

## Integrating Language and Cognition: A Cognitive Robotics Approach

Angelo Cangelosi, Adaptive Behaviour and Cognition Research Group, University of Plymouth, Drake Circus, Plymouth PL4 8AA, UK ([acangelosi@plymouth.ac.uk](mailto:acangelosi@plymouth.ac.uk))

Vadim Tikhanoff, Adaptive Behaviour and Cognition Research Group, University of Plymouth, Drake Circus, Plymouth PL4 8AA, UK ([acangelosi@plymouth.ac.uk](mailto:acangelosi@plymouth.ac.uk))

José Fernando Fontanari, Instituto de Física de São Carlos, Universidade de São Paulo, 13560-970 São Carlos SP, Brazil

Emmanouil Hourdakis, Institute of Computer Science, University of Crete, Greece

**Abstract**— In this paper, we present some recent cognitive robotics studies on language and cognition integration to demonstrate how the language acquired by robotic agents can be directly grounded in action representations. These studies are characterized by the hypothesis that symbols are directly grounded into the agents' own categorical representations, while at the same time having logical (e.g. syntactic) relationships with other symbols. The two robotics studies are based on the combination of cognitive robotics with neural modeling methodologies, such as connectionist models and modeling field theory. Simulations demonstrate the efficacy of the mechanisms of action grounding of language and the symbol grounding transfer in agents that acquire a lexicon via imitation and linguistic instructions. The paper also discusses the scientific and technological implications of such an approach.

### I. INTRODUCTION

Recent advances in cognitive psychology, neuroscience and linguistics support an embodied view of cognition, i.e. the fact that cognitive functions (perception, categorization, reasoning and language) are strictly intertwined with sensorimotor and emotional processes (Wilson 2002). This is particularly evident in recent studies on the grounding of language in action and perception (Pecher & Zwann 2004). For example, in psycholinguistics, Glenberg & Kaschak (2002) have demonstrated the existence of Action-sentence Compatibility Effects. In sentence comprehension tasks, participants are faster to judge the sensibility of sentences, implying motion toward the body (e.g. "Courtney gave you the notebook") when the response requires moving toward the body (i.e. press a button nearer body). When the sentence implied movement away from the body, participants were faster to respond by literally moving away from their bodies (press a button farther from body). The data support an embodied theory of meaning that relates the meaning of sentences to human action and motor affordances. This is also consistent with neuroscientific studies on action and language, such as the involvement of the mirror neuron system for action and language learning (Rizzolatti & Arbib 1998), and brain imaging studies where words (e.g. action verbs) activate cortical areas (e.g. motor and premotor cortex) in a somatotopic fashion (Pulvermuller 1993). In linguistics, the link between the properties of language and their relationship with cognitive processes has been formalized by cognitive and constructivist linguistic theories (e.g. Talmy, 1980).

This growing empirical evidence is consistent with recent advances in artificial intelligence and robotics, where the design of the capabilities of the artificial cognitive agents is based on an integrated cognitive approach (Perlovsky, this volume). For example, the design of the linguistic capabilities of interactive systems for human-robot communication are built (grounded) onto the robot's other sensorimotor and cognitive skills (Cangelosi et al. 2005; Feldman & Narayanan 2004). Robots acquire words through direct interaction with their physical and social world, so that linguistic symbols do not

exist as arbitrary representations of some notion, but are intrinsically connected to behavioral or cognitive abilities, based on the properties of the reference system they belong to. This task of connecting the arbitrary symbols used in internal reasoning with external physical stimuli is known as Symbol Grounding (Harnad 1990).

In this paper, we will present some recent cognitive robotics studies on language and cognition integration to demonstrate how the language acquired by robotic agents can be directly grounded in action representations. These studies are characterized by the hypothesis that symbols are directly grounded into the agents' own categorical representations, while at the same time having logical (e.g. syntactic) relationships with other symbols. First, each symbol is directly grounded into internal categorical representations. These representations include perceptual categories (e.g. the concept of blue color, square shape, and male face), sensorimotor categories (e.g. the action concept of grasping, pushing, and carrying), social representations (e.g. individuals, groups and relationships) and other categorizations of the agent's own internal states (e.g. emotions and motivations). These categories are connected to the external world through our perceptual, motor and cognitive interactions with the environment. Second, symbols also have logical (e.g. syntactic) relationships with the other symbols of the lexicons used for communication. This allows symbols to be combined, using compositional rules such as grammar, to form new meanings. For example, the combination of the two symbols "stripes" and "horse," which are directly grounded into the agents' own sensorimotor experience of striped objects and horses in its environment, produces the new concept (and word) "zebra." This new symbol becomes indirectly grounded in the agents' experience of the world through the process of "symbol grounding transfer." An example of symbol grounding transfer will be demonstrated in the cognitive robotics model for the acquisition and combination of names of actions.

