A cost and information-based account of epistemic *must*

We show how a general model of rational inference in communication delivers the puzzlingly weak interpretation of the necessity modal *must*. At issue is the failed inference in (1): How could *must p* (1b) not entail that *p* (1a) ? Since karttunen1972, linguists have debated the meaning of this word, arguing about its semantic strength. Rather than engineering weakness into the meaning of the word *must*, our account derives its weakness as an M-implicature: *must p* is marked (i.e., costly) relative to the bare form (1a); the bare form is sufficiently strong already, so listeners weaken the interpretation of *must p*.

(1) a. It's raining.

b. It must be raining.

We begin with a careful comparison of the meanings of the two statements in *inference*, asking whether a speaker's choice between (1a) and (1b) is affected by the strength of her evidence for whether it is raining (q); whether listeners' interpretations of (1a) and (1b) differ with respect to the strength of their resulting belief in q; and whether these beliefs are determined in part by the evidence they attribute to the speaker's choice between (1a) and (1b).

In **Exp. 1 (n=40)**, we collected estimates of evidence strength. Participants on Amazon's Mechanical Turk rated the probability of q (e.g., of rain) given a piece of evidence e (e.g., *You hear the sound of water dripping on the roof*) on a sliding scale with endpoints labeled "impossible" and "certain". These estimates were used for analysis in Exps. 2 and 3.

Exp. 2 (n=40) tested how likely speakers are to use the marked *must p* utterance as evidence strength decreases. On each trial, participants were presented with a piece of evidence (e.g., *You see a person come in from outside with wet hair and wet clothes*) and were asked to choose one of four utterances—bare (1a), *must p* (1b), *probably p*, *might p*—to describe the situation to a friend. Participants were more likely to choose the more marked *must* form over the bare form as the strength of evidence decreased ($\beta = 5.4, SE = 2.4, p < .05$), even when controlling for evidence type (e.g., perceptual, reportative, inferential).

Exp. 3 (n=120) tested whether listeners' estimates of a) the probability of q and b) the strength of speakers' evidence for q differ depending on the observed utterance; i.e. whether listeners take into account their knowledge of speakers' likely utterances in different evidential states as they interpret the bare and *must* forms. On each trial, participants were presented with an utterance (e.g. *It's raining*), and were asked a) to rate the probability of q on a sliding scale with endpoints labeled "impossible" and "certain"; and b) to select one out of five pieces of evidence that the speaker must have had about q. Participants' believed q was less likely after observing the *must* utterance ($\mu = .65, sd = .21$) than after observing the bare utterance ($\mu = .86, sd = .15, \beta = -.21, SE = .02, t = -10.1, p < .0001$). In addition, average strength of evidence was lower after *must* ($\mu = .78, sd = .12$) than after the bare utterance ($\mu = .87, sd = .1, \beta = -.08, SE = .01, t = -6.8, p < .0001$).

Taken together, these results support an M-implicature account of the choice and interpretation of epistemic *must*: the longer, marked, *must* is interpreted by listeners as conveying the marked meaning that the speaker arrived at the conclusion that q via an evidentially less certain route than if they had chosen the shorter, unmarked, bare form. A structured probabilistic model of rational inference in communication in the Bayesian/game-theoretic/information-theoretic tradition (?,?,?,?,?) formalizes this pragmatic weakening.

Following (?,?,?), we present an extension of the Bayesian Rational Speech Acts framework (cite) using lexical uncertainty (cite) to derive the implicature. In this model, the semantics of the bare utterance and *must* q are relatively unconstrained. We define the semantics of the utterances such that $p(q|\textit{bare}) > \theta_b$ and $p(q|\textit{must}) > \theta_m$, where the pragmatic listener is uncertain about θ_b

and θ_m and infers the values through pragmatic reasoning. When the cost of uttering *must* q is greater than the bare form, the pragmatic listener infers that $p(q)$ is smaller than when the utterance is the less costly *bare* q.Given the weakened certainty of q, the listener may then infer that the speaker has weak or imperfect evidence of q. Our empirical results and computational model support this account and provide a new perspective on the meaning of *must*: its weakened meaning derives from straightforwardly from an M-implicature.

References