

An experimental investigation into the weakness and evidentiality of epistemic *must*

Languages evidence a variety of ways to communicate both about beliefs and the evidence used to form them. For example, languages like Turkish and Quechua feature specialized evidential morphology with which a speaker marks an assertion for whether the evidence supporting that assertion was obtained, for example, directly, via hearsay, or through inferential means. Without specialized morphology, the English evidential system is relatively impoverished, or rather, the items used to mark evidentiality in English usually have other uses as well, which makes it difficult to tease apart a given lexical item's evidential contribution from its assertoric one. Here we focus on one such lexical item, the English epistemic necessity modal *must*. Since [3], linguists have debated the meaning of this word. The clear empirical fact is that an utterance of *must q* (1b), does not entail that *q* (1a).

- (1) a. It's raining.
- b. It must be raining.

This failed entailment is surprising under a strong necessity semantics for *must*: if it is necessarily raining, ought it not to follow that it is raining? Different accounts of the weakness of *must* have been put forward. Many have proposed that the weakness is in the semantics of *must* itself [4, 5], while others have proposed that the semantics of *must* is strong, but its weak interpretation is the result of an inference about the evidential status of *q* [8]. In particular, the claim has been that *must q* communicates that the speaker's evidence for *q* is indirect or inferential (as defined in [9]'s evidentiality tree), or that the evidence is below a certain strength or trustworthiness threshold [7]. [8] frame the semantic strength of *must* as an issue of speaker commitment: a speaker who utters *must q* is committed to the truth of *q*, given indirect evidence for *q* [8]. However, citing naturalistic corpus examples, [5] observes that speakers need not be fully committed to *q*; the strength of their belief in *q* must simply be greater (by some large margin) than the belief in any alternative, given the available indirect evidence.

One obstacle for the analysis of *must* is that many of the different ways of treating its weakness and evidentiality cannot be teased apart empirically. Here we make headway on the parts that can, as well as proposing an alternative formal, implemented, model of *must* that derives its weakness as an M-implicature: *must q* is marked (i.e., costly) relative to the bare form (1a); since the bare form is sufficiently strong to communicate *q*, listeners weaken the interpretation of *must q*. This account is implemented within the Rational Speech Act framework [1, 2], which has the advantage of being explicit about representing speaker and listener beliefs, and in turn making predictions for both production and interpretation choices.

Empirically, we address the following questions: (i) Is the use of *must q* only felicitous with indirect evidence, or with evidence whose strength is below a certain threshold, or is the probability of using *must q* probabilistically modulated by evidence strength (Exp. 2)? (ii) Does *must q* result in weak listener belief in *q* compared to bare *q* (Exp. 3a)? (iii) Does *must q* commit the speaker to *q* (Exp. 3b)?

Exp. 1 (n=40) collected estimates of evidence strength. Participants on 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 q* 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 *q* (1a), *must q* (1b), *probably q*, *might q*—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). Importantly, there were cases of direct perceptual evidence for *q* in which participants nevertheless chose the *must q* utterance (6%, replicated in a free production paradigm).

Exp. 3a (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 for *q*. Participants' believed *q* was less likely after observing *must q* ($\mu = .65$, $sd = .21$) than after observing bare *q* ($\mu = .86$, $sd = .15$, $\beta = -.21$, $SE = .02$, $t = -10.1$, $p < .0001$). In addition, average strength of evidence was judged lower after *must q* ($\mu = .78$, $sd = .12$) than after bare *q* ($\mu = .87$, $sd = .1$, $\beta = -.08$, $SE = .01$, $t = -6.8$, $p < .0001$).

Exp. 3b (n=60) tested listeners’ judgments of the speaker’s commitment to q . The procedure was as in Exp. 3a, with a minor variation in the dependent measure: participants rated the probability of the speaker believing q on a sliding scale with endpoints labeled “impossible” and “certain”. Speaker belief in q was rated higher after observing the bare utterance ($\mu = .96, sd = .07$) than after observing *must q* ($\mu = .78, sd = .18, \beta = -.2, SE = .02, t = -12.42, p < .0001$). A comparison of the data from Exps. 3a and 3b revealed that participants judged speaker commitment to be stronger than their own resulting belief in q ($\beta = .11, SE = .02, t = 5.39, p < .0001$). Nevertheless, speaker commitment for *must q* was not judged at ceiling, supporting [5] but not [8].