The two cognitive robotics models presented below will demonstrate the mechanisms of action grounding of language and the symbol grounding transfer in agents that acquire a lexicon via imitation and linguistic instructions. These models are based on the combination of cognitive robotics with neural modeling methodologies, such as connectionist models and modeling field theory.

## II. COGNITIVE ROBOTICS AND CONNECTIONIST MODELLING OF SYMBOL GROUNDING TRANSFER

Neural networks have been proposed as an ideal cognitive modeling methodology to deal with the symbol grounding problem (Harnad 1990). For example, connectionist models, such as multilayer perceptrons (MLP), permit a good implementation of the process of grounding output symbolic representations in the (analogical) input representation of external stimuli (Plunkett et al. 1992; Cangelosi 2005). The same feedforward models can be extended to simulate the process of grounding transfer (Cangelosi et al. 2000). More recently, these connectionist models have been incorporated in studies based on cognitive agents and robots. Cognitive robotics refers to the field of robotics that aims at building autonomous cognitive systems capable of performing cognitive tasks such as perception, categorization, language and sensorimotor problems. Cognitive robotics approaches include epigenetic robotics and autonomous mental development systems (Weng et al. 2001), as well as evolutionary robotics (Nolfi & Floreano 2000). Here we briefly present a cognitive robotics model for the acquisition of a lexicon of words of action and for the grounding transfer. This is an extension of the first cognitive robotics model for symbol grounding in language comprehension tasks originally developed by Cangelosi and Riga (2006). The new model presented below extends the previous study by considering both linguistic comprehension and production capabilities.

### A. *The Robot*

The robotics model consists of two simulated agents (teacher and learner) embedded within a virtual simulated environment (Figure 1). Each robot consists of two three-segment arms attached to a torso (6

Degrees of Freedom). This is further connected to a base with four wheels, which were not used in the present simulation. Through the two arms, the robot can interact with the environment and manipulate objects placed in front of it. Three objects were used in the current simulation: a cube, a horizontal plane and a vertical bar. The agent can receive in the input retina different views (perspectives) of each object. The agent has to learn six basic actions: lower right shoulder, lower left shoulder, close right upper arm, close left upper arm, close right elbow, and close left elbow. They will also learn the name of such basic actions: “LOWER\_RIGHT\_SHOULDER,” “LOWER\_LEFT\_SHOULDER,” “CLOSE\_RIGHT\_UPPER ARM,” “CLOSE\_LEFT\_UPPER ARM,” “CLOSE\_RIGHT\_ELBOW,” AND “CLOSE\_LEFT\_ELBOW.” Each action will be associated with some of the above objects that are put in front of the agent. The close left and close right shoulder actions are associated with different views of the cube.

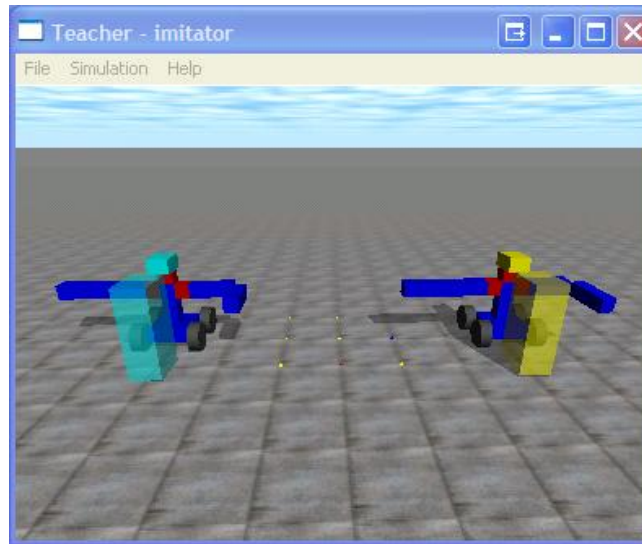


Figure 1. Simulation setup with the two robots. The teacher robot is on the left and the learner on the right. The agents are performing the close left elbow action.

