

Integration of Action and Language Knowledge: Key Challenges for Developmental Robotics Research

Angelo Cangelosi, Tony Belpaeme, Davide Marocco

Adaptive Behaviour and Cognition Research Group
University of Plymouth
italkproject.org

1. Introduction

Recent 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 (among others Cappa & Perani, 2003; Glenberg & Kaschak, 2002; Pulvermuller et al., 2003; Rizzolatti & Arbib, 1998). For example, neuroscientific studies of the mirror neurons system (Fadiga, Fogassi, Gallese, & Rizzolatti, 2000; Gallese, Fadiga, Fogassi, & Rizzolatti, 1996) and brain imaging studies on language processing provide an abundance of evidence for intertwined language-action integration. Hauk et al. (2004) used fMRI to show that action words referring to face, arm or leg actions (e.g. to lick, pick, or kick) differentially activate areas along the motor cortex that either were directly adjacent to or overlapped with areas activated by actual movement of the tongue, fingers, or feet. This demonstrates that the referential meaning of action words has a correlate in the somatotopic activation of the motor and premotor cortex. Neuroscientific evidence supports a dynamic view of language according to which lexical and grammatical structures of language are processed by distributed neuronal assemblies with cortical topographies that reflect lexical semantics (Pulvermuller et al., 2003). The mastery of fine motor control, such as non-repetitive action sequences involved in making complex tools, is also seen as a precursor of Broca's area in the modern brain, which is adjacent to the area that governs fine motor control in the hand. This is consistent with Rizzolatti & Arbib's (1998) hypothesis that area F5 of the monkey's brain, where mirror neurons for manual motor activity have been identified, is a precursor of Broca's area involved in language processing and speech production and comprehension.

In addition, developmental psychology studies based on emergentist and constructivist approaches (e.g. Bowerman & Levinson, 2001; Macwhinney, 2005; Tomasello, 2003) support a view of cognitive development strongly dependent on the contribution of various cognitive capabilities. They demonstrate the gradual emergence of linguistic constructs built through the child's experience with her social and physical environment. This is consistent with cognitive linguistics (cf. Lakoff, 1987; Langacker, 1987) where syntactic symbols and functions are constructed by reference to other cognitive representations.

All these studies on action-language integration have important implication for the design of communication and linguistic capabilities in cognitive systems and robots (Cangelosi et al. 2005). Amongst the various approaches to design communication capabilities in interactive agents, some provide a more integrative vision of language and treat it as an integral part of the whole cognitive system (Cangelosi & Harnad 2000). The agent's linguistic abilities are strictly dependent on, and grounded in, other behaviours and skills. Such a strict action-

language interaction supports the bootstrapping of the agent's cognitive system, e.g. through the transfer of properties of action knowledge to that of linguistic representations (and vice versa).

Below we discuss some of the key research challenges on language and action integration and on the associated representational issues for mental categories. These issues constitute some of the main challenges for research in cognitive robotics.

2. Research Issues

2.1 Hierarchical and Compositional Actions

Research on action development in the ITALK project will focus directly on the acquisition of hierarchical, compositional actions. The typical experimental scenario will involve robotic agents that use proprioceptive and visual information to actively explore the environment. This will allow agents to build embodied sensorimotor categories of object-body interactions. Tasks will include manipulation actions such as touch/move/modify objects. In addition, more advanced experiments will look at action patterns based on combination and sequences of movements. For example, simulations will consider tasks in which the robot agent learns to use a tool (e.g. "stick") to push an object. Other tasks might include a cascade of inter-dependent actions, such as making a composite tool (e.g. combine a stick with a cuboid object – as with the handle and head of a "hammer") and using this tool on a third object (e.g. to crack open a spherical object – "nut"). Tasks will be inspired by object manipulation and tool making/use observed abilities in primates and humanoids, and their relationship with the development of linguistic capabilities (e.g. Corballis 2003; Greenfield 1991). A possible starting point would be to attempt object manipulation in order to get an agent to relate one object with another in a particular combination, as a young infant would (Tanaka & Tanaka 1982). In conjunction with the research undertaken by Hayashi and Matsuzawa (2003) on the development of spontaneous object manipulation in apes and children, we are planning to include the use of language in order to carry out a series of simulations to investigate the following possible tasks: (1) Inserting objects into corresponding holes in a box; (2) Serializing nested cups; (3) Inserting variously shaped objects into corresponding holes; (4) Stacking up wooden blocks. A first instance of the experiments could be to isolate the agent from the human, as to let it calibrate its joints and hand eye coordination, recognizing colour, form/shapes and moving objects. The second part would be to introduce the agent to a "face to face" situation where a user would use linguistic instructions in order to expand the object "knowledge acquisition", taking the form of some kind of symbolic play. A third part would be to let the agent watch the human performance and then freely demonstrating the autonomous learning process either via gestures or language.

