Symbol Grounding Transfer with Hybrid Self-Organizing/Supervised Neural Networks

Thomas Riga Angelo Cangelosi Alberto Greco
Adaptive Behaviour & Cognition Research Group
School of Computing, Communication & Electronics
University of Plymouth
Drake Circus, PL4 8AA Plymouth, UK
thomas.riga@plymouth.ac.uk angelo.cangelosi@plymouth.ac.uk
Psychology Division
Department of Anthropological Sciences
University of Genoa
vico S. Antonio 7, Genoa, Italy
greco@unige.it

***Abstract* – This paper reports new simulations on an extended neural network model for the transfer of symbol grounding. It uses a hybrid and modular connectionist model, consisting of an unsupervised, self-organizing map for stimulus classification and a supervised network for category acquisition and naming. The model is based on a psychologically-plausible view of symbolic communication, where unsupervised concept formation precedes the supervised acquisition of category names. The simulation results demonstrate that grounding is transferred from symbols denoting object properties to newly acquired symbols denoting the object as a whole. The implications for cognitive models integrating neural networks and multi-agent systems are discussed.**

I. INTRODUCTION

Semantic interpretability is of great importance to modeling cognitive systems. In fact it has been the central theme in cognitivism, which considers the brain to be a symbol system and explains cognition as a manipulation of symbols governed by rules [6,7,16]. This approach doesn't resolve a fundamental question though: how these symbols are connected to their meanings. The cognitivist symbols don't have an inherent meaning: they reside in the mind of the interpreter and always require the semantic interpretation of an external observer. This is the symbol grounding problem [9], which affects all cognitivist models that deal with language learning and evolution.

The grounding of symbols can be tackled using a connectionist approach. Various neural network models exist that are able to extract invariant characteristics from input stimuli and build categories. Some models [e.g. 15, 17] explicitly involve the learning of labels (names) for the categories. However, attaching individual labels to conceptual categories is not sufficient for the development of a symbolic language. The names of categories have to be combined to form propositions using syntactic or other logic (e.g. boolean) constraints to be considered symbols. This would permit the expression of complex meanings by means of new combinations of symbols (i.e. linguistic propositions). Some connectionist models have made use of multiple symbols, although with limited compositional capabilities [2,5,8]. For example, Cangelosi et al. [2] studied the transfer of grounding from names for geometric shapes to the superposed categories "symmetric" and "asymmetric". This model was then extended [8] to include three grounding transfer simulations with both extensional (category information) and intensional (property information) categorical structures, e.g. object = superposed object category ("Circle = Symmetric"), object + property = object category ("Horse + Stripes = Zebra") & property + property = object category ("Red & Square = DAX").

An additional requirement for solving the symbol grounding problem is that of anchoring symbols to the sensorimotor abilities of organisms. Cognition is not exclusively explained by means of isolated symbolic and mental processes, but it also involves embodiment factors. Symbols have a physical grounding in the bodily experience of the environment with which a cognitive agent interacts [1,14]. In computational cognitive modeling, the embodied approach has been extensively used for studying low-level cognitive functions like sensorimotor coordination and active vision. Only recently has this approach been extended to include high-level cognitive abilities such as language [4]. While the connectionist models discussed simulate the grounding of language based solely on perceptual experience, embodied models integrate the perceptual and motor experience. To achieve it, neural networks have been combined with models of artificial life and adaptive behavior to simulate language evolution and language use in simulated agents and robots [e.g. 3,12]. For example, Marocco et al. [12] use a population of evolutionary robots that learn to manipulate objects in a simulated physical environment. The robots also learn to name objects through a process of communication using an evolutionary algorithm. The sensorimotor, cognitive and linguistic behavior of robots is controlled by a neural network. This permits the grounding of symbols (names) in their sensorimotor experiences.

Other embodied models study the acquisition of grounded language in robots without the use of connectionist systems. For example, Steels [18] uses robots that learn a grounded symbolic language through language games. He emphasizes the fact that language emerges as part of a specific interactive situation. Vogt [19] follows a similar approach defining symbols as structural connections between reality and sensorimotor activation deriving from the interaction between agent and environment. He uses mobile robotic agents that play adaptive language games. These studies point out that a situated and embodied multi-agent system permits us to study aspects of the evolution and acquisition of language otherwise out of reach.

