

Probabilistic Listener: A Case of Reflexive ziji "self" Ambiguity Resolution in Mandarin



Fengyue Zhao¹, Brian Dillon², Ming Xiang³

Department of Linguistics. ¹ Cornell University, ² University of Massachusetts Amherst, ³ University of Chicago

University of
Massachusetts
Amherst

INTRO

- What mechanisms do people use to resolve ambiguous pronouns? => Bayesian models [1]
- Ambiguous reflexive ziji in Mandarin:** permits local and non-local referents.

[小红]_i 说 [小明]_j 总 把 [自己]_{i/j} 弄糊涂
[Hong]_i says [Ming]_j always BA [ziji]_{i/j} confuses.
[Hong]_i says that [Ming]_j always confuses [self]_{i/j}.

- Previous studies ask this question using cross-sentential pronouns [2]. **What about within-sentence pronoun dependencies?**

METHODS

- Two Bayesian models:

1. Simple Bayesian Model (SBM) [3]:

$$P(\text{ref} | \text{utt}) \propto P(\text{ref}) \times P(\text{utt} | \text{ref})$$

2. Rational Speech Act (RSA) [4]:

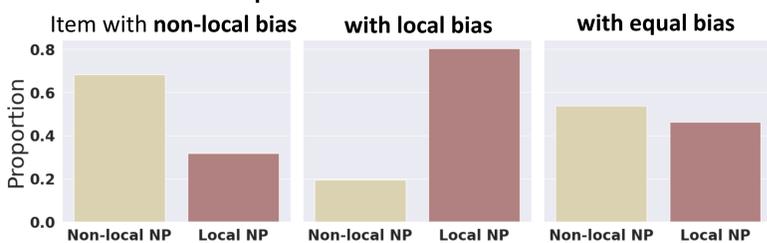
$$P(\text{ref} | \text{utt}) \propto P(\text{ref}) \times \text{Rational Speaker}$$

$$\text{Rational Speaker} = \text{Literal listener} - \text{Cost}(\text{utt})$$

$$\text{Literal listener} = \text{Binding convention} \times P(\text{ref})$$

To estimate the colored parameters:

Experiment 1 ($N_{\text{subj}}=135$, $N_{\text{item}}=30$): an antecedent selection task shows preferences and extent of preferences differ across items.



Experiment 2 ($N_{\text{subj}}=28$, $N_{\text{item}}=30$): a fill-in-the-blank task to test world knowledge bias.

Experiment 3 ($N_{\text{subj}}=65$, $N_{\text{item}}=30$): a selection task, select a pronoun from a drop-down menu to convey a given referent.