2.2 Transition from single-word lexicons to compositional languages

One important issue in language development is that of the transition from single-word lexicons to multi-word utterances. Steels (2005) has recently proposed a model for the emergence of language and grammaticalisation based on sequential stages. We plan to use such a theoretical model of language emergence stages for its value in providing a clear operational definition of qualitative changes in language development that can be easily tested in robotic experiments. Robotic experiments will specifically address the emergence of proto-linguistic categories such as names of objects, of action, and of properties of objects, and their dependence from sensorimotor knowledge and representations. It will also address the issue of the grounding of proto-function words, such as spatial terms (e.g. in, on, over). The experimental plan will follow Steels' (2005) grammaticalisation stage III on simple compositionality and subsequently stage IV on situation-specific grammar constructions for

multiple object and predicate meanings. For example, some experiments will look at constructions involving Multiple Objects + Predicates. These refer to syntactic patterns such as Subject + Predicate + DirectObject + PrepObject (e.g. “Robot puts stick on cube”) with a semantic frame of the type TRANSFER_TO_TARGET + Agent + Patient + Target (Goldberg 1995, 2006; see also Steels 2005 for a discussion on computational/robotic modelling of such constructs). In this stage agents need to be able to recognise/reproduce the syntactic patterns that are used in a particular context and apply them to situations with similar meaning patterns (Steels 2005). In the experiments agents will be capable to construct sequential patterns that link a set of predicate-argument structures and are able to combine patterns by an appropriate mapping of the variables. These grammaticalisation experiments might involve the use of recurrent neural network architectures (e.g. Batali 2002 and Sugita & Tani 2005). Experiments will also include the manipulation of the neural network architecture to test the effects of different modular topologies on different levels of complexity of the grammatical structures (cf. Deacon 1997; Elman 1990). Depending on the quality and robustness of the results achieved in Stage III and IV experiments, some additional simulations will look at the further stages of language development, such as Steels’ stage V on meta-grammar competence. This is the stage where agents move from the *ad hoc* use of syntactic patterns to more systematicity and the development of abstract syntactic categories (e.g. noun, verb, nominative, past) and abstract semantic categories (e.g. agent, beneficiary, source, cause-transfer).

2.3 Evolutionary origins of action and language compositionality.

The relationship between language and action is particularly important when we consider the striking similarities and parallels that have been demonstrated to exist between the linguistic structure and the organisation of action knowledge. As previously discussed (e.g. theme ACTION), action knowledge can be organized into compositional and hierarchical components. Language has two core characteristics: Compositionality and Recursion. Compositionality refers to the fact that a series of basic linguistic components (i.e. word categories such as nouns, verbs, adjectives etc.) can be combined together to construct meaningful sentences. Recursion refers to the fact that these words and sentences can be recursively combined to express new sentences and meanings. These mechanisms create a parallel between the structure of language and that of meaning (including sensorimotor representations). When considering such remarkable similarities between language and action, some fundamental questions arise: Why do language and action share such hierarchical and compositional structure and properties? Is there a univocal relationship between them (e.g. the structure of action influences that of language, or vice versa), or do they affect each other in a reciprocal way? Do these two abilities share common evolutionary, and/or developmental, processes?