Not all symbols can be directly grounded in our own perceptual and sensorimotor experience. We acquire most concepts from descriptions by others, rather than from direct experience with their referents. Moreover, we possess concepts that have no reference to anything actually existing in the outside world. We can invent new concepts on our own and talk about them thereby transferring the concepts into somebody else's mind. It is clear that language can't rely exclusively on a direct grounding of symbols on sensorimotor projections.

Linguistic descriptions are essential for knowledge acquisition and transfer. Typically a description of a concept consists of its name together with a list of properties, such as shape, size, color, functionality, usage, etc. (e.g. "A zebra is white with black stripes"), and/or with references to other concepts (e.g. "A zebra is a horse with stripes"). In this paper we will take into consideration the case in which a concept is defined by a list of its properties. These properties can be either grounded directly through interaction with the environment or can derive their grounding from descriptions. We hypothesize a process of grounding transfer from the names of the properties (directly grounded during the network training) to the names of the new concepts being defined (indirectly grounded during language learning). Furthermore, we state that this process essentially consists of a neural linking of the linguistic sign to sensorimotor representations of its defined properties. This implies that representations of symbols are essentially amodal: they are based on sensorimotor representations distributed across different parts of the neural network.

The new simulation reported in this paper further extends the work by Cangelosi and colleagues [2,8] by postulating a process of self-organization in which the ability to classify and categorize stimuli is acquired in an unsupervised manner. This process constructs analogue sensorial representations in a feature map on which names for properties are grounded. Only the relations between symbols and the objects they refer to and those between symbols and their symbolic definitions are acquired through supervised learning. In [17] a supervised algorithm (error backpropagation) was used for the acquisition of categories and their names. The present model is based on a more psychologically plausible view of symbolic communication: unsupervised concept formation precedes the supervised association of a category name [17]. The grounding of symbols is initially based on sensorimotor cognitive functions that spontaneously emerge from the agent's interaction with its environment. Only subsequently a supervised algorithm is employed for the acquisition of the basic symbols and for the transfer of knowledge through language.

II. EXPERIMENTAL SETUP

The model proposed here implements an autonomous cognitive system, immune to the symbol grounding problem. Its basic symbols (names of shapes and colors) are intrinsically connected to the categories being acquired through direct interaction with the environment. These symbols are successively applied to construct descriptions of new categories of stimuli (e.g. individuals made by specific combinations of one shape and color). New symbols are in this way defined without having a direct experience of their referents. This process of grounding transfer enables the system to express meanings that go beyond immediate experience. New symbols, acquired exclusively from symbolic descriptions, are ultimately grounded in the interaction of the system with its environment.

During the first phase, we present images of objects distinguished by different colors and shapes. The networks learn to discriminate between different stimulus categories by constructing a feature map in an unsupervised manner. This feature map expresses the intrinsic order of the stimulus set. The networks acquire in this way analogue sensorial representations of their environment that enable them to categorize the stimuli along the dimensions of shape and color.

During the second phase, we also present symbolic stimuli together with the images. Every picture is associated with symbols, consisting of arbitrary bit sequences that represent names for its color and shape properties. The networks are required to reproduce the same symbols in output. During this phase, symbols denoting color and shape properties are directly grounded in the sensorial representations acquired during the first phase.

During the third phase, the training input is exclusively symbolic. We use descriptions containing the previously acquired symbols in combination with a new symbol that denotes an object category. A sample description could be "Red & Square = DAX". Descriptions of both known and unknown objects are presented to the networks. For example, we can present images of red and blue squares and of green and red triangles, but never show a green square or a blue triangle. In any case, we can present the symbolic definition of a green square ("Square & Green =

SOD") during this learning phase. Only later, during the test phase, we will actually project on the retina the images of these previously unseen objects, to check if grounding transfer has occurred. If grounding has been transferred from the basic names of shapes/colors to the new names of individual objects, the network will effectively activate the output unit corresponding to name of the new images.

