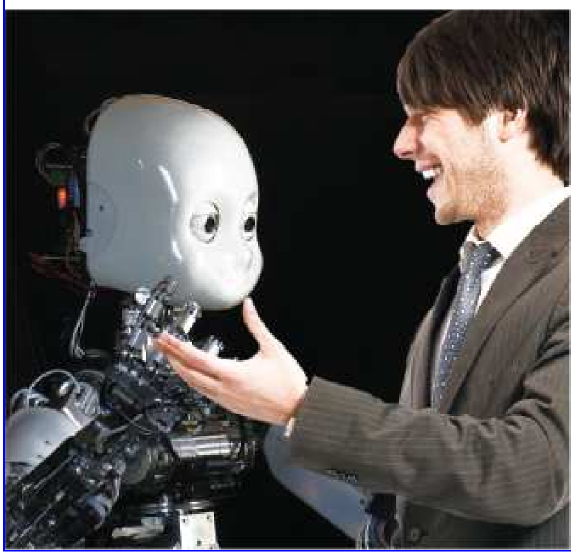
1.1 Publishable summary

"*Stir the coffee*" is a simple verbal request for an everyday activity. When addressed to an artificial agent though, the complexity of the involved motor program, the perceptual skills employed and the cognitive processes required for bridging the gap between the verbal request and the actual behaviour are far from trivial. State of the art observation learning and imitation techniques can be used successfully for teaching a robot the corresponding behaviour, such as moving a spoon in a circular motion within a container. Machine learning algorithms can be employed to train the robot to associate the specific words with the implied behaviours and visual objects. Even the visually difficult to perceive '*coffee*' concept can be identified with proper training images (coffee in a mug), while olfactory and tactile sensors could be employed for identification of the substance. However, developing and integrating cutting edge technologies for teaching an agent how to 'stir coffee' stumbles on the need to go beyond the learned motor program (and the involved objects). Could our agent generalise over this learned behaviour and '*stir the soup*'? If no *spoon* is available, could it find and use another appropriate object to perform the task and if we ask the agent to '*stir the pot*', what kind of action will it perform and on which objects?

POETICON**++** argues that **robots need natural language** for *controlled* ***generalisation*** *of learned behaviours* and for ***creativity***. Its main objective is the development of an innovative *computational mechanism* for robust generalisation *of motor programs and visual experiences for robots* that will utilise the hierarchical and generative nature of language for 'indexing' (labelling) sensorimotor experiences at different levels of abstraction. The mechanism will integrate natural language and visual action and object recognition tools with advanced manipulation and mobility skills, affordance-based self-exploration abilities and a bio-inspired action-language learning module. All tools will be integrated in the iCub humanoid and will be employed in novel action generalisation and creativity experiments in two scenarios of 'everyday activities', comprising of (a) *behaviour generation through verbal instruction*, and (b) *visual scene understanding*.




POETICON**++** advocates that language is **necessary** for robots to generalise behaviours and perceptions and to create new ones and puts forward a concrete methodology for developing the corresponding cognitive computational mechanism. There is increasingly growing evidence that language is inherently connected to action and perception which has impinged in a significant part of computational and engineering research and cognitive robotics in particular; however, the POETICON++ statement is not a conventional statement among sensorimotor scientists and roboticists that have concentrated on control aspects of robotics. This is because there is a large variety of biological systems (birds, insects, cats, mammals, etc.) possessing amazing robotic capabilities (sensorimotor competences), which have no language with the hierarchical and compositional characteristics of human language. This is of course true, but not directly related to our central thesis. Robots need language for helping them to structure their knowledge and learn new behaviours of any level of complexity, something very challenging for sensorimotor systems without such language. However, contrary to most approaches that study language learning and how systems can reach the point of learning language, we start with the premise that our robot *has* language. Cognitive robotics research is geared towards closing the loop between robot sensing, robot acting and robot learning, with language kept as a communication interface with humans. We suggest that language (through its hierarchical and compositional structure) can play a more dynamic role and should be included in the loop. Our project will show why this is significant leading to the next generation of robots.

