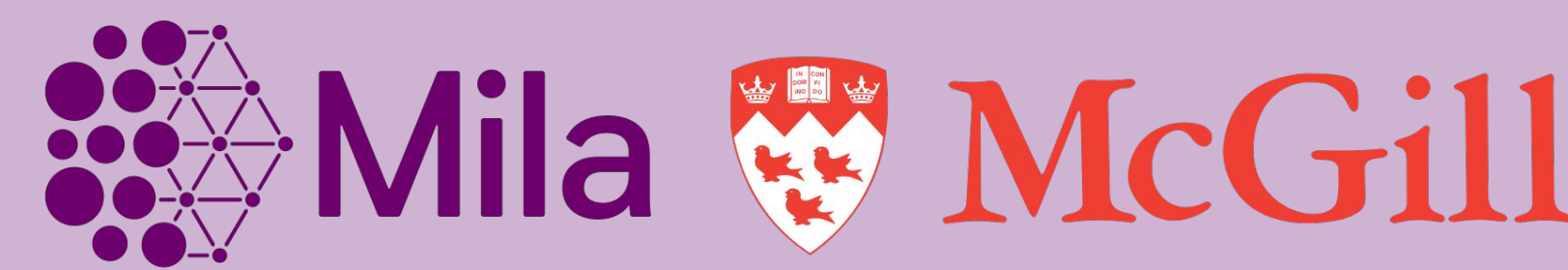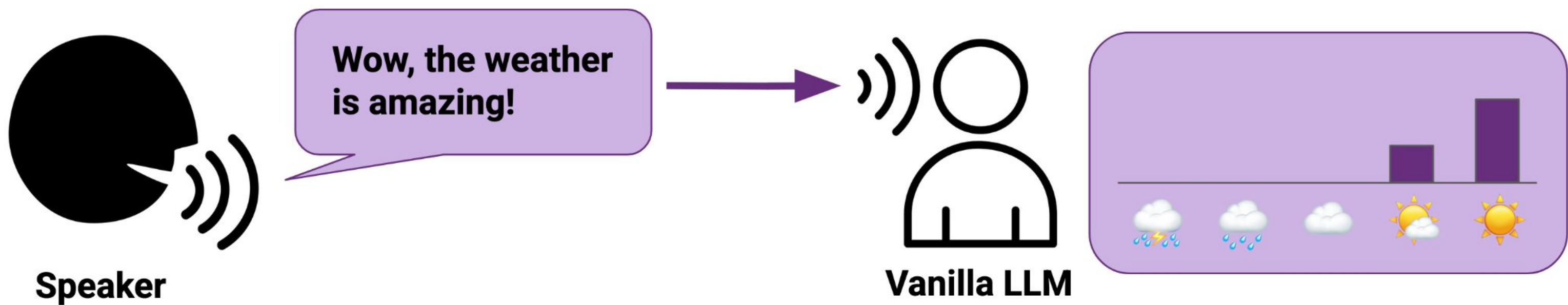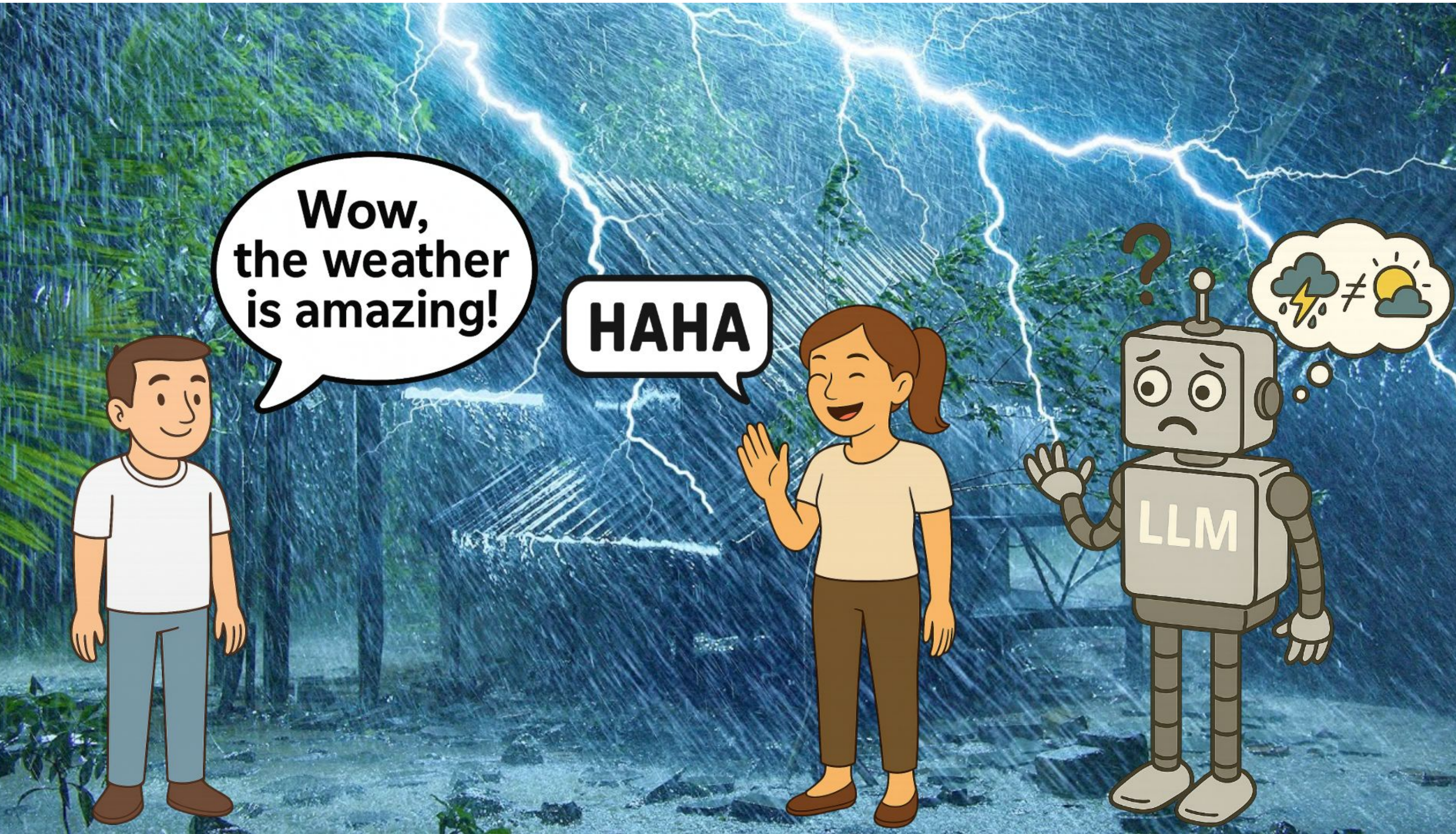# $(RSA)^2$: A Rhetorical-Strategy-Aware Rational Speech Act Framework for Figurative Language Understanding

