to the object that is currently located at X or to the one that is habitually located at X, though not in the present trial). The final step then involves the introduction of the social dimension. The experiment proceeds as in step 2, but at a certain point, the agent's instructor changes. In the social setting, the same expression *X's Y* can be resolved differently depending on interlocutors' shared experience in their interaction history: if both interlocutors are aware of an ownership relation between a particular X and a particular Y that has emerged from their habitual co-occurrence, responses corresponding to both 'ownership' and 'disposal' readings would be appropriate for the learner (since both are conventional interpretations of the construction). By contrast, if the speaker is not aware of such a relationship, the learner should invariably choose the 'disposal' variant. Hence, the goal of the agent in this setting is not merely to keep track of its own categorisation history, but also of that of its interlocutor.

B. General properties of the computational model

Behavioural and linguistic skills are grounded in the sensory-motor experience of the agent by using integrated artificial neural networks as controllers, and artificial evolution to define the network parameters [7]. Our robotic platform of reference is the iCub [8]. The learning scenario is similar to the one illustrated in [9], with populations of agents repeatedly playing a language game in which they are incrementally evaluated for their ability to correctly parse and carry out the changing linguistic instructions with which they are presented. The agent is equipped with a simple visual system so as to be able to perceive salient object features such as colour and shape, as well as to extract the relative position of the objects with respect to the agent's own body. The agent also attends to the position of the arm that it uses to point to objects with proprioceptive sensation. Linguistic items are modeled as binary vectors which uniquely identify either physical properties of the objects or spatial locations relative to the agent. The structure of the instruction is modeled as follows: when both X and Y are not-null vector, then the structure is meant to be *Point to X's Y*. A combination of a null vector and a not-null vector represents *Point to X*/*Y* type instructions. The agent learns the meaning of the linguistic vectors by experiencing them in different circumstances. The controller activates the joints that move the arm in order to point to the target object. The presence/absence of a given interlocutor is a binary input.

II. CONCLUSION

Complementing the sophisticated work on robotic language games in approaches that make use of a symbolic grammar formalism [10], the present approach explores key extensions to grounded language learning with neural network architectures: in these extensions, the robotic learner is required to (i) transfer grounded language knowledge across two different grammatical constructions; (ii) form semantic abstractions such that perceptually different members of the same underlying category can be identified using the same linguistic label; (iii) strategically employ this grounded language knowledge in communication with changing partners on the basis of assumptions about the current dyad's common ground. In this, the study offers a contribution to the discussion of perspectives and limitations of current neural network approaches to grounded robotic language learning.

ACKNOWLEDGMENT

This research was supported by the ITALK project (EU ICT Cognitive systems & robotics integrating project, grant 214668).

REFERENCES

[1] M. Tomasello, *Constructing a language. A usage-based theory of language acquisition*. Cambridge/Mass.: Harvard University Press, 2000.
[2] A. Goldberg, *Constructions at work: The nature of generalization in language.* Oxford: Oxford University Press, 2006.
[3] H. Clark, Language and language users. In: *The handbook of social psychology*, 3rd edition, ed. G. Lindzey & E. Aronson. Harper Row, 1985.
[4] B. Heine, Possession. *Cognitive sources, forces and grammaticalization*. Cambridge: Cambridge University Press, 1997.
[5] E. Lieven, D. Salomo and M. Tomasello, Two-year-old children's production of multiword utterances: a usage-based analysis. *Cognitive Linguistics* 20(3): 481–507, 2009.
[6] M. Tomasello, One child's early talk about possession. In: The Linguistics of Giving, John Newman (ed.), 349-373. Amsterdam/Philadelphia: John Benjamins, 1998.
[7] S. Nolfi, & D. Floreano. Evolutionary Robotics: The Biology, Intelligence, and Technology of Self-Organizing Machines. Cambridge, MA: MIT Press/Bradford Books, 2000
[8] G. Sandini, G. Metta, & D. Vernon, "The icub cognitive humanoid robot: An open-system research platform for inactive cognition," in *50 Years of Artificial Intelligence*, M. Lungarella, F. Iida, J. Bongard, and R. Pfeifer, Eds. Berlin, Germany: Springer-Verlag, pp. 358–369, 2007
[9] E. Tuci, T. Ferrauto, A. Zeschel, G. Massera and S. Nolfi, An Experiment on the Evolution of Compositional Semantics and Behaviour Generalisation in Artificial Agents. Special Issue on "Grounding Language in Action", *IEEE TAMD* 3(2), pp 1-14, 2011.
[10] L. Steels, The role of construction grammar in fluid language grounding. Unpublished manuscript, Sony Computer Science Laboratory, 2005.