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

Cesare Spinoso-Di Piano[1,2] **David Austin**[1,2]
Pablo Piantanida[2,3,4] **Jackie Chi Kit Cheung**[1,2,5]
[1]McGill University, [2]Mila - Quebec AI Institute
[3]International Laboratory on Learning Systems (ILLS)
[4]CNRS, CentraleSupélec - Université Paris-Saclay, [5]Canada CIFAR AI Chair, Mila
{cesare.spinoso,david.austin,pablo.piantanida,cheungja}@mila.quebec

1 Introduction

Figurative language, in which expressions are meant to be interpreted non-literally, is ubiquitous in human communication. For example, speakers will often generate utterances which are literally incompatible with the world but which still convey meaning. While the utterance "It is so nice outside." may appear infelicitous when uttered during a blizzard[1], it may still carry meaning if interpreted ironically. As a result, it is critical to be able to account for figurative uses of language — predicting its use and deciphering its intended meaning — when designing computational models of human language and communication.

One of the most popular computational approaches to modeling human communication is the Rational Speech Acts (RSA) framework (Frank and Goodman, 2012). In RSA, an utterance's meaning is interpreted probabilistically by positing the existence of three communicative agents which reason about each other recursively. First, the *literal listener*, denoted $P_{L_0}(m|c,u)$, which reasons about the literal meaning, m, of an utterance u produced in a context c. Second, the *pragmatic speaker*, $P_{S_1}(u|c,m)$, which computes an utterance's expected utility based on its ability to convey the desired meaning to the literal listener. Third, the *pragmatic listener*, $P_{L_1}(m|c,u)$, which interprets the intended meaning of an utterance by reasoning about the pragmatic speaker's likelihood of generating it. Altogether, these three layers of recursive reasoning have been shown to be effective at reducing a listener's uncertainty regarding the interpretation of literally ambiguous utterances such as "some" in scalar implicatures.

While RSA offers a way to refine the interpretation of literally ambiguous utterances, this framework cannot handle non-literal interpretations of utterances. In particular, it can be shown that interpretations of utterances which are not compatible with the literal meaning of an utterance will be assigned zero probability by the pragmatic listener.

As a result, efforts to model human use of figurative language have sought to address this limitation by incorporating the affect of an utterance within the RSA framework. In Kao et al. (2014); Kao and Goodman (2015), the authors propose an affect-aware RSA formulation where an utterance's communicative goal is extended to include both the literal information it conveys as well as the affect it might be communicating (e.g., joy, annoyance, etc.). By explicitly modeling the connection between meaning and affect, the authors are able to assign non-zero probability mass to non-literal meanings because of the association between those meanings and a modeled affect.

2 The Rhetorical-Strategy-Aware RSA Framework

While affect-aware RSA is successful in enabling the RSA framework to obtain non-literal interpretations of utterances, it results in a framework which we believe too tightly couples the notion of non-literal meaning and affect. As a result, we introduce a rhetorical-strategy-aware RSA framework, **(RSA)**^{2}, where a speaker's rhetorical strategy is explicitly modeled and used to interpret an utterance's potentially non-literal intended meaning (Spinoso-Di Piano et al., 2025). To do so, we define a rhetorical strategy variable r along with a corresponding rhetorical function $f_r(c,m,u)$ which generalizes the RSA indicator function $\mathbb{1}_{m\in[\![u]\!]}$. Thus, for example, the *ironic* rhetorical strategy function, $f_r(c,m,u)$, might return 1 if the literal meaning of u is the opposite of m. Crucially, **(RSA)**^{2} enables a pragmatic listener to infer non-literal interpretations of utterances *without* having to model an utterance's intended affect. The **(RSA)**^{2} equations

[1]Blizzards are not typically referred to as being "nice".

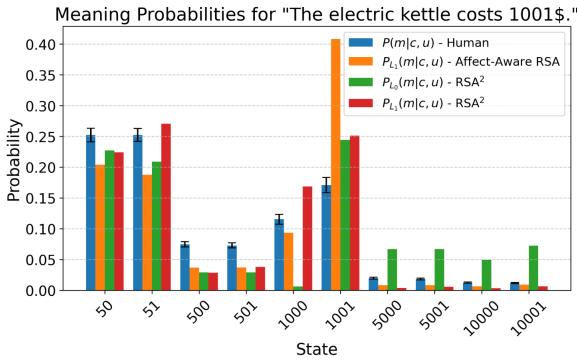

Figure 1: Meaning probability distributions from human judgments along with the listeners of both affect-aware RSA and **(RSA)**^{2} for the utterance "The electric kettle costs 1001 dollars".