This system is implemented using ODE (Open Dynamics Engine, [www.ode.org](http://www.ode.org)), an open source, high performance library for simulating rigid body dynamics. ODE is useful for simulating vehicles, objects in virtual reality environments and virtual creatures, and it is being increasingly used for simulation studies on autonomous cognitive systems.

The first agent, the teacher, is pre-programmed to perform and demonstrate a variety of basic actions, each associated with a linguistic signal. These are demonstrated to the second robot, the learner, which attempts to reproduce the actions by mimicking them. First, the agent acquires basic actions by observing the teacher. Then, it learns the basic action names (direct grounding). Subsequently, it autonomously uses the linguistic symbols that were grounded in the previous learning stage to acquire new higher-order actions (symbol grounding transfer).

#### *B. Neural network controller and training procedure*

The imitator robot is endowed with an MLP neural network (Figure 2) with input units for vision, proprioceptive and linguistic input and output units for motor control and linguistic output. For the robot motor control, the motor output units encode the force that is being applied on each joint. Each action consists of a sequence of 10 steps of motor activations.

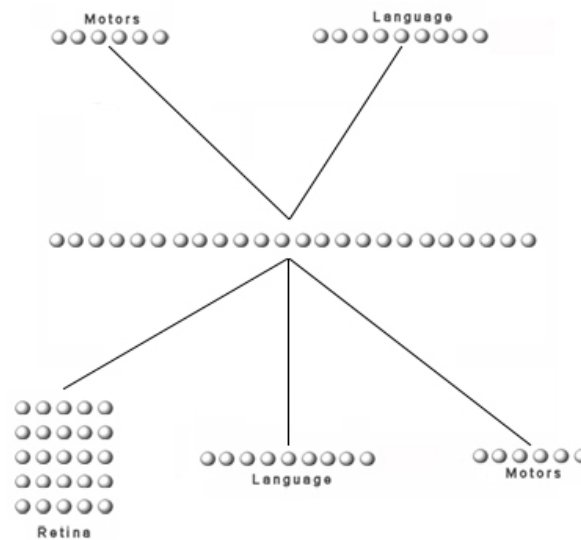


Figure 2. Architecture of the learner robot's neural network controller.

We attain the grounding transfer, using a three stage training process: (1) BA – Basic Action learning, (2) EL – Entry-Level naming and (3) HL – Higher-Order learning.

During the Basic Action learning stage, the agent learns to execute all six basic actions in association with the view of the different objects. No linguistic elements are used at this stage. The imitation algorithm is used to adjust the weights contributing to the activation of the motor units using supervised learning (see Cangelosi et al. 2006 for the learning algorithm details).

The second learning stage, Entry Level naming, was concerned with associating the previously acquired behaviors to linguistic signals. It features three sequential activation cycles. The first EL cycle, Linguistic Production, trains the learner to name the six basic actions. Motor (proprioceptive) and visual (object view) information are given in input to the network. The agents learn to correctly activate the output linguistic nodes corresponding to the basic action names. This is based on a standard backpropagation algorithm. This linguistic production cycle implements the process of basic symbol grounding, by which the names (symbols) that the agent is learning are directly grounded on its own perceptual and sensorimotor experience. In the second EL cycle, Linguistic Comprehension, learner agents are taught to correctly respond to a linguistic signal consisting of the name of the action, without having the ability to perceive the object associated to the action. To accomplish this, the retinal units in the network were set to 0, while we activate the input units corresponding to the action name. In the final EL cycle, Imitation, both motor and linguistic inputs were activated in input, and the network learns to reproduce the action in output and activate the corresponding action name unit. This third cycle is necessary to permit the linking of the production and the comprehension tasks in the hidden units activation pattern (Cangelosi et al. 2000).

The final training stage, Higher-Level learning, allows the learner agents to autonomously acquire higher-order actions without the need of a demonstration from the teacher. This is achieved only through a linguistic instruction strategy and a “mental simulation” strategy similar to Barsalou’s perceptual symbol system hypothesis (Barsalou 1999). The teacher has only to provide new linguistic instructions consisting of the names of two basic actions and the name of a new higher-order action. For example, the three higher-order actions, “LOWER\_RIGHT\_SHOULDER+LOWER\_LEFT\_SHOULDER=PLACE.”

