**The weakness of epistemic *must*: A pragmatic reasoning approach**

Since Karttunen (1972), semanticists have debated how to account for the relatively weak meaning of epistemic *must* (1b) compared to bare utterances without a modal (1a). Rather than engineering the weakness of *must* into the lexical semantics (Lassiter, 2014), and rather than analyzing its interpretive contribution as part of the propositional content (von Fintel & Gillies, 2007 *et seq.*), we propose a pragmatic approach and derive its interpretation as an M-implicature (Grice, 1989; Levinson, 2000): *must q* (1b) is marked relative to the bare form (1a); the bare form conveys already that it is raining (*q*), so listeners take the marked form to convey the marked meaning that the speaker arrived at the conclusion *q* via an evidentially less certain route than if they had chosen the cheaper bare form. We present experiments demonstrating (i) that *must*-statements coincide with less certain belief states formed on the basis of weaker evidence, and (ii) that epistemic *must* patterns cross-linguistically with non-propositional means of expressing weakened commitment towards the truth of the proposition. Over the course of three experiments, we assessed evidence strength, speaker commitments, and listener understanding; we tested both English and German to compare *must* with the behavior of discourse particles, a non-propositional means for expressing weakened commitment (e.g. Potts, 2005; Zimmermann, 2008). In **Exp. 1 (n=40 XXX GERMAN)**, we collected estimates of evidence strength. Participants rated the probability of *q* (e.g., of rain) given a piece of evidence (e.g., *You hear the sound of water dripping on the roof*) on a sliding scale with endpoints labeled ‘impossible’ and ‘certain’; the German task was a direct translation of the English. These estimates were used for analyses in Exps. 2 and 3. **Exp. 2 (n=40 XXX GERMAN)** tested how likely speakers are to use utterances as a function of evidence strength. On each trial, participants were presented with a piece of evidence from Exp. 1 and asked to choose one of four utterances. The English task compared bare (1a) with *must p* (1b), as well as *probably p* (1c), *might p* (1d). In the German task, we compared bare utterances (2a) with *muss* (2b), as well as the modal adverb *vermutlich* (2c) and the discourse particle *wohl* (lit `well) (2d). English participants were more likely to choose the more marked *must* form over the bare form as the strength of evidence decreased (β=5.4, SE=2.4, p<.05), XXX GERMAN. **Exp. 3 (n=120 XXX GERMAN)** 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. On each trial, participants saw an utterance and were asked a) to rate the probability of *q* on a sliding scale with endpoints labeled ‘impossible’ and ‘certain’; and b) to determine which of five pieces of evidence the speaker likely had. Participants believed *q* was less likely after observing the *must* utterance (μ=.65, sd=.21) than after observing the bare utterance (μ=.86, sd=.15, β =-.21, SE=.02, t=-10.1, p<.0001). Average strength of evidence was lower after *must* (μ=.78, sd=.12) than after the bare utterance (μ=.87, sd=.1, β =-.08, SE=.01, t=-6.8, p<.0001). XXX GERMAN. Our findings qualitatively confirm Karttunen’s original observation, and quantitatively demonstrate the relative weakness of *must*. Moreover, they highlight the role of evidence strength in the computation of *must*’s meaning. The cross-linguistic comparison demonstrates similar weakening with non-propositional discourse particles, further supporting our pragmatic account: the relatively weak interpretation of *must* does not require encoding weakness or indirectness into the semantics of modals. When the cost of uttering *must q* is greater than the bare form, a pragmatic listener jointly infers that p(*q*) is smaller than when the utterance is the less costly bare *q* and that the speaker has weak or imperfect evidence of *q*.

(1) a. It’s raining.

b. It must be raining.

c. It probably is raining.

d. It might be raining.

(2) XXX GERMAN

XXX PLOTS AND REFERENCES