

# Connecting Language to Robot Behaviour: An Architecture of Object Schemas

Rony Kubat, Kai-yuh Hsiao and Deb Roy

[kubat, eepness, dkroy]@media.mit.edu

August 1, 2008

Our shared visions of future robots interacting, cooperating and aiding humans carry a common element: a primary mode of interactivity via natural, human language. Moreover, the situations in which we envision our robotic aids and companions are full of an ever-changing and richly dynamic world. These two observations place strong requirements of any robotic system: an ability to cope with—and respond to—a continuously evolving environment, and a power to understand and respond to a partner using natural language. This talk addresses both these requirements in concert by offering an architecture of dynamic interactive processes which ground language to physical objects in the world. An implementation of the architecture on a physical robot demonstrates how these two core requirements are confronted with a unified approach.

Many early robotic systems interacted with the world through a chain of components: sensing the world, building a simplified internal model of it, planning an action based on this model and finally acting upon this plan (classically: [9]). Language, when integrated into systems of this architecture, commonly connected to the world-model and planning phases. As the complexity of the world model in an SMPA system increases, the greater become the challenges of maintaining synchronization between model and world. Behaviour-based systems, an alternative to the Sense-Model-Plan-Act (SMPA) approach popularized in the robotics community by Brooks [1, 2], addressed this problem by layering collections of concurrent processes each of which vie for control of limited output resources. In the classic behaviour-based approach, there is no explicit model of the world, and as a result, grounding language—which by making reference to the world implicitly demands a kind of model—is an awkward and inelegant task [5]. Many modern systems [3, 4, 6, 7] work with a hybrid between these two approaches, using lower-level behaviours for safety and collision avoidance, and using SMPA to make plans and attend to goals.

Our approach, based on the theoretical framework presented in [11, 12] and influenced by Piaget’s sensorimotor schema [10], explicitly models the world by use of concurrent behaviour-like processes called *object schemas*. The coupling of these processes encode beliefs about physical objects in the environment and provide a substrate upon which planning operates. The system remains respon-

sive to the environment through the concurrency of the object schemas.

The basic building blocks of the architecture are *interaction processes*: concurrent elements that communicate with each other through shared memory. *Interaction processes* (or processes, for brevity) perform tasks such as processing sensor data, commanding motors, generating verbal output, and manipulating internal records. Each process posts data to the shared memory as an *interaction history*. Processes may also post anticipated data to their interaction history to aid other elements. For example, a grasping process may post an anticipated completion status of success with a completion time of twelve seconds. The anticipated success can then be leveraged by planning processes to choose appropriate actions. The collection of currently running processes is called the *belief context*.

Processes are divided into seven classes:

**sensory:** Responsible for monitoring raw sensor values and reifying them to the interaction history.

**action:** Motor commands are sent when these processes are triggered by the planning system.

**plan fragments:** Components of a hierarchical planning system, providing preconditions and anticipated postconditions.

**condition:** Continuously evaluate a condition, triggering the creation of new plan fragments by the planning system.

**translation:** Manage the transformation from one coordinate space to another, and transformations between continuous values and discrete categories.

**coordination** Maintain the integrity of an object schema by coupling other processes.

**reference** Connect noun phrases from speech input to the object schemas which best match the linguistic input.

Actions are taken by the robot by executing hierarchical plans whose root nodes represent prioritized “motivations”. Leaf nodes in the hierarchy send commands to motors or activate a text-to-speech verbal response process. Other interior nodes in the hierarchy represent higher-level plan subcomponents. For example, an interior plan node may require that the hand be at the location of an object before the fingers close.

Linguistic input enters the system as a parse tree, whose tokens are then bound to processes in the belief context. Reference processes couple noun phrases to object schemas present in the belief context, and verbs are likewise tied to nodes in the plan hierarchy.

The object schema architecture was implemented on Trisk, a ten degree of freedom robot working in a desktop domain. Using an arm (6 DOF) with three-fingered hand (4 DOF), Trisk can respond to simple language instructions such as “Pick up the heavy. . . no, red, ball” or “group the red apple and green block.”

The object-schema architecture has several benefits relative to SMPA, behaviour-based and hybrid approaches. The architecture is computationally scalable because its components are concurrent and often independent. That objects in the world are represented as dynamic processes allows the system to remain responsive to change. This common currency of process-based representation conveniently allows natural language interaction to be integrated. Finally, object affordances [8] such as “liftability” and “graspability” are easily represented using the interaction histories.

**Acknowledgments** This paper is partly based upon work supported under a National Science Foundation Graduate Research Fellowship.

## References

- [1] R.A. Brooks. A robust layered control system for a mobile robot. *IEEE Journal of Robotics and Automation*, 2:14–23, 1986.
- [2] R.A. Brooks. Intelligence without representation. *Artificial Intelligence*, 47(pp.139–160), 1991.
- [3] J. J. Bryson. *Intelligence by Design: Principles of Modularity and Coordination for Engineering Complex Adaptive Agents*. PhD thesis, Massachusetts Institute of Technology, 2001.
- [4] J. J. Bryson and L. A. Stein. Modularity and design in reactive intelligence. In *Proceedings of the International Joint Conference on Artificial Intelligence*, pages 1115–1120, 2001.
- [5] A. Clark and J. Toribio. Doing with representing. *Synthese*, 101:401–431, 1994.
- [6] E. Gat. Integrating planning and reaction in a heterogeneous asynchronous architecture for controlling mobile robots. In *Proceedings of the Tenth National Conference on Artificial Intelligence (AAAI)*, 1992.
- [7] E. Gat. Three-layer architectures. In D. Krotenkamp, R.P. Bannasso, and R. Murphy, editors, *Artificial Intelligence and Mobile Robots*. AAAI Press, 1998.
- [8] J. J. Gibson. *The Ecological Approach to Visual Perception*. Erlbaum, 1979.
- [9] N. J. Nilsson. Shakey the robot. Technical Report 323, AI Center, SRI International, 1984.
- [10] J. Piaget. *The Construction of Reality in the Child*. Basic Books, 1955.
- [11] D. Roy. Semiotic schemas: A framework for grounding language in action and perception. *Artificial Intelligence*, 167(1-2):170–205, 2005.

- [12] D. Roy. A mechanistic model of three facets of meaning. In M.D. Vega, G. Glennberg, and G. Graesser, editors, *Symbols and Embodiment*. Oxford University Press, 2008.