These scientific questions will be investigated through new robotic experiments based on the combination of evolutionary algorithms and ontogenetic/developmental learning algorithms. These experiments will be based on robotic simulations due to time constraints involved in evolutionary computation (i.e. parallel testing of many robots within one generation, to be repeated for hundred of selection/reproduction cycles). Experiment will directly address some of the language origins hypothesis on action/language interaction. For example, one study will consider Corballis (2002) hypothesis that language evolved from the primates’ ability to use and make tools and the corresponding cognitive representation that such a compositional behavior requires. Evolutionary simulations will first look at the evolution of tool use and object manipulation capabilities, as in THEME1. Subsequently, agents will be allowed to communicate about their action and object repertoire. The analysis

of evolutionary advantages in pre-evolving object manipulation capability will be considered. Another simulation will consider Greenfield (1991) study on sequential sorting behaviour and its relationship to language and motor development (evolutionary and ontogenetic). Children use different dominant strategies in sequential tasks such as nesting cups, e.g. from an early “pot” strategy (move one cup at a time) to a later “subassembly” strategy (moved pairs or triples of stacked cups). Greenfield suggests that language and sorting task processes are built upon an initially common neurological foundation, which then divides into separate specialized areas as development progresses. Such a hypothesis will be studied in simulation on the manipulations of the topology of the neural network controlling the agents’ linguistic and motor behaviour. Simulations will provide further insights on the evolutionary relationship between action and language structure, as well as providing new methodologies for the combination of evolutionary and ontogenetic learning mechanisms in communicating cognitive systems.

2.4 Action basis of language processing.

Psycholinguistic data on Action-Compatibility Effects (ACE) during language comprehension tasks (Glenberg & Kaschak, 2002) support an embodied theory of language that strictly relates the meaning of sentences to human action and motor affordances. Glenberg & Robertson (2000) have proposed the Indexical Hypothesis to explain the detailed interaction of language and action knowledge. This suggests that sentences are understood by creating a simulation of the actions that underlie them. When reading a sentence, the first process is to index words and phrases to objects in the environment or to analogical perceptual symbols. The second process is deriving affordances from the object or perceptual symbol. Finally, the third process is to mesh the affordances into a coherent set of actions. The mesh process is guided by the syntax of the sentence being processed. This suggests a parallel between syntax and action. Syntax has the role of combining linguistic components into an acceptable sentence. Motor control has the role of combining movements to produce the desired action. Moreover, Glenberg (personal communication) suggests that syntax emerges from using linguistic elements to guide mechanisms of motor control to produce effective action or a simulation of it. Such a view is compatible with construction grammar hypothesis that suggests that linguistic knowledge consists of a collection of meaning/symbol pairs reflecting, amongst other things, action roles and properties. In this project we will carry out robotic simulations of language comprehension and ACE. For example, experiments using the i-cub platform will train robots to acquire an action repertoire producing various motor affordance representations and constructs (e.g. give-object-to, receive-object-from, lift-object etc.). In parallel the robot will learn the names of actions in such a way to reproduce ACE phenomena due to action/language simulations.

2.5 Mental representation of categories and concepts

There have been two orthogonal approaches to representing categories in artificial systems: one commonly known as the symbolic approach, the other as the subsymbolic approach. In the symbolic approach, conceptual information is represented as a symbolic expression containing recursive expressions and logical connectors, while in the subsymbolic approach concepts are represented in a continuous domain, for example in connectionist networks or semantic spaces (cf. Gärdenfors, 2000). Both approaches serve their purpose, but none seems to resonate well with human conceptualisation. Human use symbolic knowledge in representations for communication and reasoning (Deacon, 1997), but these symbols are implemented in neural tissue, which is unsymbolic and imprecise. There have been few attempts to reconcile both, and one of the aims of the project is to design a conceptual

representation which has the precision of logic symbols, but the plasticity of human concepts. This representation should also support the acquisition of concepts through sensorimotor interactions with the environment and other agents and through linguistic interactions.