Istituto Italiano di Tecnologia, Department of Robotics, Brain & Cognitive Sciences (IIT) – Italy
Giulio Sandini, Luciano Fadiga, Giorgio Metta
Role in the project: Project Administrative Coordination, Development of advanced manipulation and mobility skills for iCub, Neurophysiological Experiments

"Athena" Research Centre (ATHENA R.C), Institute for Language and Speech Processing – Greece
Katerina Pastra
Role in the project: Project Scientific Coordination, Development of the language-based behaviour generalisation mechanism, System integration, Dissemination

Cognitive Systems Research Institute (CSRI), Embodied Language Processing Group – Greece
Katerina Pastra
Role in the Project: Development of embodied language parsers

University System of Maryland Foundation, University of Maryland College Park, Computer Vision Lab (USMF) – USA
Yiannis Aloimonos
Role in the project: Development of visual object and action parsers

University of Plymouth, Centre for Robotics and Neural Systems (UOP) – United Kingdom
Angelo Cangelosi
Role in the project: Development of a bio-inspired language-action learning architecture

Instituto Superior Tecnico, Computer and Robot Vision Lab (IST) – Portugal
Jose Santos Victor
Role in the project: Development of a language-augmented self-exploration model for humanoids

Figure 1: Project Partners – For updated information on all project results please visit the project website at: http://www.poeticon.eu

The scientific and technological objectives of the project that will be monitored to achieve the general aim are:

1. To develop a cognitively-plausible computational learning model and mechanism for language-mediated behaviour generalisation and creativity in interactive cognitive systems;
2. To demonstrate, through humanoid robotic experiments, the feasibility and validity of this model in two dynamic, real-life scenarios: behaviour generation through verbal instruction, and visual scene understanding.
3. To advance our understanding of the neuroscientific, cognitive and linguistic phenomena and mechanisms supporting the generalisation and creation of new behaviour through the essential role of language and action-language hierarchical representations;

In particular, POETICON++ will develop a number of basic technologies for cognitive artificial agents, which will enable them to generalise their behaviours and cope with uncertaintenty and unexpected situations in real life environments. The technologies will comprise of:

(a) A set of *embodied, generative language processing tools* that will bridge the gap between verbal communication and the sensorimotor space. The tools will comprise a new generation of language technology that will mine everyday, common sense information from large language resources. They will be engaged in a cognitive dialogue with visual object and action parsers and will incorporate corresponding reasoning mechanisms for the tasks.

(b) A set of *generative visual object and action analysis tools* that will be engaged in a cognitive dialogue with the language tools in the above mentioned tasks. They will analyse visual action scenes of any complexity, taking advantage of prior knowledge and reasoning (provided by language) as well as self-exploration abilities of an agent. The latter will contribute to active vision strategies in visual object and action analysis.

(c) A *self-exploration model* that will integrate motoric skills, multisensory perception skills (visual and tactile) and verbal labelling of self-acquired sensorimotor experiences for artificial agents.

(d) Improved *grasping skills* for a humanoid *via learning and affordances.*

(e) A *word-level articulation-based automatic speech recognition system;* this module will allow an agent to go beyond traditional speech recognition abilities using motor information.

The above technologies will be integrated and tested within two real life environment scenarios/applications:

(a) *Verbal-based human-robot interaction*: in this demonstrator, all cognitive tools developed in the project will be integrated within a cognitive dialogue framework for optimal generalisation and for creativeness in unexpected real life interaction cases. The demonstrator will guide the development of the tools and their integration, and will also be used as proof of their usability.

(b) *Visual scene understanding:* in this demonstrator, all cognitive tools developed in the project are integrated within a cognitive dialogue framework for dealing with un-seen scenes generalising an artificial agent's perceptual experiences and creating new ones. The demonstrator will guide the development of the tools and their integration, and will also be used as proof of their usability.

