A computational model for projection inferences in clause-embedding predicates

Interrogatives with clause-embedding predicates as in (1) frequently give rise to so-called *projection inferences* – that the speaker is committed to the truth of the embedded content, in this case that Julian dances salsa [1]. Multiple factors modulate projection inferences, including the lexical meaning of the predicate (factive predicates like “know” tend to give rise to projection inferences while non-factive ones like “think” tend not to [1-4]), prior knowledge about the embedded content [5], and the at-issueness of the embedded content [6]. To date there is no account that captures the interactions between these factors. Here we present a first step towards a computational model, partially couched in the Rational Speech Act (RSA) framework [6,7], that provides a *pragmatic account* of the above described patterns of projection inferences. We conducted a behavioral experiment to elicit ascribed speaker belief ratings to inform the model and evaluate its predictions.

Computational model. We model projection inferences as listeners inferring the degree to which a speaker believes the embedded content upon observing one of the five possible interrogatives in (1)-(3). Each predicate is associated with a belief threshold θ sampled from a beta distribution, and the threshold for “BARE” is set to 0.5. Under the assumed threshold semantics, the utterance can be felicitously used if the belief in the embedded content exceeds the threshold (“think”, “know”) or is close to it (“BARE”). The pragmatic speaker in Eq.(b) soft-maximizes the utility of utterances to balance utterance cost and informativeness as captured by the literal listener model in Eq.(a). What distinguishes this model from a vanilla RSA model is the pragmatic listener model (Eq.(c)): instead of using Bayes’ rule, it mixes the prior belief distribution and the inferred belief distribution based on the speaker production distribution according to how likely the embedded content is at-issue, which is parametrized via mean at-issueness ratings collected in [9].

Behavioral experiment. *Methods*. Participants (Prolific, n=345 after exclusions) rated speaker belief in the embedded content for sentences as in (1-3). Eighteen critical items from [5] were used as the embedded contents and paired with facts that made them either likely or unlikely a priori (norms taken from [5]). ***Results***. Fig.1 shows mean belief ratings. We conducted Bayesian mixed effect linear regressions predicting the ascribed speaker belief rating from the main effect of predicate (reference: BARE), mean-centered prior belief rating, and their interaction, with the maximal random effects structure that allowed the model to converge. There was a significant main effect of predicate and prior, replicating existing findings in [4,5].

Model evaluation. We conducted a Bayesian Data Analysis to estimate the value of three free parameters and used the estimated values to generate predictions. The mixture model captures the empirical results well (see Fig. 1 and 2; 30 training items: R^2=0.78; 6 test items: R^2=0.82): in particular, it captures the main effects of prior and predicates. However, the model does not accurately capture the interaction between prior and predicates, as it under- or overestimates the effect of prior for certain utterances.

Discussion. We provided a pragmatic account for projection inferences, using a computational model to capture the interaction between different modulating factors. The proposed mixture model is able to qualitatively capture the effects of predicates and prior, but fails to predict the interaction between prior and embedded clause type for certain utterances. Future work should further investigate modeling at-issueness effects in projection inferences.

Example sentences

(1) *know-p? / think-p?:* Does John know/think that Julian dances salsa?

(2) *know-not-p? / think-not-p?:* Does John know/think that Julian doesn't dance salsa?

(3) *BARE-p?:* Does Julian dance salsa?

Equations

(a) Literal listener

$$P_{L_0}(b_{SP}|u) \propto \begin{cases} 1 & \text{if } b_{SP} > \theta_u \\ 0 & \text{otherwise} \end{cases} \quad \text{(for positive embedded propositions)}$$

$$P_{L_0}(b_{SP}|u) \propto \begin{cases} 1 & \text{if } b_{SP} < 1-\theta_u \\ 0 & \text{otherwise} \end{cases} \quad \text{(for negated embedded propositions)}$$

$$P_{L_0}(b_{SP}|u) \propto P(1-|b_{SP}-\theta_u|^x) \text{ (for unembedded interrogatives)}$$

$$\text{where } x = \begin{cases} 0.1 & \text{if } |b_{SP}-\theta_u| > 0.1 \\ 2 & \text{otherwise} \end{cases}$$