In order to understand how humans represent knowledge, we believe that much can be learned from studying how infants and young children acquire concepts. There are many experimental studies and theories on concept acquisition in young children (Rakison & Oakes, 2003). Children, for example, seem to employ a number of strategies to facilitate concept acquisition, such as mutual exclusivity, where a word is only related to one object in a context and not to others (Markman, 1989), or the preference to bind unfamiliar words with unfamiliar perceptual input. Also, language seems to play a crucial role in concept acquisition. There is support for this still controversial issue from recent evidence of linguistic relativism (for example Gentner & Boroditsky, 2001) and from the experimental evidence of the critical importance of linguistic input on concept acquisition (e.g. Bowerman & Levinson, 2001; Choi, McDonough, Bowerman, & Mandler, 1999). In addition, children employ world knowledge, or theories, during the acquisition of knowledge as well as biases on the perception which helps to restrict the number of possibilities while learning (see Rakison & Oakes 2003 for an overview).

3. Implications for Road Map on Action-Language Integration in Developmental Robotics

The above research issues constitute some of the key challenges for research in developmental cognitive robotics, in particular regarding ongoing and future work on (linguistic) communication between interactive robots. Other core issues in developmental robotics regard additional linguistic/communicative capabilities (such new developments in phonetic and articulatory systems, or new insights in concept acquisition and the influence of language on the process) as well as other cognitive and behavioural abilities. These include research on motivation and emotions, on perception and action, on social interaction, and on higher-order cognitive skills such as decision making and planning.

In addition to research specifically addressing individual cognitive skills and their interaction, other core cognitive robotics research issues regard general cognitive capabilities. In particular, two main challenges regard the further development of learning techniques (e.g. development of new, scalable learning algorithms) and the design of brain-inspired techniques for robot control.

If we consider future advancements on developmental robotics and the parallel progresses in the various cognitive and behavioural capabilities, we can identify the following sequence of milestones for what regards specifically research on action and language learning and integration:

- Milestone 1: Developmental learning of simple actions, and subsequent acquisition of hierarchical and compositional actions (and their internal representations)
- Milestone 2: Developmental acquisition of a capacity to categorise and name objects, events and states (and their internal representations)
- Milestone 3: Integration of action and naming representations and emergence of shared representation roles for both actions and names
- Milestone 4: Bootstrapping of linguistic capabilities with the acquisition of situation-specific grammar constructions and compositional language structures

Milestone 5: Development of general-purpose grammatical constructions and full blown syntactic competence

The above list of milestones provides a possible sequential list of goals and test-scenarios. However, we do not propose a fully temporal and differentiated approach to these milestones, especially as there will be overlap of cognitive capabilities development in the transition between milestones/stages. This milestone list, together with other proposals on language development stages (for example Steels, 2005, grammaticalisation stages), can contribute to the definition of the Road Map for developmental cognitive robotics research.