A. The stimulus set

The training and test stimuli consist of images of objects that vary along the dimensions of color, shape and position (Fig. 1 Left). Each image is a 5x5 pixel drawing. The objects depicted can be red, green or blue. The shapes are a square, a cross or a group of four dots. Every image (i.e. combination of a color and shape) is presented in nine different positions in the retina (Fig. 1 Right). Every pixel of the image is represented using three input units, with a continuous activation ranging from 0 to 1, encoding the red, green and blue primary components.

In total there were 81 images, 54 for training purposes and 27 reserved for testing. The names of properties and object categories are encoded with localist symbolic input and output units. The symbolic input may contain the names of the categories of different visual features (e.g. "Red", "Square"), the name of the object as a whole, or a full description of the objects (e.g. "Red + Square = DAX").

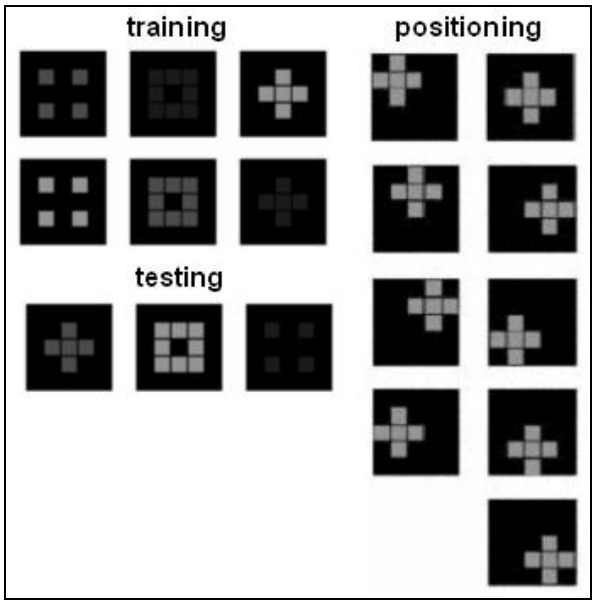

Fig. 1. (Left) The training and stimulus sets. Different levels of gray correspond to the colors red, green or blue. Stimuli are plotted in a 5x5 retina. (Right) Example of the 9 positions of one of the training stimuli.

B. The network architecture

The current model implements neural networks composed of two modules and a retina for input (Fig. 2). The first module is a two-layer Self-Organizing Map (SOM), while the second is a standard multi-layer perceptron (MLP). This new model is a further extension of previous connectionist models of symbol grounding and transfer [2,8]. Images are projected on the retina and are subsequently categorized by the first module. The second module receives as input the analogue sensorial representations of the first module and some additional symbolic input stimuli in the form of combinations of arbitrary bit sequences.

Two learning algorithms coexist within the model. The first module classifies stimuli using a Kohonen Self-Organizing Feature Map [11]. The second module connects the analogue representations emerged in the first network to the discrete symbolic stimuli via the standard error backpropagation algorithm [13]. SOMs are the result of a vector quantization algorithm that generates a mapping from the multi-dimensional stimulus space to a bi-dimensional matrix within which the similarity between different stimuli is expressed. In this way, the first module autonomously constructs an activation matrix in which the intrinsic order of the stimulus set is expressed.

The second module receives a symbolic input together with the activation matrix of the first module. Visual input, symbolic input and output stimuli are related through a direct learning process in which stimuli are presented simultaneously. Learning is supervised and consists in computing the error with respect to a teacher input and propagating this error backwards from output to input units and correcting the weight distribution. In this way the symbols are grounded on the representations of the first unsupervised module and are therefore grounded in the interaction between the system and the environment.

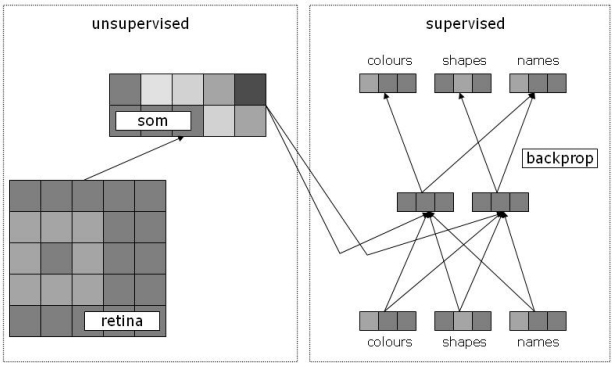

Fig. 2. The network architecture

C. Training and testing procedures

Learning is incrementally organized in three stages. During the various training phases, the networks resolve progressively more difficult tasks by basing their learning

on the representations that emerged during the previous phase. It is the modularity of the hidden layer that permits this kind of learning since representations for different categorization criteria are localized in different areas of the network. In this way we obtain a combination of a neural architecture and a learning procedure that enables the network to combine elementary constituents into higher-level representations.

