

## **“Language-Action Tools for Cognitive Artificial Agents: Integrating Vision, Action and Language”**

**Katerina Pastra and Yiannis Aloimonos**

The “Language-Action Tools for Cognitive Artificial Agents” Workshop was a two-day satellite event of the 25<sup>th</sup> Conference in Artificial Intelligence (AAAI 2011) in San Francisco, California from March 7-8, 2011. The goal of this workshop was to investigate the interface of computational linguistics, computer vision, robotics, neuroscience and cognitive psychology in developing biologically-inspired tools that endow artificial agents with language, action and perception integration abilities.

Endowing artificial agents with language and action abilities has been a quest in many Artificial Intelligence (AI) subfields. Recent years have witnessed great advances in different disciplines that provide the theoretical and technological framework for an interdisciplinary approach to language-action integration. Neuroscience research provides more and more evidence on a common neural basis for language and action, both in perception and in production. The goal of this workshop was to shed light on the different aspects of *developing biologically-inspired language and action technology for artificial agents*.

The workshop brought together researchers from a variety of subfields of AI such as computational linguists, computer vision researchers, roboticists, computational neuroscientists as well as experimental neuroscience and cognitive psychology researchers. It comprised fifteen invited talks, five paper presentations and a panel discussion and attracted funding from *AAAI*, *Google Inc.* and the *EuCognition Network*.

One major theme of invited talks was related to action information as expressed through language; a talk given by *Jerry Feldman* (UC Berkeley) presented *simulation semantics* as a cognitively motivated computational framework for analysing verbally expressed events, while *Barbara Landau* (John Hopkins University) focussed on comparative data of spatial action language descriptions and the clues revealed by such comparisons regarding basic spatial cognition. *Stanley Peters* (Stanford University) presented the challenges in grounding even simple verbal references to agents of an event and related work in the analysis of spoken language meeting data. Katerina Pastra (Cognitive Systems Research Institute) presented a generative grammar of action (in the motoric space) and the role of language as a labelling system for action syntax trees which is generative itself and which requires a new generation of ‘embodied language technology’. Yiannis Aloimonos (University of Maryland) introduced the notion of a ‘cognitive dialogue’ between perceptual executives, the motor executive and language executive as a computational framework for developing cognitive artificial agents that are ‘active interpreters’.

Another major theme was that of learning language-perception associations. *Tamara Berg* (SUNY Stony Brook) presented work on labelling images and discovering visual attribute terms from web collections. In her turn, *Evelyn Tzoukerman* (MITRE) presented work on using transcripts and online data for developing joint models of visual action/objects and

accompanying text. *Ray Mooney* (University of Texas), presented language-learning systems that learn from sentences paired with ambiguous, naturally-occurring perceptual environments, such as sportscast in simulated robocup games and navigation directions in virtual world settings. The use of language as a tutoring tool for learning the meaning of actions that cannot be inferred by observation alone and its use in robot tutoring was elaborated in *Britta Wrede's* (University of Bielefeld) invited talk. Along similar lines, *Gabriela Vigliocco* (University College London) advocated the integration of embodied (perceptual, motoric and affective) experience and information expressed through language in word meaning representations and discussed likely mechanisms for the acquisition of such representations in childhood.

A number of invited talks comprised demonstrations of robots integrating –among other cognitive abilities- perception, action and natural language; *Jeffrey Siskind* (Purdue University) demonstrated two robots engaged in playing a board game while a third one observes the play to infer the game rules and a Lincoln logs assembly task undertaken by a robot, while a second one observes that activity and communicates those observations, in natural language, to a third robot who must replicate that assembly. *Vadim Tikhanoff* (Italian Institute of Technology) demonstrated the POETICON-iCub humanoid succeeding when verbally instructed to perform an everyday task such as ‘stirring the coffee’ without knowing in advance what ‘coffee’ is and deprived of the most commonly used tools for the task.

Last, neural perspectives of the language-perception-action integration issue were addressed; *Tom Dean* (Google Inc.) gave an overview of principles that derive from the study of areas of the brain other than neocortex and suggested ways of employing them in computer vision. *Max Garagnani* (MRC Cambridge) presented a neuroanatomically grounded computational model of sensorimotor circuits for language and action, while *Jun Tani* (Riken Institute) presented a neurodynamical model of compositionality for goal-directed action generation, mental simulation and planning, free-decisions, and language-action associative learning.

The participants discussed the challenges and future directions in this emerging field within a panel discussion moderated by *Ruzena Bajcsy* (UC Berkeley) and *Vincent Müller* (Anatolia College). Meetings of such highly interdisciplinary nature were considered by all participants to be the best means for cross-fertilization of ideas, methods and practices.

Katerina Pastra and Yiannis Aloimonos served as cochairs of this workshop. The papers of the workshop were published as AAAI Press Technical Report WS-11-14.

Katerina Pastra is the Director of the Cognitive Systems Research Institute, Athens, Greece. Yiannis Aloimonos is a Professor in the Department of Computer Science at the University of Maryland, College Park and the Director of the Computer Vision Laboratory.