Editorial of e-book on action and language integration

Angelo Cangelosi*

Centre for Robotics and Neural Systems, Plymouth University, Plymouth, UK
Correspondence: a.cangelosi@plymouth.ac.uk

Increasing theoretical and experimental research on action and language processing in humans and animals clearly demonstrates the strict interaction and co-dependence between language and action. This has been extensively demonstrated in neuroscientific investigations (e.g., Rizzolatti and Arbib, 1998; Cappa and Perani, 2003; Pulvermuller, 2003), psychology experiments (e.g., Glenberg and Kaschak, 2002; Pecher and Zwaan, 2005; Barsalou, 2008), evolutionary psychology (e.g., Corballis, 2002), and computational modeling (e.g., Cangelosi and Parisi, 2004; Massera et al., 2007; Cangelosi, 2010). All these studies have important implication both for the understanding of the action basis of cognition in natural and artificial cognitive systems, as well as for the design of cognitive and communicative capabilities in robots (Cangelosi et al., 2010).

The journal "Frontiers in Neurorobotics" published a collection of articles on the topic of action and language integration both in natural cognitive systems (e.g., humans and animals) and in artificial cognitive agents (robots and simulated agents). These articles are now collected in an e-book, for wider dissemination. This set of chapters provides an up to date overview of current advances in the grounding of language into sensorimotor knowledge. The first chapters primarily focus on experimental evidence from cognitive psychology (Symes et al., 2010), cognitive neuroscience studies (Borghi et al., 2010), and comparative experimental/simulation studies (Greco and Caneva, 2010). Two chapters then use neural network simulation for motor chains for sentence processing (Chersi et al., 2010) and a computational model of gaze planning in word recognition and reading (Ferro et al., 2010). Finally, four chapters use cognitive systems and robotics methodologies to investigate general principles of action–language grounding (Parisi, 2010), teleological representations of action and language for human–robot interaction experiments (Lallee et al., 2010), verbal and non-verbal communication in neurorobotics models (Bicho et al., 2010), and action bases of action words (Marocco et al., 2010).

EXPERIMENTAL STUDIES

Borghi et al. (2010) focus on language comprehension as an embodied simulation of actions. This hypothesis is supported by embodied and grounded cognition theories (Barsalou, 2008; Pezzulo et al., 2011) and the neural underpinnings in neural substrates involve canonical and mirror neurons (Rizzolatti et al., 1996). Borghi et al. review their recent behavioral and kinematic studies to characterize, and evidence, the relationship between language and the motor system. This review leads to three consistent findings: (i) the simulation evoked during sentence comprehension is fine-grained, and shows sensitivity to the different effectors used to perform actions; (ii) linguistic comprehension also relies on the representation of actions in terms of goals and of the chains of motor acts necessary to accomplish them; and (iii) the goals are modulated by both the object features the sentence refers to, as well as by social aspects such as the characteristics of the agents implied by sentences. The authors also explicitly discuss the implications of these studies for embodied robotics.

Symes and colleagues present a cognitive psychology study on the integrating action and language through biased competition. This is based on previous psychological investigations that have demonstrated that planning an action biases visual processing, as in Symes et al.'s (2008) findings reporting faster target detection for a changing object amongst several non-changing objects. This new experimental study investigates how this effect might compare to, and indeed integrate with, effects of language cues. Using the same change-detection scenes as in Symes et al. (2008), two effective sources of bias are identified: (i) action primes, and (ii) language cues. For example, a sentence as "Start looking for a change in the larger objects" cues object size, and these successfully enhanced detection of size-congruent targets. Additional experiments explore the biases' co-occurrence within the same task, such as action prime (participants plan a power or precision grasp) *and* a language (a sentence) cue preceding stimulus presentation. Experimental results support the authors' predictions from the biased competition model by Desimone and Duncan (1995), in particular reliably stronger effects of language, and concurrent biasing effects that were mutually suppressive and additive.