Cesare Spinoso-Di Piano, David Austin, Pablo Piantanida, Jackie Chi Kit Cheung
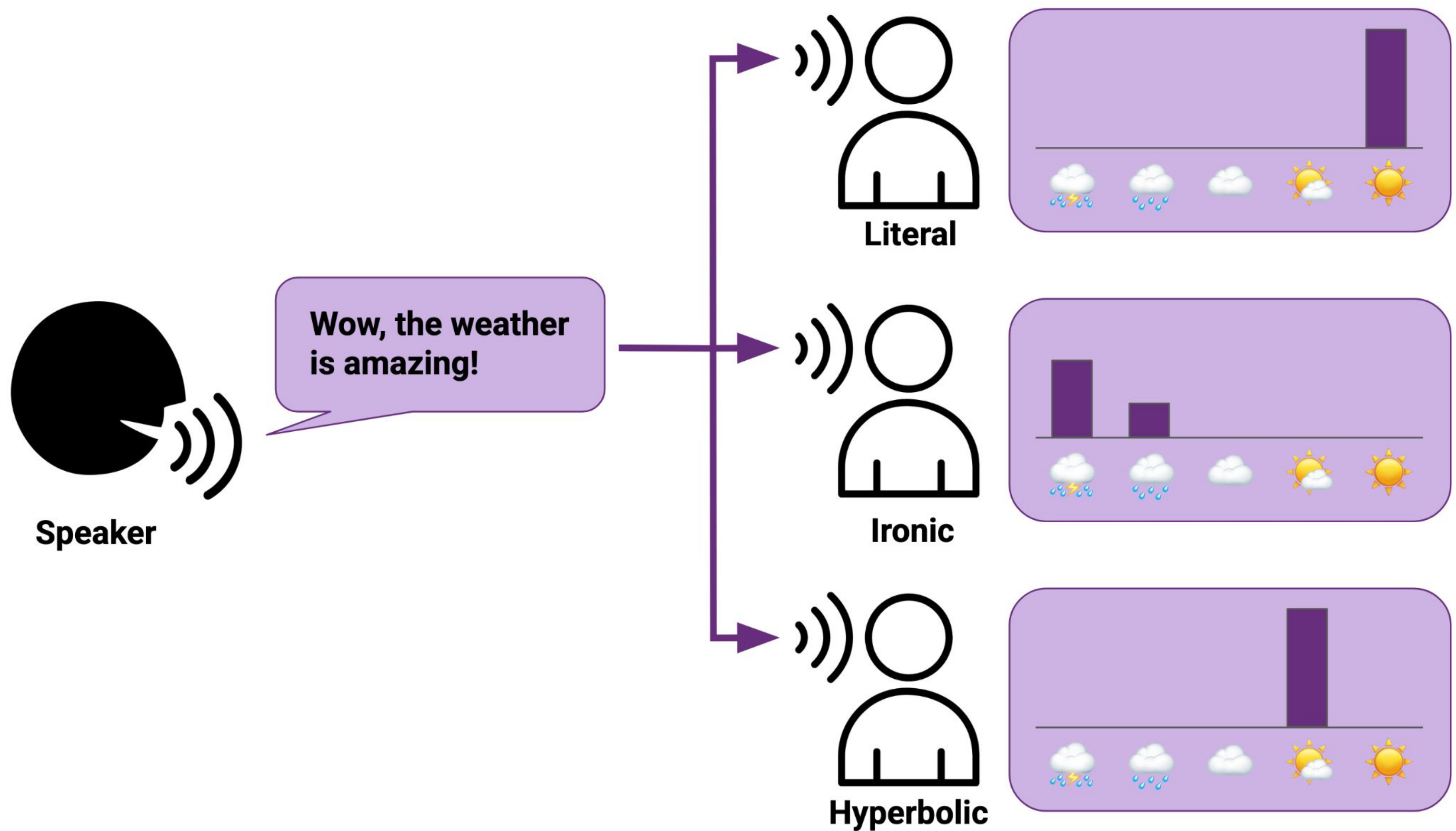
Mila **McGill**