Once the teacher (or a human instructor) provides a higher-order instruction, the learner goes through four HL learning cycles. First, it activates only the input unit of the first basic action name to produce and store (“memorize”) the corresponding sequence of 10 motor activation steps. Second, it activates in

input the linguistic units for the first basic action name and the new higher-order action. The resulting 10 motor activations are compared with the previously stored values to calculate the error and apply the backpropagation weight corrections. The next two cycles are the same as the first two, except that the second basic action name unit is activated as well.

The Higher-Order stage permits the implementation of a purely autonomous way to acquire new actions through the linguistic combination of previously-learned basic action names. The role of the teacher in this stage is only that of providing a linguistic instruction, without the need to give a perceptual demonstration of the new action. The motor imitation learning, such as in the Basic Action training stage, is a slow process based on continuous supervision, trial-and-error feedback-based learning. The acquisition of a new concept through linguistic instruction is a quicker learning mechanism because it requires the contribution of fewer units (the localist linguistic units) and corresponding weights. Moreover, in a related symbol grounding model on language (symbolic) vs. error-and-trial (sensorimotor toil) learning of categories, the linguistic procedure consistently outperforms the other learning method (Cangelosi & Harnad 2000).

To establish if the agent has actually learned the new higher-order actions and transferred the grounding from basic action names to higher order names, a test phase is performed. This grounding transfer test aims at evaluating the aptitude of the imitator agent to perform a new composite action with any of the objects previously associated, in the absence of the linguistic descriptions of the basic actions. Thus, the agent is requested to respond solely on the signal of the composite action (e.g. Grab) and selectively to the different view of the objects. In addition, while the imitator was taught only the motion of the dissected action for each composite behavior, the test evaluated the performance of the higher-order composite action. This was a behavior never seen before by the robot. The stage was comprised of two basic trials per behavior, using the different view of the objects. All inputs were propagated through the network with no training occurring.

### *C. Results*

We replicated the simulation experiment as above with five agents. Each agent had a different set of random weights initialized in the range  $\pm 1$ . The three learning stages, Basic Action, Entry-Level and Higher-Level learning, respectively, lasted for 1000, 3000, and 1500 epochs. This was the approximate minimum number of epochs necessary to reach a good learning performance. The parameters of the backpropagation algorithm were set as follows: BA stage, momentum 0.6 and learning rate 0.2; EL stage, momentum 0.6 and learning rate 0.3; HL stage, momentum 0.8 and learning rate 0.2. The weights were updated at the end of every action.

Overall, results indicate that all agents were able to learn successfully the six basic actions and the three higher-order behaviors. At the end of the stage, the imitator was able to execute all actions flawlessly when presented with an object (final error of 0.004). The overall average error on the final epoch of the Entry-Level stage was 0.03. Finally, in the grounding transfer test, the agent was requested to perform a new composite action by giving input of only the new action name or the new name together with the basic action names (error 0.018). These results confirm our hypothesis that previously grounded symbols are transferred to the new behaviors.

## III. ACTION AND LEXICON SCALING WITH MODELING FIELD THEORY

In this study, we aim at extending the behavioral and linguistic capabilities of the robot by scaling its action repertoire. Perlovsky (2001; this volume) has recently proposed the use of the Modeling Field Theory learning algorithm to deal with the issue of the combinatorial complexity (CC) of linguistic and cognitive modeling based on machine learning techniques, such as multilayer perceptrons. The Modeling Field Theory algorithm uses dynamic logic to avoid CC and computes similarity measures

between internal concept-models and the perceptual and linguistic signals. By using concept-models with multiple sensorimotor modalities, an MFT system can integrate language-specific signals with other internal cognitive representations. Perlovsky's proposal to apply MFT in the language domain is highly consistent with the grounded approach to language modeling discussed above. That is, both accounts are based on the strict integration of language and cognition. This permits the design of cognitive systems that are truly able to "understand" the meaning of words being used by autonomously linking the linguistic signals to the internal concept-models of the word constructed during the sensorimotor interaction with the environment. The combination of MFT systems with grounded agent simulations will permit the overcoming of the CC problems currently faced in grounded agent models, and scale the lexicons in terms of high numbers of lexical entries and syntactic categories.

