Loyal as a Fox: Formalizing the Role of Typicality, Salience, and Alternative Utterances in Metaphor Interpretation

Alexandra Mayn | s8almayn@stud.uni-saarland.de
Department of Computational Linguistics and Phonetics, Saarland University

Introduction & Research Questions

We propose a computational model of metaphor understanding within the Rational Speech Act framework[1].

Our work builds on the model proposed by Kao et al.[2] but addresses the following questions:

- Incorporate **gradient salience** of features
  - Kao et al. define an animal as a vector of 3 binary features
- Role played by **alternative utterances**
- Can metaphors be used "**inversely**"? e.g. "John is an ox" to mean "John is *not* weak"
- What makes a metaphor **easy vs. hard to interpret**?

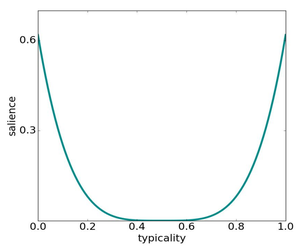
The Model

From Typicality to Salience

We define **salience** of a feature *f* for a referent *o* as follows:

$$salience(f, o) = |typicality(f, o) - \mu)|^{\kappa}$$

where (normalized) typicality is a real number between 0 and 1, and hyperparameters μ and κ can be fit to the data.



Salience for a given typicality value with μ=0.5 and κ=4.

We make the assumption that **both ends** of the typicality spectrum have high salience.

Metaphor Interpretation in RSA

Idea: The listener hears and interprets an utterance of type **"X is Y"**, where X is a male name and Y is an animal category.

- The **literal listener** hears the utterance "John is a shark" and interprets it literally:

$$L_0(c, f|u) = \begin{cases} \text{salience(f,u)} & \text{if } c = u \\ 0 & \text{otherwise} \end{cases}$$

- The **pragmatic speaker**'s choice of utterance is governed by a softmax choice rule based on utility:

$$U(u|f) = log(L_0)$$

$$S_1(u|f) \propto e^{\lambda U(u|f)} = L_0^{\lambda}$$

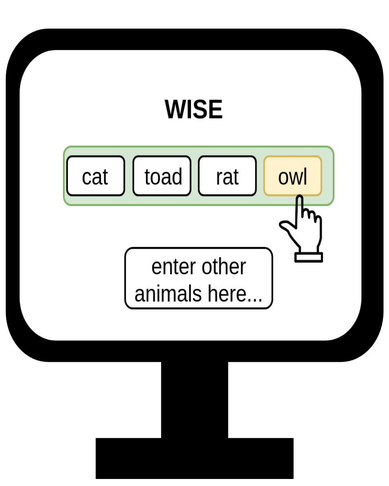
- Finally, the **pragmatic listener** interprets the utterance in the *conversational context*:

$$L_1(c, f|i, u) \propto S_1(u|f) \cdot L_1(f, i|c) \cdot L_1(c)$$

c	category of the referent
f	feature intended by the speaker
i	conversational context, vague or specific
u	speaker's utterance

Behavioral Experiments

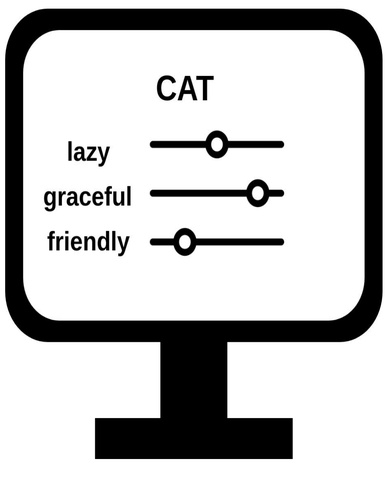
1 Association Elicitation



20 participants* were asked to associate animals with adjectives.

The lists were then post-edited to exclude rarely used animals and high-entropy adjectives, resulting in **28** adjectives and **27** animals.