References

- Batali, J. (2002) The negotiation and acquisition of recursive grammars as a result of competition among exemplars. In Ted Briscoe (ed.), *Linguistic Evolution through Language Acquisition: Formal and Computational Models*. Cambridge University Press.
- Bowerman, M., & Levinson, S. (2001). *Language Acquisition and Conceptual Development*. Cambridge: Cambridge University Press.
- Cangelosi A. & Harnad S. (2000). The adaptive advantage of symbolic theft over sensorimotor toil: Grounding language in perceptual categories, *Evolution of Communication*, 4, 117-142.
- Cangelosi A., Bugmann G. & Borisyuk R. (Eds.) (2005). *Modeling Language, Cognition and Action*. Singapore: World Scientific.
- Cappa, S. F., & Perani, D. (2003). The neural correlates of noun and verb processing. *Journal of Neurolinguistics*, 16(2-3), 183-189.
- Choi, S., McDonough, L., Bowerman, M., & Mandler, J. M. (1999). Early sensitivity to language-specific spatial categories in English and Korean. *Cognitive Development*, 14(2), 241-268.
- Corballis, M. C. (2002). *From hand to mouth: The origins of language*. Princeton, NJ: Princeton University Press
- Deacon, T. W. (1997). *The symbolic species: the co-evolution of language and the brain*. New York: W.W. Norton.
- Elman J.L. (1990). Finding structure in time. *Cognitive Science*, 14, 179-211.
- Fadiga, L., Fogassi, L., Gallese, V., & Rizzolatti, G. (2000). Visuomotor neurons: ambiguity of the discharge or 'motor' perception? *International Journal of Psychophysiology*, 35(2-3), 165-177.
- Fillmore C. (1988), The mechanisms of construction grammar. *Proceedings of the 14th Berkeley Linguistic Society Meeting*, pp. 35-53.
- Gallese, V., Fadiga, L., Fogassi, L., & Rizzolatti, G. (1996). Action recognition in the premotor cortex. *Brain*, 119(2), 593-609.
- Gärdenfors, P. (2000). *Conceptual spaces: The geometry of thought*. Cambridge, MA: The MIT Press, Bradford Books.
- Gentner, D., & Boroditsky, L. (2001). Individuation, relational relativity and early word learning. *Language acquisition and conceptual development*,
- Glenberg, A. M., & Kaschak, M. P. (2002). Grounding language in action. *Psychonomic Bulletin & Review*, 9(3), 558-565.
- Glenberg, A. M., & Robertson, D. A. (2000). Symbol grounding and meaning: A comparison of high-dimensional and embodied theories of meaning. *Journal of Memory and Language*, 43(3), 379-401.

- Goldberg A. (1995) *Constructions. A Construction Grammar Approach to Argument Structure*. University of Chicago Press.
- Goldberg A. (2006). *Constructions at Work: The Nature of Generalization in Language*. Oxford University Press.
- Greenfield P.M. (1991). Language, tools and the brain: the ontogeny and phylogeny of hierarchically organized sequential behavior. *Behavioral and Brain Sciences*, 14, 531-595.
- Hauk, O., Johnsrude, I., & Pulvermuller, F. (2004). Somatotopic representation of action words in human motor and premotor cortex. *Neuron*, 41(2), 301-307.
- Hayashi M. & Matsuzawa T. (2003) Cognitive development in object manipulation by infant chimpanzees. *Animal Cognition*, 6(4), 225-233.
- Lakoff, G. (1987). *Women, Fire, and Dangerous Things: What Categories Reveal about the Mind*. Chicago: University of Chicago Press.
- Langacker, R. W. (1987). *Foundations of cognitive grammar*. Stanford, Calif.: Stanford University Press.
- Macwhinney, B. (2005). The emergence of linguistic form in time. *Connection Science*, 17(3-4), 191-211.
- Markman, E. M. (1989). *Categorization and naming in children: problems of induction*. Cambridge, MA: The MIT Press.
- Pecher, D. & Zwaan, R.A. (2005). *Grounding Cognition: The Role of Perception and Action in Memory, Language, and Thinking*. Cambridge University Press.
- Pulvermuller, F., Hauk, O., Shtyrov, Y., Johnsrude, I., Nikulin, V., & Ilmoniemi, R. (2003). Interactions of language and actions. *Psychophysiology*, 40, S70-S70.
- Rakison, D. H., & Oakes, L. M. (2003). *Early Category and Concept Development: Making Sense of the Blooming, Buzzing Confusion*: Oxford University Press.
- Rizzolatti, G., & Arbib, M. A. (1998). Language within our grasp. *Trends in Neurosciences*, 21(5), 188-194.
- Steels, L. (2005). The emergence and evolution of linguistic structure: from lexical to grammatical communication systems. *Connection Science*, 17(3-4), 213-230.
- Sugita Y. & J. Tani (2005). Learning semantic combinatoriality from the interaction between linguistic and behavioral processes. *Adaptive Behavior*, 13(1), 33-52.
- Tanaka M. & Tanaka S. (1982) Developmental diagnosis of human infants: from 6 to 18 months Otsuki, Tokyo.
- Tomasello, M. (2003). *Constructing a language :a usage-based theory of language acquisition*. Cambridge, MA: Harvard University.