Setting

Vanilla LLMs struggle to interpret non-literal language where listeners' expectations are violated (e.g. irony).

Our Approach

We extend the Rational Speech Act (RSA) framework by modeling different *rhetorical strategies* (e.g. irony, hyperbole) for each pragmatic listener (L1).

Our framework **marginalizes** different **rhetorical listeners** to induce a **distribution over possible meanings** given a context and an utterance.

$$P_{L_1}(m|c,u) = \sum_{r'} P_{L_1}(m|c,u,r') P_{R|CU}(r'|c,u)$$
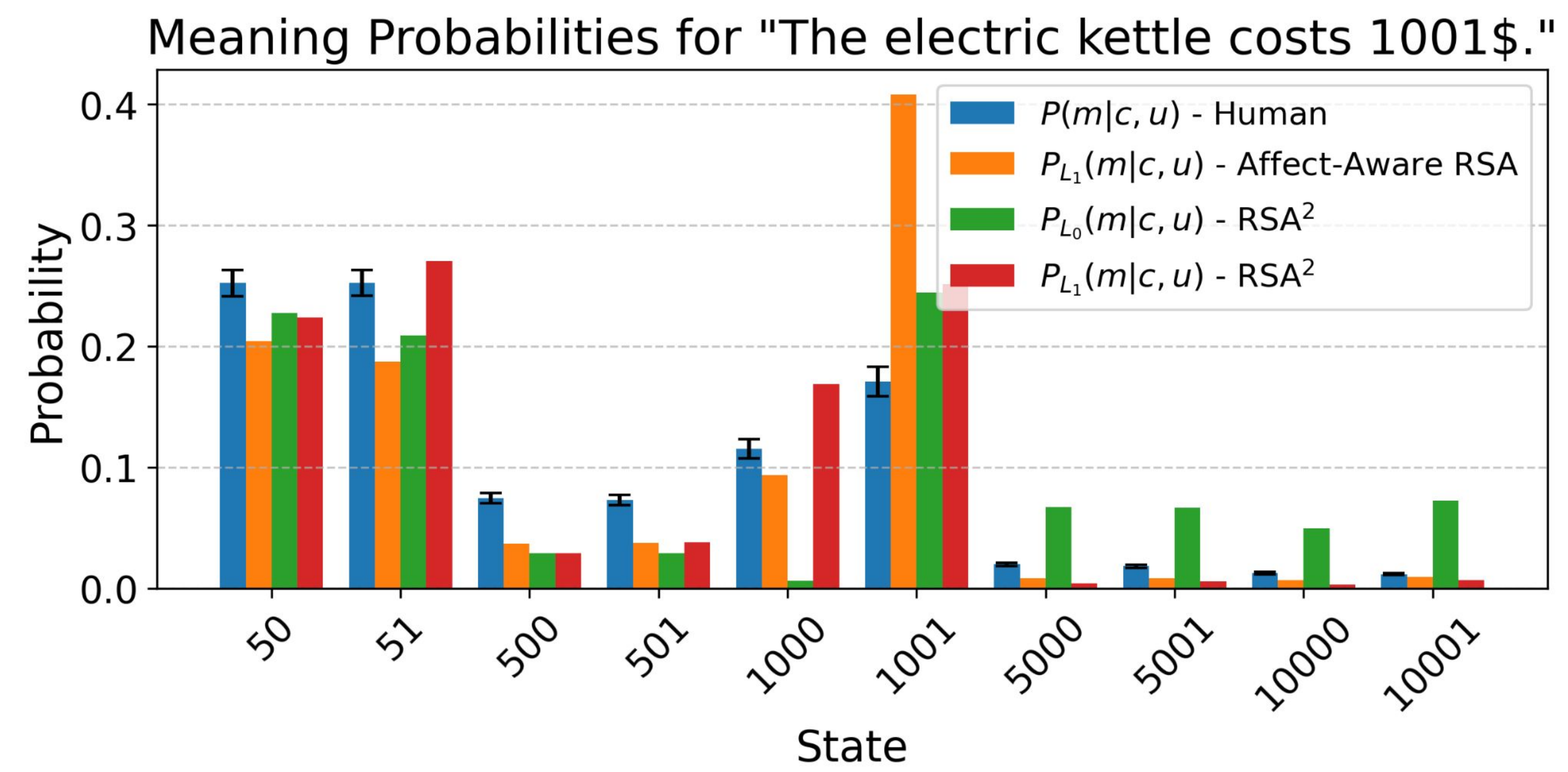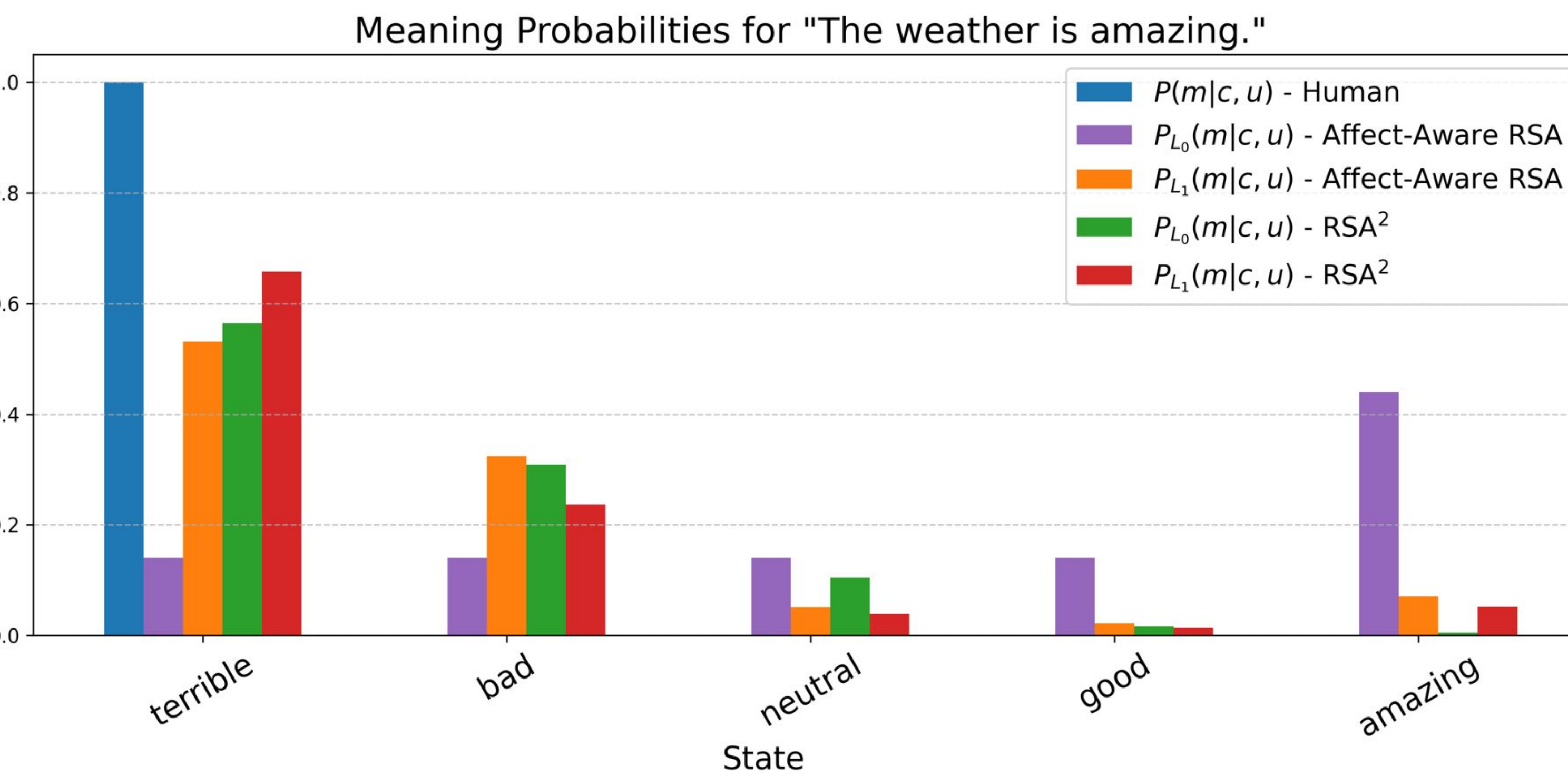
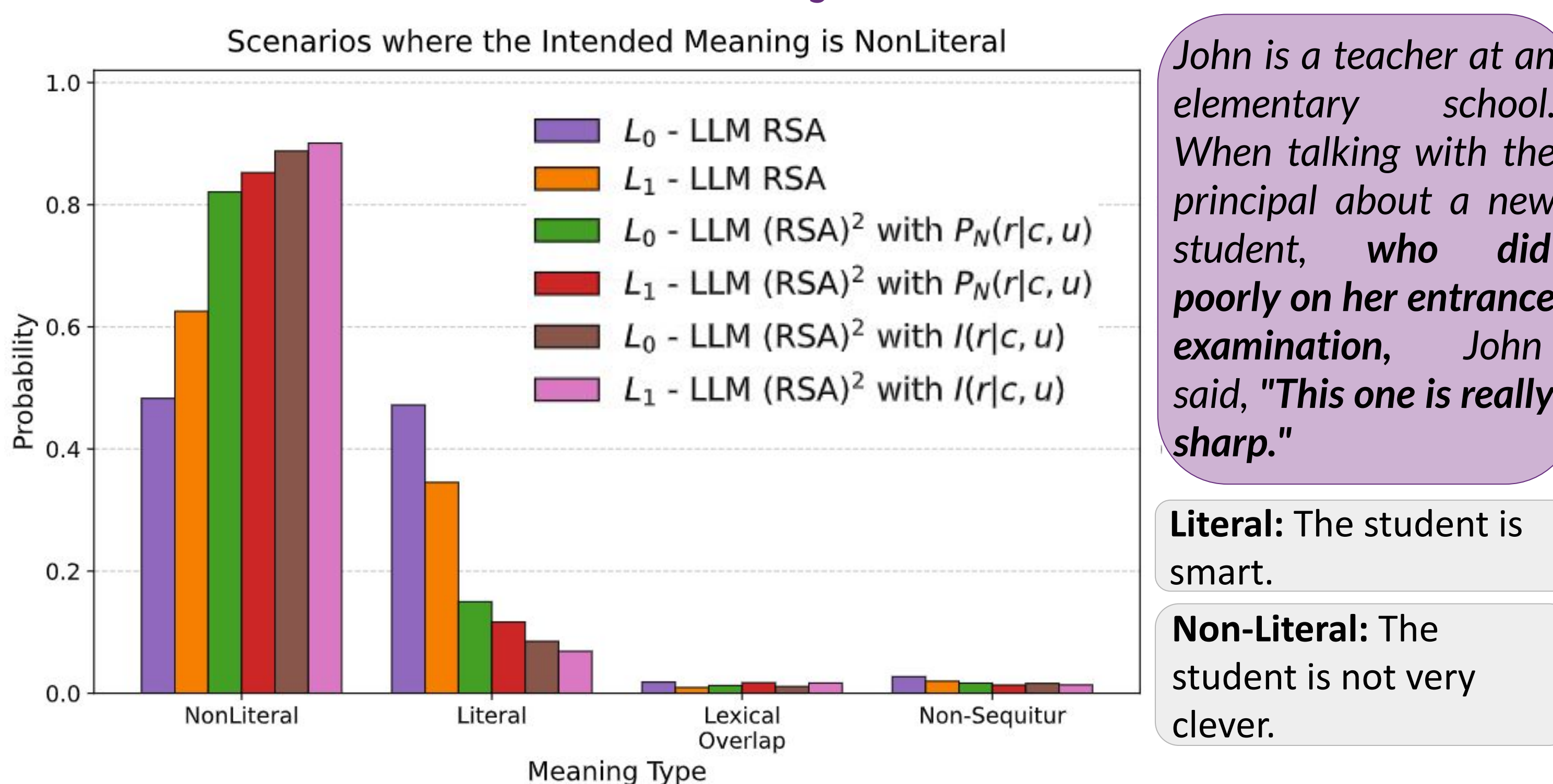| Marginalized Pragmatic Listener | Rhetorical Pragmatic Listener | Rhetorical Strategy Posterior |