During the first learning stage, the SOM network acquires categories for the stimuli projected on the retina through the unsupervised algorithm. The network generates a matrix which expresses the degree of similarity of the stimulus set and thus implicitly contains a division into categories.

In the second phase, the MLP networks learn to connect the SOM stimulus representations produced by the first module to the symbolic input and output units of the second module. These symbolic stimuli correspond to the names of the categories acquired during the first phase. The networks receive as input, at the same time, both visual and symbolic stimuli and learn to produce the corresponding symbolic output, consisting of the correct names of the property categories to which the visual stimuli belong.

In the third training phase, the networks exclusively receive symbolic input consisting of simple descriptions. These contain the previously-learned property names together with a new symbol denoting the object as a whole. In this way new symbols can be defined without having a direct experience of their referent, by just describing their form and color properties. This is the process of grounding transfer. Grounding is transferred from the symbols acquired during the second phase to the new symbols defined in the propositions. The new symbols are now indirectly grounded in the network's perceptual and cognitive (categorization) experience.

During the test phase, novel images, depicting the referents of the new symbols, are projected onto the retina for the first time. This enables us to verify if grounding has effectively transferred. If the networks consistently respond with the correct symbol for every new image presented, then we can conclude that grounding transfer has taken place.

III. Results

The training and testing phases were repeated with 30 networks with different random initial weights. The learning rate was 0.2 for the first phase and 0.5 for the following two phases. The momentum was set invariably to 0.9 in all training phases. The stimuli were presented in random order during the unsupervised SOM categorization, and in sequential order during the supervised MLP phases.

All 30 networks completed the training procedure successfully. The networks acquired names for the color and shape properties of the training stimuli correctly, with a success rate of 100%. After the final training stage, 27 images depicting new objects were presented to the networks for the first time, in order to check if grounding transfer had taken place from directly grounded property names to names for the objects. The percentage of correct object naming during the grounding transfer test (i.e. producing the correct name for an image containing an object never seen before) was computed using a winner-takes-it-all approach. The unit with highest activation determines the name of the input image. The rate of correct responses for the 30 nets was 89.7%. Without having ever seen the test images before, the nets were able to categorize and name them correctly in the great majority of cases. These results are very similar to that of the previous grounding transfer models using only MLPs and the supervised learning algorithms.

IV. Discussion and Conclusions

This simulation demonstrates that grounding is transferred from symbols denoting object properties to newly acquired symbols denoting the object as a whole. In general terms, it shows that a connectionist model enables knowledge acquisition, through the combination of previously-grounded symbols, that goes beyond direct experience with the environmental stimuli.

The original aspect of this simulation is the use of a hybrid unsupervised (SOM) and supervised (MLP) model. Previously, such a hybrid, modular approach had been used to study category acquisition and labeling [17]. Schyns calls it "mapped functional modularity". His model contains an unsupervised module that categorises the stimulus set, while a supervised module connects labels to their representations. Like in our model, unsupervised concept formation precedes the supervised association of a category name. However, Schyns's model is limited to the direct grounding of basic category names. No names of higher-order categories are learned via symbolic instructions, and therefore the grounding transfer mechanism does not apply. Instead, he concentrates on prototype effects and conceptual nesting of hierarchical category structures. Symbols are only used as indicators of knowledge and facilitators of concept extraction. On the other hand, the present work builds on Schyns's [17] and our [2,8] previous models by focusing on the transfer of grounding. This can better permit the development of scaled-up connectionist models that can deal with various properties of language, such as that of generativity. In particular, future research with the current grounding transfer model will consider the use of more complex rules for combining previously-grounded symbols to generate and describe new meanings. The current simulation only uses an implicit AND logical connective to link the two symbols referring to the color and shape of the new object. New work will include the use of more complex

compositional languages. These will require the use of more logical connectives, up to the level of combining symbols through syntactic rules.

