

Multimodal Communication as a Dynamic Network

Rick Dale
Cognitive and Information Sciences
University of California, Merced
cognition.org/rick

Max M. Louwerse
Institute for Intelligent Systems
The University of Memphis
www.madresearchlab.org

During a bout of face-to-face interaction, such as sharing a joke, a whole suite of dynamic processes unfolds between people. For example, the speaker may gesture and smile while the listener nods in eager anticipation. At the punch line, the two may engage in shared laughter. These bouts of interaction, stretching just over several seconds or minutes, are constituted by multiple behavioral and linguistic dimensions. This seems to be one of the most impressive properties of our ability to communicate: Human language makes use of the many degrees of freedom of our minds and bodies to engage in shared experience. In just a few moments of interaction, the cognitive system juggles numerous channels, such as joint gaze, body posture and gesture, physical properties of speech, all weaved into discourse management in order to succeed in a goal as cognitively complex as making one of our conspecifics laugh.

Studying these cognitive and behavioral channels has been the subject of diverse subfields of cognitive science, from speech processing to discourse comprehension. In typical research in these subfields, relevant channels are isolated, sometimes individually or in pairs, to identify the manner in which the channels systematically relate during language behavior. Consider the channels relevant to naturalistic, face-to-face interaction. Several studies suggest that speaker and listener knowledge influences language production and comprehension. For example, if you know someone is new to your region of the country, you will adapt how you give directions to a landmark. These so-called “audience-design” effects found in discourse suggest that there is an important mutual influence between knowledge states of interaction partners and their language production and processing systems.

A fuller picture of this process will come from an integration of the many channels that are adapted simultaneously during discourse. Here we present a formal characterization and analysis of this mutual influence that permits exploration of a very large number of channels simultaneously: *sequential dynamical systems*. This approach, derived from widely used methods of network analysis and graph theory, offers a new level of analysis in conceptualizing complex linguistic interaction. We import ideas from network analysis, applied in a wide variety of domains such as social network analysis, brain structure and function, and genomics. In this framework, a multimodal interaction is conceptualized as a graph structure. In this structure, different channels (e.g., gesture, eye movements, lexical and syntactic choice, discourse moves) are represented as nodes. Edges between these nodes are defined by data analytic considerations: E.g., *what is the functional relationship between gesturing and nodding during interaction?* The result is a holistic representation of the *overall* multimodal interaction. This “bird’s-eye view” of discourse offers the view of face-to-face interaction as a loosely coupled, adaptive multimodal network.