$$m - \text{Meaning} \quad c - \text{Context} \quad u - \text{Utterance}$$
$$r - \text{Rhetorical Strategy}$$

Results

Interpretations of irony & hyperbole induced by $(RSA)^2$ (green & red) are closest to human distributions (blue).

Meaning Probabilities for "The weather is amazing."

Meaning Probabilities for "The electric kettle costs 1001$."

$(RSA)^2$ can mitigate LLM bias towards interpreting ironic utterances too literally.

Scenarios where the Intended Meaning is NonLiteral

*John is a teacher at an elementary school. When talking with the principal about a new student, **who did poorly on her entrance examination**, John said, "This one is really sharp."*

Literal: The student is smart.

Non-Literal: The student is not very clever.

Ablation Study: Does pragmatic reasoning help? We isolate the effect of pragmatic reasoning by ablating the utterance $P(u|c)$ & meaning $P(m|c)$ priors.

Average probability of the correct meaning

Model	L_i	w/o $P(m|c)$	w/o $P(u|c)$
LLM RSA	L_1	0.44 (-42.7%)	0.78 (+1.8%)
LLM $(RSA)^2$ with $P_N(r|c,u)$	L_1	0.44 (-44.8%)	0.80 (+0.3%)
LLM $(RSA)^2$ with $I(r|c,u)$	L_1	0.51 (-39.4%)	0.84 (+0.2%)

$(RSA)^2$: A probabilistic pragmatic framework which induces human-like distribution of figurative language! **More findings in paper!**