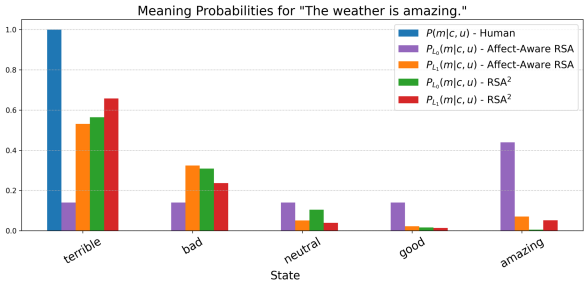

Figure 2: Meaning probability distributions from human judgments along with the listeners of both affect-aware RSA and **(RSA)**^{2} for the utterance "The weather is amazing." in the context of a blizzard.

are the following:

$$P_{L_0}(m|c,u,r) \propto f_r(c,m,u) \cdot P(m|c), \quad (1)$$

$$P_{S_1}(u|m,c,r) \propto P_{L_0}(m|c,u,r) \cdot P(u|c), \quad (2)$$

$$P_{L_1}(m|c,u,r) \propto P_{S_1}(u|m,c,r) \cdot P(m|c), \quad (3)$$

$$P_{L_i}(m|c,u) = \sum_{r'} P_{L_i}(m|c,u,r') \cdot P(r'|c,u). \quad (4)$$

3 Experimental Results

We evaluate our formulation on two previously studied settings: non-literal number price expressions (e.g., "This kettle costs 1001$.") (Kao et al., 2014) and ironic weather utterances (e.g., "The weather is amazing." during a winter blizzard) (Kao and Goodman, 2015). We provide plots for the human, affect-aware RSA and **(RSA)**^{2} meaning distributions on both settings in Figures 1 and 2. In addition, we evaluate predictive accuracy by computing the mean absolute difference (MAD) between the human-derived meaning probability distributions and those generated by the listeners in both affect-aware RSA and **(RSA)**^{2} (Table 1). We note that **(RSA)**^{2} is competitive with affect-aware RSA, surpassing it on the ironic weather utterances dataset. Overall, we believe these results demonstrate that **(RSA)**^{2} can induce listener meaning distributions for figurative language that are at least as compatible with human interpretations as those produced by existing affect-aware RSA models.

4 Conclusion

In conclusion, we presented a rhetorical-strategy-aware formulation of the RSA framework and showed that it is able to achieve human-compatible non-literal interpretations of utterances *without* the need to explicitly motivate its use through affect. Future work in this direction will involve expanding the applicability of **(RSA)**^{2} to other figurative language phenomena and integrating the formulation of **(RSA)**^{2} in modern language technologies such as large language models which we begin to do in Spinoso-Di Piano et al. (2025).

Model	L_i	Non-literal Numbers ↓	Weather Utterances ↓
Affect-Aware RSA	L_0	-	0.2377
	L_1	**0.0436**	0.1278
(RSA)^2	L_0	0.0438	0.1647
	L_1	0.0467	**0.1229**

Table 1: Mean absolute differences between listener meaning distributions, $P_{L_i}(m|c,u), i = 0, 1$, and the human posterior, $P(m|c,u)$, for both affect-aware and **(RSA)**^{2} on both the non-literal numbers and ironic weather utterances datasets.

References

Michael C. Frank and Noah D. Goodman. 2012. Predicting Pragmatic Reasoning in Language Games. *Science*, 336(6084):998–998. Publisher: American Association for the Advancement of Science.

Justine T Kao and Noah D Goodman. 2015. Let's talk (ironically) about the weather: Modeling verbal irony. In *Proceedings of the Annual Meeting of the Cognitive Science Society*, volume 37.

Justine T Kao, Jean Y Wu, Leon Bergen, and Noah D Goodman. 2014. Nonliteral understanding of number words. *Proceedings of the National Academy of Sciences*, 111(33):12002–12007.

Cesare Spinoso-Di Piano, David Austin, Pablo Piantanida, and Jackie Chi Kit Cheung. 2025. **(RSA)**^{2}: A rhetorical-strategy-aware rational speech act framework for figurative language understanding. In *Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (ACL)*. To appear.