Taken together, these results support a pragmatic 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. Furthermore, the results suggest that the interpretation of *must* involves three related components: the listener’s inference about the speaker’s belief in q , the listener’s inference about the speaker’s evidence for q , and the listener’s inference about the truth of q . Our model seeks to clarify the relationships among these three components as well as their roles in the interpretation of *must*.

Following [6], we present an extension of the Bayesian Rational Speech Act framework using lexical uncertainty to derive the M-implicature that weakens *must* and contributes its evidential information. In this model, the semantics of the bare utterance and *must q* are relatively unconstrained. We define the semantics of the utterances such that $p(q|bare) > \theta_b$ and $p(q|must) > \theta_m$, where the pragmatic listener is uncertain about θ_b and θ_m and infers the values through pragmatic reasoning. We assume that given whether q is true in the world, the speaker is likely to have access to different kinds of evidence about q , which in turn shapes the speaker’s belief about q . When the cost of uttering *must q* is greater than the bare form, the pragmatic listener infers that $p(q)$ —the speaker’s belief about q —is less likely than when the utterance is the less costly bare q . Given the speaker’s weakened belief in q , the listener then infers that the speaker is likely to have had weak evidence of q , which leads to the inference that q is less likely to be true in the world. The model produces the pattern of results observed in our experiments, suggesting that the weakened meaning of *must* can be derived straightforwardly from an M-implicature.

Our approach has been to measure the discrete components of the meaning of *must*, use these empirical measurements to inform a computational pragmatic model of its meaning, then compare the predictions of the model with the behavior we observe in speakers and listeners. This approach allows us to move beyond the qualitative predictions of unimplemented formal models to quantitative predictions of implemented formal models. By making explicit the various aspects of *must*’s meaning, together with how they arise, we arrive at a much more conservative semantics: instead of encoding information about semantic strength and evidence strength, we derive these facts via pragmatic inference within an articulated formal model of communication. We discuss decisions at choice points in the modeling process as reflecting different assumptions about the interaction of the discrete components of the meaning of *must q* and bare q .

References

- [1] M. C. Frank and N. D. Goodman. Predicting pragmatic reasoning in language games. *Science*, 336:998, 2012.
- [2] N. D. Goodman and A. Stuhlmüller. Knowledge and implicature: modeling language understanding as social cognition. *Topics in Cognitive Science*, 5(1):173–84, 2013. ISSN 1756-8765.
- [3] L. Karttunen. *Possible and must*. In J. Kimball, editor, *Syntax and Semantics*, volume 1, pages 1–20. Academic Press, New York, 1972.
- [4] A. Kratzer. Modality. In A. von Stechow and D. Wunderlich, editors, *Semantics: An international handbook of contemporary research*, pages 639–650. De Gruyter, Berlin, 1991.
- [5] D. Lassiter. The weakness of *must*: In defense of a mantra. In T. Snider, editor, *Proceedings of SALT 24*, to appear.
- [6] D. Lassiter and N. D. Goodman. Context, scale structure, and statistics in the interpretation of positive-form adjectives. In *Proceedings of SALT 23*, pages 587–610, 2013.
- [7] L. Matthewson. Evidential restrictions on epistemic modals. In L. Alonso-Ovalle and P. Menendez-Benito, editors, *Epistemic Indefinites*. Oxford University Press, Oxford, in press.
- [8] K. von Stechow and A. S. Gillies. Must... stay... strong! *Natural Language Semantics*, 18:351–383, 2010.
- [9] T. Willett. A cross-linguistic survey of the grammaticalization of evidentiality. *Studies in Language*, 12:51–97, 1988.