(b) Pragmatic speaker

$$P_{S_1}(u|b_{SP}) \propto \exp(\alpha \cdot U(u; b_{SP}))$$

$$U(u; b_{SP}) = \ln P_{L_0}(b_{SP}|u) - C(u), \text{ where } C(u) = C_{\text{Neg}}(u) + C_{\text{Embed}}(u)$$

(c) Pragmatic listener

$$P_{L_1}(b_{SP}|u) \propto \underbrace{P_{S_1}(u|b_{SP})}_{\text{speaker model}} \cdot P(q_{MC}|u) + \underbrace{P(b_{SP})}_{\text{prior belief}} \cdot P(q_{CC}|u)$$

$$\text{where } P(q_{MC}|u) + P(q_{CC}|u) = 1$$

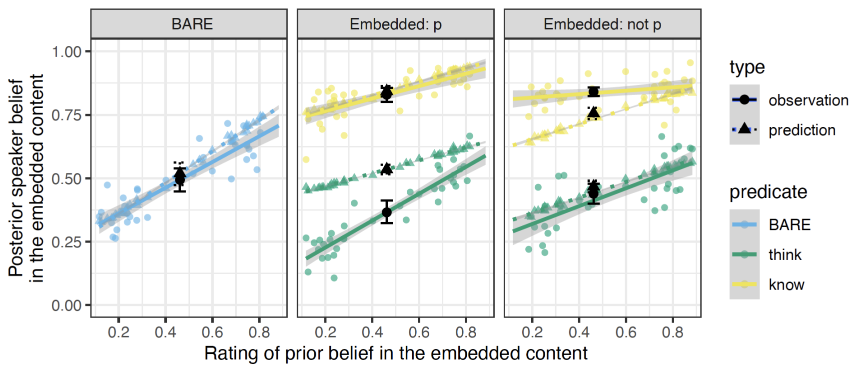

Fig.1 Mean empirical and model predicted speaker belief ratings against the prior ratings. The empirical results are represented by solid lines and the model predictions are presented by dashed lines. Each translucent dot (empirical)/triangle (predicted) represents an item mean, and each large black solid dot/triangle represents the grand mean in that condition.

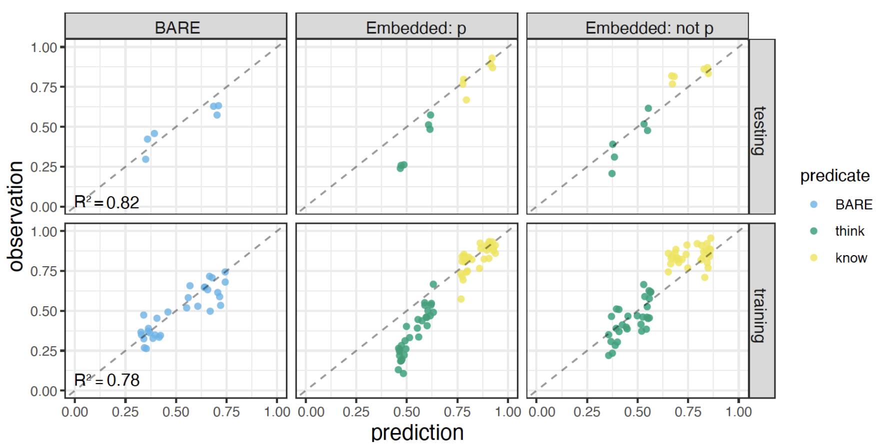

Fig.2 Empirically collected speaker belief ratings against the model-predicted means, faceted by embedded clause type and training vs. testing conditions.

[1] Kiparsky & Kiparsky (1970) *Progress in Linguistics*. [2] Karttunen (1971) *Papers in Linguistics*, [3] Heim (1983) *WCCFL2*, [4] Degen & Tonhauser (2022) *Language*, [5] Degen & Tonhauser (2021) *Open Mind*, [6] Tonhauser et al. (2018) *Journal of Semantics*. [7] Frank & Goodman (2012) *Science*, [8] Goodman & Frank (2016) *Trends in Cognitive Sciences*, [9] Tonhauser & Degen (under review)