Modeling Field Theory is based on the principle of associating lower-level signals (e.g., inputs, bottom-up signals) with higher-level concept-models (e.g. internal representations, categories/concepts, top-down signals) avoiding the combinatorial complexity inherent to such a task. This is achieved by using measures of similarity between concept-models and input signals together with a new type of logic, so-called dynamic logic. MFT may be viewed as an unsupervised learning algorithm, whereby a series of concept-models adapt to the features of the input stimuli via gradual adjustment dependent on the fuzzy similarity measures.

#### A. Extended action and lexicon repertoire

The robotic scenario is based on the same simulated robotic agents described in the previous section (see Figure 1). The teacher robot is pre-programmed to demonstrate an extended action repertoire of 112 actions. The learner robot uses MFT to learn to reproduce those actions as well as to learn the action names.

The main difference with respect to the previous model is the use of 112 different actions. These are inspired by the semaphore flag, signaling the alphabet. For the encoding of the actions, we collected data on the posture of the teacher robots using six features, i.e. three pairs of angles for the two joints of the shoulder, upper arm and elbow joints. In this simulation, objects are not present.

When performing the action, the teacher agent can emit a three-letter word labeling the action. Each label consists of a Consonant-Vowel-Consonant word, such as "XUM," "HAW," or "RIV." All consonants and letters of the English alphabet are used. Each letter is encoded using two real value features in the interval [0,1]. Therefore, each action word is represented by six features. Each word is unique to the action performed.

#### B. MFT algorithm

We use a multidimensional MFT algorithm (Tikhanoff et al. 2006) with 112 input fields (concept-models) randomly initialized. We consider the action learning problem as that of categorizing  $N=112$  objects (actions)  $i=1, \dots, N$ , each of which is characterized by  $d=12$  features  $e=1, \dots, d$ . These features are represented by real numbers  $O_{ie} \in (0,1)$  – the input signals. These 12 features correspond to the six joint rotation angles and six phonetic encoding values. Moreover, we assume that there are  $M=112$   $d$ -dimensional fields (i.e. concept-models of the prototype of actions/words to be learned)  $k=1, \dots, M$  described by real valued fields  $S_{ke}$ , with  $e=1, \dots, d$  as before. The concept models will tend to match the input object features  $O_{ie}$  during learning by maximizing the global similarity function

$$L = \sum_i \log \sum_k l(i | k)$$

where

$$l(i | k) = \prod_{e=1}^d (2\pi\sigma_{ke}^2)^{-1/2} \exp[-(S_{ke} - O_{ie})^2 / 2\sigma_{ke}^2]$$

is the similarity measure between object  $i$  and concept  $k$ . Here,  $\sigma_{ke}$  is the fuzziness parameter that gradually decreases over time. Full details on the learning algorithm can be found in Tikhanoff et al. 2007. See also Perlovsky (this volume) for an overview of the MFT algorithm.

### C. Results

The simulation lasts for 25000 training steps. In the first 12500 cycles, only the six action features (angles) are provided. This is enough for the agents to learn to reproduce the action repertoire. At cycle 12500 (half of the training time), all 12 feature sets (six for actions/angles, six for phonetic sounds) are considered when computing the MFT fuzzy similarity functions. The re-initialization of the fuzziness parameter  $\sigma_{ke}$  at timestep 125000 allows the agent to learn the new sound features and create a concept model of the labels.

Results demonstrate that the robot is able to categorize 95% of actions and learn their unique labels. Figure 3 shows the evolution of the 112 concept-model fields during training. Note the resetting of the fields at timestep 12500, when words are introduced and the fuzziness  $\sigma_{ke}$  is reinitialized.