Greco and Caneva (2010) focus on compositional symbol grounding for motor patterns. They propose a new comparative experimental/simulative paradigm to study the learning of compositional grounded representations for motor patterns. In a psychology experiment, participants learn to associate non-sense arm motor patterns, performed in three different hand postures, with non-sense words. Two experimental conditions are carried out: (i) in the *compositional* condition, each pattern was associated with a two-word (verb–adverb) sentence; (ii) in the *holistic* condition, each pattern was associated with a unique word. Experimental results show that the compositional group achieved better results in naming motor patterns, especially for patterns where hand postures discrimination was relevant. In order to ascertain the differential effects of memory load and of systematic grounding, neural network simulations were also carried out. After a basic simulation reproducing the default participants' performance, in some simulations the number of stimuli (motor patterns and words) was increased and the systematic association between words and patterns was disrupted, while keeping the same number of words and compositionality. Simulation results show that in both conditions the advantage for the compositional condition significantly increased. This indicates that the advantage for the compositional condition may be related to systematicity rather than to mere informational gain. Overall, both experimental and simulation data support the hypothesis of a shared action/language compositional motor representation.

NEURAL NETWORK STUDIES

Chersi et al. (2010) investigate the relationship of language to motor chains for sentence processing. As in Borghi et al. (2010), they also start from embodied theories of language grounding in the sensorimotor system, and language understanding as a process based on a mental simulation process (Jeannerod, 2007; Gallese, 2008; Barsalou, 2009). This hypothesizes that during action words and sentence comprehension the same perception, action, and emotion mechanisms implied during interaction with objects are recruited. Their aim is to identify the precise dynamics underlying the relation between language and action, e.g., to disentangle experimental evidence reporting both either facilitation or interference effects between language processing and action execution. This chapter presents a new neural network reproducing experimental data on the influence of action-related sentence processing on the execution of motor sequences. Chersi et al.'s modeling framework is based on three main principles: (i) the processing of action-related sentences causes the resonance of motor and mirror neurons encoding the corresponding actions; (ii) a varying degree of crosstalk exists between neuronal populations depending on whether they encode the same motor act, the same effector, or the same action-goal; (iii) neuronal populations' internal dynamics, which results from the combination of multiple processes taking place at different time scales, can facilitate or interfere with successive activations of the same or of partially overlapping pools. Interactions between sensory and motor modalities are modeled as a crosstalk between neuronal pools in motor and mirror chains. Results show also that the neural dynamics governing the activation of the pools can qualitatively reproduce the timings observed in behavioral experiments.

Ferro et al. (2010) propose a computational model of gaze planning in word recognition And the theory that reading is an active sensing process. Their computational model of gaze planning during reading consists of two main components: (i) a *lexical representation network*, acquiring lexical representations from input texts from the Italian CHILDES database; (ii) a *gaze planner* capable to recognize written words by mapping strings of characters onto lexical representations. Thus the model implements an active sensing strategy that selects which characters of the input string are to be fixated, depending on the predictions dynamically made by the lexical representation network. The analyses investigate the developmental trajectory of the system in performing the word recognition task as a function of both increasing lexical competence, and correspondingly increasing lexical prediction ability.

NEUROROBOTICS STUDIES

Parisi (2010) discusses a general neural modeling approach to language grounding in robots, consistent with the same literature on embodiment and grounding theories. The paper proposes a neural model of language according to which the robot's behavior is controlled by a neural network composed of two sub-networks: (i) the network controlling non-linguistic interaction between the robot and its environment; and (ii) a network for the processing of linguistic comprehension and production. Parisi reviews results of a number of computational simulations and suggests that the model can be extended to account for variety of language-related phenomena such as disambiguation, the metaphorical use of words, the pervasive idiomaticity of multi-word expressions, and mental

life as talking to oneself. This modeling approach implies a view of the meaning of words and multi-word expressions as a temporal process that takes place in the entire brain and has no clearly defined boundaries. This can be further extended to emotional words, considering that an embodied view of language should consider not only the interactions of the robot's brain with the external environment, but also the interactions of the brain with what is inside the body such as motivational and emotional processes.