2 Typicality Elicitation

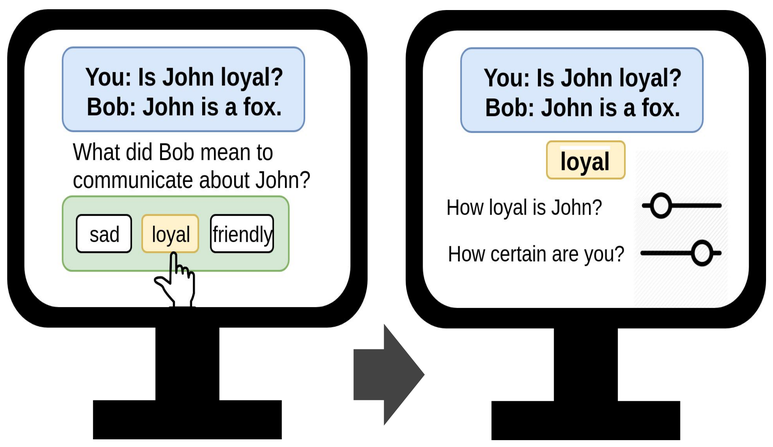


20 participants* will rate the typicality of each feature for each animal on a 7-point scale.

Typicalities will be averaged, normalized and converted to salience.

3 Metaphor Interpretation

Context	Typicality	Example question	Example utterance
vague		"What is John like?"	"He is an *ox*."
specific	high	"Is John loyal?"	"He is a *dog*."
specific	average	"Is John loyal?"	"He is a *dolphin*."
specific	low	"Is John loyal?"	"He is a *fox*."



24 participants* will each see one data point in each specific condition and all data points in the vague condition.

*native English speakers

SCAN ME

Scan this QR-Code to read the paper and/or try out the model!

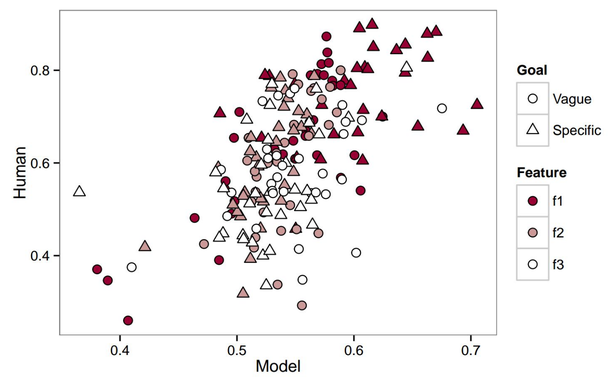
https://github.com/sashamayn/rsa_metaphors

Expected Results

Next Steps:

- obtain typicality priors (Exp. 2)
- run the model and obtain posterior probabilities
- correlate the model's predictions with human judgements from Exp. 3

Kao et al.'s model achieved a fit of r=0.6. We predict that by taking alternative utterances and gradient salience of features into account our model will be able to achieve a significantly better fit.



Predictions of Kao et al.'s model.. They define an animal as a vector of size 3 and evaluate the posterior probability for each feature. The fit of the model is r=0.6.

Predictions for Experiment 3:

- **Correct Inference of Feature:** Low salience (= average typicality) leads to more interpretation errors
- **Certainty:** similarly, lowest for the average typicality condition.
- **High average salience** of features in the vague condition, which would suggest metaphors are used to convey an animal's salient features

Conclusions

Contributions

- A gradient approach to metaphor interpretation
- Taking alternative utterances into account
- Explanation of interpretation difficulty

Limitations & Directions for Future Work

- Extend the model to account for alternative literal utterances
- Speaker's intention may be to communicate multiple features

References

1. Frank, M. C., & Goodman, N. D. (2012). Predicting pragmatic reasoning in language games. *Science*, *336*(6084), 998-998.

2. Kao, J., Bergen, L., & Goodman, N. (2014). Formalizing the pragmatics of metaphor understanding. In *Proceedings of the annual meeting of the Cognitive Science Society* (Vol. 36, No. 36).