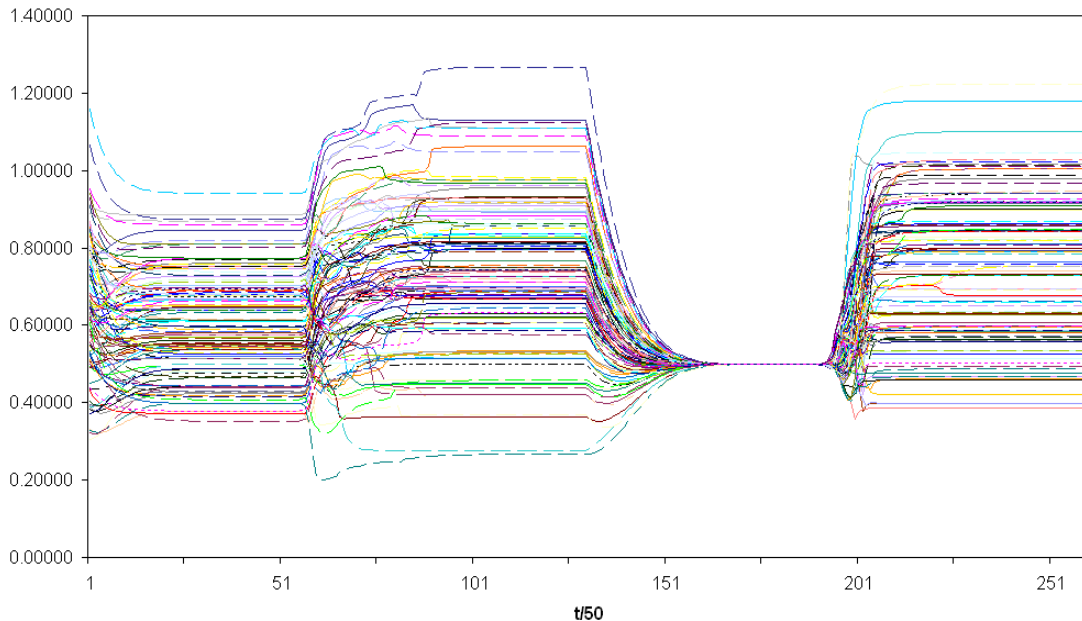


Figure 3 – Evolution of 112 concept-models during training. Vertical axis represents a compressed one-dimensional

representation of the concept-models using the amplitude  $S_k = \sqrt{\sum_{e=1}^d (S_{ke})^2} / d$ .

## IV. DISCUSSION AND CONCLUSIONS

The simulation experiments above concern the study of the language grounding in action and the symbol grounding transfer in cognitive robotic agents. The positive results of the grounding transfer simulation demonstrate that it is possible to design autonomous linguistic agents capable of acquiring new grounded concepts.

The use of MFT to overcome the CC limitations of connectionist models demonstrates that it is possible to scale the action and lexicon repertoire of the cognitive robotic agents. Perlovsky's (2004) proposal to apply MFT in the language domain is highly consistent with the grounded approach to language modeling discussed above. That is, both accounts are based on the strict integration of language and cognition. This permits the design of cognitive systems that are truly able to "understand" the meaning of words being used by autonomously linking the linguistic signals to the internal concept-

models of the word constructed during the sensorimotor interaction with the environment.

The potential impact of this grounded cognitive robotic approach for the development of intelligent systems is great, both for cognitive science and for technology. In cognitive science, the area of embodied cognition regards the study of the functioning and organization of cognition in natural and artificial systems. For example, the Higher-Order learning procedure is inspired by Barsalou's "reenactment" and "mental simulation" mechanism in the perceptual symbol system hypothesis. Barsalou (1997) demonstrates that during perceptual experience, association areas in the brain capture bottom-up patterns of activation in sensory-motor areas. Later, in a top-down manner, association areas partially reactivate sensory-motor areas to implement perceptual symbols simulators. A simulation platform like the one used here can be used to test further embodied cognition theories of language, such as Glenberg and Kaschak's (2002) action-compatibility effects. In addition, such an approach can be used to study the development and emergence of language in epigenetic robots (Weng et al. 2001; Metta et al. 2006).

For the technological implications of such a project, the model proposed here can be useful in fields such as that of defense systems, service robotics and human-robot interaction. In the area of defense systems, cognitive systems are essential for integrated multiplatform systems capable of sensing and communicating. Such robots can be beneficial in collaborative and distributed tasks, such as multiagent exploration and navigation in unknown terrains. In service and household robotics, future systems will be able to learn language and world understanding from humans, and also to interact with them for entertainment purposes (e.g. Tikhanoff & Miranda, 2005; Steels & Kaplan 2000). In human-robot communication systems, robots will develop their lexicon through close interaction with their environment while communicating with humans. Such a social learning context can permit a more efficient acquisition of communication capabilities in autonomous robots, as demonstrated in Steels & Kaplan (2000).