Lallee et al. (2010) link embodied and teleological representations of action and language for humanoid robotic experiments with the iCub platform. In this chapter the authors extend their framework for embodied language and action comprehension to include a teleological representation of goal-based reasoning for novel actions. Both from a theoretical perspective, and via human–robot interaction experiments with the iCub robot, they demonstrate the advantages of this hybrid, embodied–teleological approach to action–language interaction. Lallee et al. first demonstrate how embodied language comprehension allows the system to develop a set of representations for processing goal-directed actions such as "take," "cover," and "give." A crucial component of the new approach is the representation of the subcomponents of these actions, which includes state–action–state (SAS) relations between initial enabling states, and final resulting states for these actions. Robotic experiments demonstrate how grammatical categories including causal connectives (e.g., because, if–then) can allow spoken language to enrich the learned set of SAS representations. The study also examines how this enriched SAS repertoire enhances the iCub's ability to represent perceived actions in which the environment inhibits goal achievement.

Bicho et al. (2010) employ a dynamic neural field architecture for human–robot interaction and the integration of verbal and nonverbal communication. Specifically they investigate how a group of people coordinate their intentions, goals, and motor behaviors whilts performing joint action tasks. Their model is inspired by experimental evidence about the resonance processes in the observer's motor system, and their involvement in our ability to understand actions of others and to infer their. Bicho et al. develop a control architecture for human–robot collaboration that exploits perception–action linkage as a means to achieve more natural and efficient communication grounded in sensorimotor experiences. The architecture consists of a coupled system of dynamic neural fields. These represent a distributed network of neural populations that encode in their activation patterns goals, actions, and shared task knowledge. Human–robot experiments consist of verbal and non-verbal communication for a joint assembly task in which the human–robot pair has to construct toy objects from their components. This dynamic neural field architecture sustain the robot's capacity to anticipate the user's needs and goals and to detect and communicate unexpected events that may occur during joint task execution.

Marocco et al. (2010) presents new experiments with a simulated model of the humanoid robot iCub (Tikhanoff et al., 2011) to investigate the embodied representation of action words. The simulated iCub robot is trained to learn the meaning of action words (i.e., words that represent dynamical events that happen in time) such as "push," "hit." The words are learned by physically interacting with the environment and linking the robot's effects of its own

actions (proprioception) with the behavior observed on the objects, before and after the action. The control system of the robot is an artificial neural network trained to manipulate an object through a Back-Propagation-Through-Time algorithm. Results show that the robot is able to extract the sensorimotor contingency of a particular interaction with an object and to reproduce its dynamics by acting on the environment. Moreover, in the absence of linguistic input, the robot is capable of associating a certain temporal sensorimotor dynamics to the learnt action words.

CONCLUSION

The collection of chapters in this volume provides a variety of methodological approaches to the experimental investigation and the neural network and cognitive robotic modeling of action and language integration. The studies address different phenomena linked to language grounding, such as sentence processing and comprehension, reading and word recognition, action word learning, compositionality of action and language representations, and language acquisition through interaction with the environment. All studies offer further support the existing evidence and theoretical stances of the grounding of language in action and perception, and the contribution of embodied cognition and mental simulation in language processing. Moreover, the multi-methodological contributions proposed in the volume and the close link between experimental data and computational and robotic modeling allows the fine investigation of behavioral, cognitive, and embodiment factors in the grounding of language in sensorimotor knowledge.

REFERENCES

Barsalou, L. W. (2008). Grounded cognition. *Annu. Rev. Psychol.* 59, 617–645.

Barsalou, L. W. (2009). Simulation, situated conceptualization, and prediction. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 364, 1281–1289.

Bicho, E., Louro, L., and Erlhagen, W. (2010). Integrating verbal and non-verbal communication in a dynamic neural field architecture for human–robot interaction. *Front. Neurorobot.* 4:5. doi: 10.3389/fnbot.2010.00005

Borghi, A. M., Gianelli, C., and Scorolli, C. (2010). Sentence comprehension: effectors and goals, self and others. An overview of experiments and implications for robotics. *Front. Neurorobot.* 4:3. doi: 10.3389/fnbot.2010.00003

Cangelosi, A. (2010). Grounding language in action and perception: from cognitive agents to humanoid robots. *Phys. Life Rev.* 7, 139–151.

Cangelosi, A., Metta, G., Sagerer, G., Nolfi, S., Nehaniv, C. L., Fischer, K., Tani, J., Belpaeme, B., Sandini, G., Fadiga, L., Wrede, B., Rohlfing, K., Tuci, E., Dautenhahn, K., Saunders, J., and Zeschel, A. (2010). Integration of action and language knowledge: a roadmap for developmental robotics. *IEEE Trans. Auton. Ment. Dev.* 2, 167–195.

