The Semantics of Doing and Seeing in Simulated Worlds

James Pustejovsky
Brandeis University

In order to facilitate communication with a computational agent, we have been pursuing a new approach to modeling the semantics of natural language: Multimodal Semantic Simulations (MSS). This framework assumes both a richer formal model of events and their participants than typically assumed in this area, as well as a modeling language for constructing 3D visualizations of objects and events denoted by linguistic expressions (a denotation not typically employed in linguistics). The Dynamic Event Model (DEM) encodes events as programs in a dynamic logic with an operational semantics, while the language VoxML, Visual Object Concept Modeling Language, is being used as the platform for multimodal semantic simulations in the context of human-computer communication, as well as for image- and video-related content-based grounding and querying. The simulation environment we describe is presently configured for joint activity and communication between a human and a computational agent. But because of the nature and design of the VoxML model, we believe it can be used as the conceptual platform for robotic representation, reasoning, and concept learning. In this talk, I discuss the semantics of events and actions in this framework, as well as an initial model of synthetic vision for simulated agent perception.