Research on computational modeling of the grounding of language has recently moved towards the integration of connectionist models with other simulation methodologies such as multi-agent models and artificial life. These techniques support an embodied and situated approach to cognitive modeling [1, 14]. For example, Honkela and Winter [10] use SOM models to control the cognitive system of agents able to perceive and act in the environment and to communicate about it. Cangelosi and colleagues [3, 12] use neural networks to control the cognitive and linguistic behavior of simulated agents and embodied robots.

The current model is being expanded and integrated into an artificial life multi-agent system. This will permit the study of the emergence of a shared grounded language and of grounding transfer in autonomous agents. Such an embodied and situated approach will enable us to study the effects of interaction in an environment and how the consequent bodily experience influences language emergence and acquisition.

REFERENCES

[1] R. A. Brooks, "Intelligence without representation," *Artificial Intelligence Journal*, vol 47, pp. 139-159, 1991.

[2] A. Cangelosi, A. Greco, and S. Harnad S, "From robotic toil to symbolic theft: grounding transfer from entry-level to higher-level categories," *Connection Science*, vol. 12(2), pp. 143-162, 2000.

[3] A. Cangelosi, and S. Harnad, "The adaptive advantage of symbolic theft over sensorimotor toil: Grounding language in perceptual categories," *Evolution of Communication*, vol. 4, pp. 117-142, 2000.

[4] A. Cangelosi, and D. Parisi, *Simulating the Evolution of Language*, London: Springer, 2000.

[5] M. G. Dyer, "Grounding language in perception," In V. Honavar, L. Uhr (Eds.), *Artificial Intelligence and Neural Networks: Steps Toward Principled Integration*. Boston: Academic Press, 1994.

[6] J. A. Fodor, *The Language of Thought*, New York: Thomas Y. Crowell, 1975.

[7] J. A. Fodor, *Psychosemantics*, Cambridge MA: MIT/Bradford, 1987.

[8] A. Greco, T. Riga, and A. Cangelosi, "The acquisition of new categories through grounded symbols: An extended connectionist model," In O. Kaynak, E. Alpaydin, E. Oja & L. Xu (Eds.), *Artificial Neural Networks and Neural Information Processing - ICANN/ICONIP 2003*, Berlin: Springer, pp. 773-770, 2003.

[9] S. Harnad, "The Symbol Grounding Problem," *Physica D*, vol. 42, pp. 335-346, 1990.

[10] T. Honkela and J. Winter, "Simulating language learning in community of agents using self-organizing maps. *Technical Report*, Helsinky University of Technology, Computer and Information Science Report A71, 2003.

[11] T. Kohonen, *Self-Organizing Maps*, Springer Series in Information Sciences, Vol 30, 1995.

[12] D. Marocco, A. Cangelosi, and S. Nolfi, "The emergence of communication in evolutionary robots," *Philosophical Transactions of the Royal Society London – A*, vol. 361, pp. 2397-2421, 2003

[13] J. L. McClelland, D. E. Rumelhart, and the PDP Research Group, *Parallel distributed processing: Explorations in the microstructure of cognition, Volume 1*, Cambridge MA: MIT/Bradford, 1986.

[14] R. Pfeifer, and C. Scheier, *Understanding Intelligence*, Cambridge MA: MIT Press, 1999.

[15] K. Plunkett, C. Sinha, M. F. Moller, and O. Strandsby, "Symbol grounding or the emergence of symbols? Vocabulary growth in children and a connectionist net", *Connection Science*, vol. 4, pp. 293-312, 1992.

[16] Z. W. Pylyshyn, *Computation and Cognition*, Cambridge MA: MIT/Bradford, 1984.

[17] P. Schyns, "A modular neural network model of concept acquisition", *Cognitive Science*, vol. 15, pp. 461-508, 1991.

[18] L. Steels, "Language games for autonomous robots," *IEEE Intelligent Systems*, vol. 16(5), pp. 16-22, 2001.

[19] P. Vogt, "Bootstrapping grounded symbols by minimal autonomous robots," *Evolution of communication*, vol. 4(1), pp. 89-118, 2000.