#### REFERENCES

- Barsalou L. (1999). Perceptual symbol systems. *Behavioral and Brain Sciences*, 22, 577-609.
- Cangelosi (2005). Approaches to Grounding Symbols in Perceptual and Sensorimotor Categories. In H. Cohen & C. Lefebvre (Eds), *Handbook of Categorization in Cognitive Science*, Elsevier, pp. 719-737
- Cangelosi A. & Harnad S. (2000). The adaptive advantage of symbolic theft over sensorimotor toil: Grounding language in perceptual categories. *Evolution of Communication* 4(1), 117-142
- Cangelosi A., G. Bugmann and R. Borisyuk (Eds.), *Modeling Language, Cognition and Action: Proceedings of the 9th Neural Computation and Psychology Workshop*. Singapore: World Scientific, 2005
- Cangelosi A., Greco A. & Harnad S. (2000). From robotic toil to symbolic theft: Grounding transfer from entry-level to higher-level categories. *Connection Science*, 12(2), 143-162
- Cangelosi A., Hourdakis E. & Tikhanoff V. (2006). Language acquisition and symbol grounding transfer with neural networks and cognitive robots. *Proceedings of IJCNN 2006*, Vancouver
- Cangelosi, A. and T. Riga, "An embodied model for sensorimotor grounding and grounding transfer: Experiments with epigenetic robots". *Cognitive science*, in press 2006, 30(4), 673-689
- Feldman, J., & Narayanan S. (2004). Embodied meaning in a neural theory of language. *Brain and Language*, 89, 385-392.
- Glenberg A., and K. Kaschak, "Grounding language in action," *Psychonomic Bulletin & Review*, vol. 9(3), pp. 558-565, 2002.
- Harnad, S. "The Symbol Grounding Problem," *Physica D*, vol. 42, pp. 335-346, 1990.
- L. Perlovsky, L. "Integrating language and cognition," *IEEE Connections*, vol. 2(2), pp. 8-13, 2004

- Metta G., Fitzpatrick P., Natale L. (2006). YARP: yet another robot platform. *International Journal on Advanced Robotics Systems*, 3, 43-48.
- Nolfi S. & Floreano D. (2000). *Evolutionary Robotics: The Biology, Intelligence, and Technology of Self-Organizing Machines*. Cambridge, MA: MIT Press/Bradford Books.
- Pecher, D., & Zwaan, R.A., (Eds.). (2005). *Grounding cognition: The role of perception and action in memory, language, and thinking*. Cambridge, UK: Cambridge University Press.
- Perlovsky L. (2001). *Neural Networks and Intellect: Using Model-Based Concepts*. Oxford University Press, New York.
- Plunkett K., Sinha C., Møller M.F., Strandsby O. (1992). Symbol grounding or the emergence of symbols? Vocabulary growth in children and a connectionist net. *Connection Science*, 4: 293-312
- Pulvermuller F. (2003). *The Neuroscience of Language. On Brain Circuits of Words and Serial Order*. Cambridge University Press.
- Rizzolatti G. & Arbib M.A. (1998). Language within our grasp. *Trends in Neurosciences*, 21(5), 188-194.
- Steels L., and K. Kaplan, "AIBO's first words: The social learning of language and meaning," *Evolution of Communication*, vol. 4(1), pp. 3-32, 2000.
- Talmy L. (2000). *Toward a Cognitive Semantics*, Vol. I. Cambridge: Cambridge University Press
- Tikhanoff V., and E.R. Miranda, "Musical Composition by an Autonomous Robot: An Approach to AIBO Interaction", in *Proceedings of TAROS 2005 (Towards Autonomous Robotic System)*, Imperial College, London, UK. 2005, pp. 181-188.
- Tikhanoff V., Cangelosi A., Fontanari J.F. & Perlovsky L.I (2007). Scaling up of action repertoire in linguistic cognitive agents. *Proceedings IEEE-KIMAS'07 Conference*. Waltham, MA
- Tikhanoff V., Fontanari J.F., Cangelosi A. & Perlovsky L.I. (2006). Language and cognition integration through modeling field theory: Category formation for symbol grounding. *ICANN06 International Conference on Artificial Neural Networks*, Athens, September 2006.
- Weng J., McClelland J., Pentland A., Sporns O., Stockman I., Sur M. & Thelen E. (2001). Autonomous mental development by robots and animals. *Science*, 291, 599-600.
- Wilson, M. (2002). Six views of embodied cognition. *Psychonomic Bulletin and Review*, 9, 625-636.