Referring Expressions: Straddling the Divide

Deb Roy (dkroy@media.mit.edu)
MIT Media Laboratory
Cambridge, MA, USA

Our lab studies language use that is grounded in immediate physical context. Physical grounding of reference is seen as the foundation of social common ground and thus of shared meanings. Our explorations into this area are guided by two methods:

1. Create interactive communication systems (physical and virtual robots) to develop mechanistic models of physically grounded reference. Reference is viewed as the product of an interactive sensory-motor process that binds referents (physical objects, properties, events) to cognitive associates. This line of work leads us to recognize a necessary link between reference and teleology (Roy, 2005; Roy, 2008) in which affordances take on a central explanatory role in connecting objects to words (Roy, 2005; Gorniak & Roy, 2007).

2. Gather and analyze empirical data of human-human language use in naturalistic social contexts. We are collecting longitudinal observations of child language development (Roy et. al., 2005; Roy, 2009) and cross-sectional data on adult language use in specific social contexts (e.g., having a meal in a restaurant) (Orkin & Roy, 2007). A key form of analysis is the discovery of recurrent patterns of non-linguistic activity embedded in observational data. Recurrent patterns provide a representation of activity contexts that in turn provide a basis for analyzing grounded language use.

A fundamental divide exists between these two methods. Interactive systems let us concretely model the connections between language users' goals and semantics. This is possible since as designers we are able to precisely engineer generative planning systems that cause synthetic agent behavior. Observation of human language in naturalistic contexts, however, provides no direct access to language users' intentions – as analysts we are left to ascribe goals on the basis of intuition. Our research straddles this methodological divide with only occasional connections made between the two sides (Roy, 1999; Orkin & Roy, 2009). To truly bridge the divide, controlled behavioral experiments and systems neuroscience must enter the picture.

Roy, D. (2005). Semiotic Schemas: A Framework for Grounding Language in the Action and Perception. Artificial Intelligence, 167(1-2): 170-205.

Roy, D. (2008). A Mechanistic Model of Three Facets of Meaning. Chapter to appear in Symbols, Embodiment, and Meaning, de Vega, Glenberg, and Graesser, eds.

Gorniak, P. and Roy, D. (2007). Situated Language Understanding as Filtering Perceived Affordances. Cognitive Science, 31(2), 197-231.

Roy, D., R. Patel, P. DeCamp, R. Kubat, M. Fleischman, B. Roy, N. Mavridis, S. Tellex, A. Salata, J. Guiness, M. Levit, P. Gorniak. (2006). The Human Speechome Project. Twenty-eighth Annual Meeting of the Cognitive Science Society.

Roy, D. (2009). New Horizons in the Study of Child Language Acquisition. Proceedings of Interspeech (Invited Keynote Paper).

Orkin, J. & D. Roy. (2007). The Restaurant Game: Learning Social Behavior and Language from Thousands of Players Online. Journal of Game Development, 3(1), 39-60.

Roy, D. (1999). Learning from Sights and Sounds: A Computational Model. Ph.D. in Media Arts and Sciences, MIT.

Orkin, J. & D. Roy. (2009). Automatic Learning and Generation of Social Behavior from Collective Human Gameplay. Proceedings of the 8th International Conference on Autonomous Agents and Multiagent Systems (AAMAS).