Last, beyond the development of technology for behaviour generalisation in artificial agents, and the related demonstrators, POETICON++ will also explore aspects of the link between action and language through experimentation, including:

(a) Exploration of the *neural activation of the motor system during generation of structured action*

(b) Exploration of *the role of the motor system in speech perception*

(c) Exploration of the optimal developmental training protocol for action-language learning in humanoids

(d) Experimentation with employing a neuroanatomical model of language and action learning.

POETICON++ will carry out inter-disciplinary research involving Computational Linguistics, Computer Vision, Robotics, Neuroscience and Cognitive Modelling. Its scientific contributions will put it in the frontline of machine intelligence and artificial cognitive agents research internationally. The tools, models and skills that the iCub humanoid will be endowed with will contribute to the development of active agent-interpreters engaged in real life interaction. Both intuitive, intelligent, multimodal human-machine/robot interaction and intelligent multimedia processing within embedded or assistive systems for interpersonal communication will benefit from this research. The action grammar we present and the related integrated cognitive dialogue has the potential to radically influence the field of Artificial Intelligence. Therefore, the expected impact will be two-fold; impact on the related research in Cognitive Science and Artificial Intelligence and on the corresponding market.

During its second year, POETICON++ has achieved the following:

- Development of a cognitive semantic similarity metric (COGSIM), which forms the core of the POETICON++ embodied reasoner (WP2, task 2.4). The metric comes 1st in evaluation against the best performing state of the art metrics.
- Development of the first ever Concept Specifity identification algorithm (CONSPEC) (WP2, task 2.2) in the form of a logistic regression model; the model reaches above 90% F-score in an experimental test-set. Its results, and in particular its identification of Basic Level Concepts form core part of COGSIM. The success of the latter is an indirect, task-based evaluation of this algorithm too.
- Experimental findings on the grounding of lexical feature concepts on exploratory actions (Eas), (important for WP2, task 2.3): contrary to long-standing experimental results by the work of Klatzky et al. (1985; 1987; 1987b), there is actually no one-to-one matching between EAs and object features; instead, a set of specific EAs can facilitate the acquisition of any given object feature.
- Object recognition in cluttered environments and during hand manipulation through enhancement of the image torque (WP3, task 3.1).

- Development of a new method for monitoring objects during action that combines segmentation and tracking and development of a manipulation action parser (WP3, task 3.2).
- Action effects recogntion module for a set of 5 action effects (WP3, task 3.3).
- Development of a tool-enabled affordance model for learning and reasoning over the geometrical characteristics of tools, including shape information (WP4, task 4.1).
- Development of a mid-level planner for mediating between high-level plans coming from the semantic memory, visual affordance prior knowledge and physical constraints imposed by the environment/setting (WP4, task 4.2).
- Redesign and preliminary implementation of the grasping of objects via learning and affordances and improvement of grasping control (WP5, task 5.2).
- A scaled-up version of the MTRNN model with proprioceptive input and two new self-organising maps encoding input from vision and language modalities (WP5, task 5.4).
- Formulation of a single model, which without switching on or off any abilities or modules, or altering any parameters, will be used to naturally progress through developmental language-action learning sequence and replicates data from various child studies in each stage (WP5, task 5.5).
- Findings showing that disruption of Broca's area selectively impairs high-order chunking during sequence learning and the existence of brain signatures of content and structure violation processing during action observation (WP5, task 5.7)
- Findings showing that motor activities become critical when subjects are presented with natural occurring inter-speaker variability; the recruitment of sensori-motor processes in speech perception is automatic, and listening to speech recruits specific tongue motor synergies as revealed by TMS and Tissue-Doppler Ultrasound Imaging (WP5, task 5.8)
- Findings showing that speaker-dependent recognition accuracy of state-of-the-art Deep Neural Networks / Hidden Markov Model phone recognition systems increases when reconstructed articulatory features are combined with acoustic features. The improvement is consistent across different corpora and languages. (WP5, task 5.9).
- Development of integrated demonstration scenarios, of increasing complexity and a related extensive and critical review of action generalisation and complexity in Robotics (WP6 & WP7).