Taking other perspectives into account: an RSA model of perspectival reasoning

Carolyn Jane Anderson and Brian Dillon, University of Massachusetts, Amherst
carolynander@umass.edu

We propose a Rational Speech Acts (RSA) model of perspectival expressions and provide experimental support for its key insight: listeners consider multiple perspectives simultaneously. Perspectival expressions like English "come", which describes motion relative to a perspective-holder, pose an interpretative problem because there are multiple possible perspective-holders: the speaker, listener, or attitude-holder (Fillmore 1997). Consequently, (1) can mean that Thelma is traveling to either Seattle or London.

1. Context: Sam, in Seattle, says to Lucy, in London: "Thelma is coming."

Because listeners must reason both about the speaker's adopted perspective and their message, the interpretation of perspectival items can be modeled as a joint reasoning process.

We model perspectival reasoning in the RSA framework, in which listeners use Bayesian inference to calculate probabilities for worlds representing possible meanings (Bergen et al., 2012; Frank and Goodman, 2012). In our perspectival version, the listener jointly infers the probability of a world and perspective according to their model of how the speaker selects an utterance-perspective pair (the Literal Speaker).

\[
\text{Literal Listener: } p(w|m,a) \propto \text{denotation}(m,a,w) \times p(w)
\]

\[
\text{Literal Speaker: } p(m,a|w) \propto \text{softmax} \left( p(w|m,a) \sum_w \text{denotation}(m,a,w) \times p(a) - \text{Cost}(a) \right)
\]

\[
\text{Pragmatic Listener: } p(w,a|m) \propto p(m,a|w) \times p(w)
\]

where a = perspective, w = world, and m = utterance

A critical component of the model is that listeners consult multiple perspectives simultaneously, in contrast with the speaker-default proposed in prior work (Harris 2012, Barlew 2017). The multiple perspectives approach predicts that given (1), the marginal posterior probability should be highest for worlds where both the speaker and listener are at the destination.

We tested this prediction in a comprehension task. Participants read a sentence with a perspectival verb or manner-of-motion verb and then saw a scene depicting both the speaker and listener at the destination; just the speaker; just the listener; or neither. Participants indicated whether the scene and sentence matched.

For scenes showing both perspective-holders at the destination, there was no difference in reaction times in the come condition relative to the control condition (walk/come difference = -19 (+/-119) 95%CI), but in all others, RTs were slower for come (speaker: 138 (+/-126); listener: 408 (+/-152); none: 138 (+/-137)). A mixed effects regression revealed a significant interaction between the both-scene and come condition. Participants are faster to recognize the both-scene for perspectival expressions, in support of the multiple-perspectives account.