Cangelosi, A., and Parisi, D. (2004). The processing of verbs and nouns in neural networks: insights from synthetic brain imaging. *Brain Lang.* 89, 401–408.

Cappa, S. F., and Perani, D. (2003). The neural correlates of noun and verb processing. *J. Neurolinguistics* 16, 183–189.

Chersi, F., Thill, S., Ziemke, T., and Borghi, A. M. (2010). Sentence processing: linking language to motor chains. *Front. Neurorobot.* 4:4. doi: 10.3389/fnbot.2010.00004

Corballis, M. C. (2002). *From Hand to Mouth: The Origins of Language.* Princeton, NW: Princeton University Press.

Desimone, R., and Duncan, J. (1995). Neural mechanisms of selective visual attention. *Annu. Rev. Neurosci.* 18, 193–222.

Ferro, M., Ognibene, D., Pezzulo, G., and Pirrelli, V. (2010). Reading as active sensing: a computational model of gaze planning in word recognition. *Front. Neurorobot.* 4:6. doi: 10.3389/fnbot.2010.00006

Gallese, V. (2008). Mirror neurons and the social nature of language: the neural exploitation hypothesis. *Soc. Neurosci.* 3, 317–333.

Glenberg, A., and Kaschak, K. (2002). Grounding language in action. *Psychon. Bull. Rev.* 9, 558–565.

Greco, A., and Caneva, C. (2010). Compositional symbol grounding for motor patterns. *Front. Neurorobot.* 4:111. doi:10.3389/fnbot.2010.00111

Jeannerod, M. (2007). *Motor Cognition: What Actions Tell the Self.* Oxford: Oxford University Press.

Lallee, S., Madden, C., Hoen, M., and Dominey, P. F. (2010). Linking language with embodied and teleological representations of action for humanoid cognition. *Front. Neurorobot.* 4:8. doi: 10.3389/fnbot.2010.00008

Marocco, D., Cangelosi, A., Fischer, K., and Belpaeme, T. (2010). Grounding action words in the sensorimotor interaction with the world: experiments with a simulated iCub humanoid robot. *Front. Neurorobot.* 4:7. doi: 10.3389/fnbot.2010.00007

Massera, G., Cangelosi, A., and Nolfi, S. (2007). Evolution of prehension ability in an anthropomorphic neurorobotic arm. *Front. Neurorobot.* 1:4. doi: 10.3389/neuro.12.004.2007

Parisi, D. (2010). Robots with language. *Front. Neurorobot.* 4:10. doi: 10.3389/fnbot.2010.00010

Pecher, D., and Zwaan, R. A. (eds). (2005). *Grounding Cognition: The Role of Perception and Action in Memory, Language, and Thinking.* Cambridge: Cambridge University Press.

Pezzulo, G., Barsalou, L. W., Cangelosi, A., Fischer, M. H., McRae, K., and Spivey, M. J. (2011). The mechanics of embodiment: a dialog on embodiment and computational modelling. *Front. Psychol.* 2:5. doi: 10.3389/fpsyg.2011.00005

Pulvermuller, F. (2003). *The Neuroscience of Language: On Brain Circuits of Words and Serial Order.* Cambridge: Cambridge University Press.

Rizzolatti, G., and Arbib, M. A. (1998). Language within our grasp. *Trends Neurosci.* 21, 188–194.

Rizzolatti, G., Fadiga, L., Gallese, V., and Fogassi, L. (1996). Premotor cortex and the recognition of motor actions. *Brain Res. Cogn. Brain Res.* 3, 131–141.

Symes, E., Tucker, M., Ellis, R., Vainio, L., and Ottoboni, G. (2008). Grasp preparation improves change-detection for congruent objects. *J. Exp. Psychol. Hum. Percept. Perform.* 34, 854–871.

Symes, E., Tucker, M., and Ottoboni, G. (2010). Integrating action and language through biased competition. *Front. Neurorobot.* 4:9. doi: 10.3389/fnbot.2010.00009

Tikhanoff, V., Cangelosi, A., and Metta, G. (2011). Language understanding in humanoid robots: iCub simulation experiments. *IEEE Trans. Auton. Ment. Dev.* 3, 17–29.