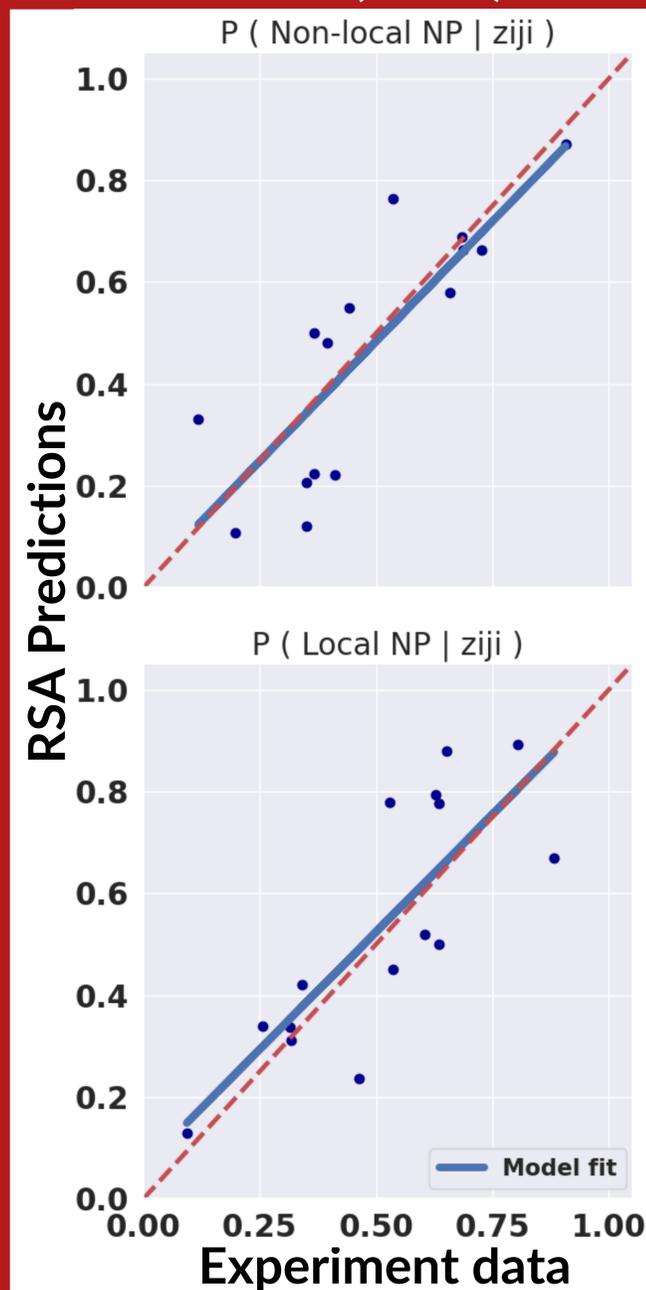
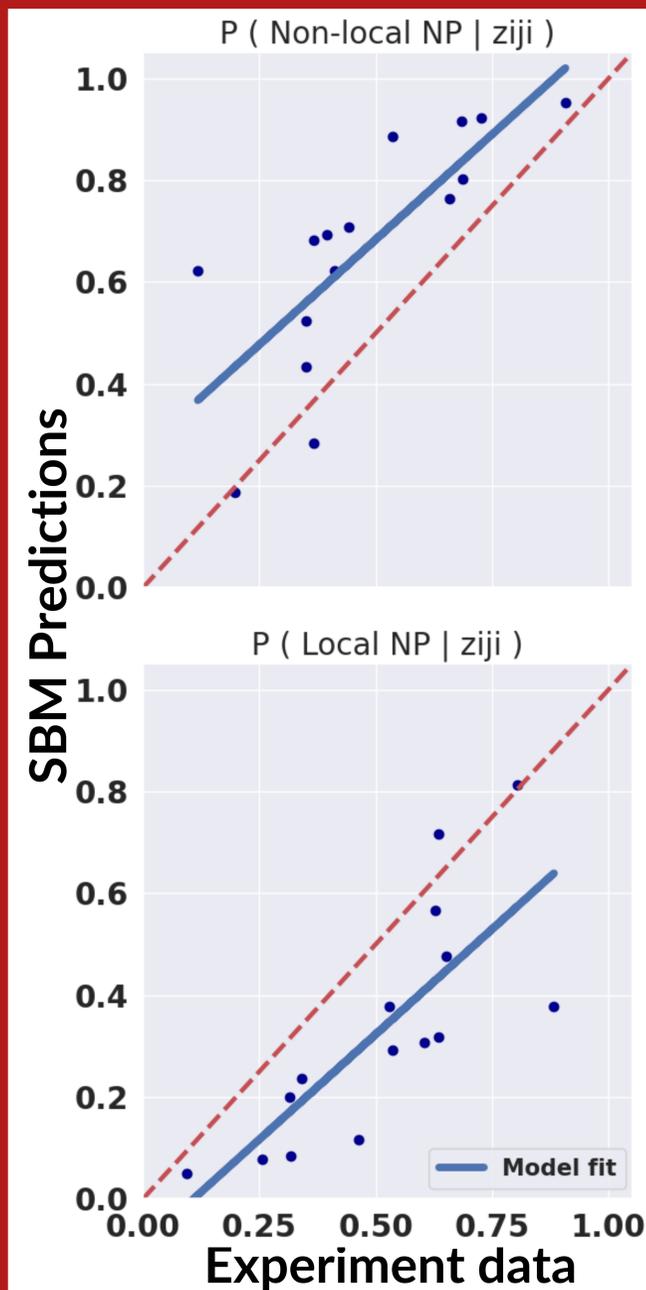
Corpus Study [1]: (Corpus of the Chinese Web)

	ziji (self)	ta (him)	taziji (himself)
Freq	29.54%	70.31%	0.15%

Resolution of long-distance reflexive can be predicted by Bayesian models.

Simple Bayesian Model (SBM)

Rational Speech Act (RSA)



MODEL EVALUATIONS

	R ²	MSE	MAE	P-value
SBM	0.598	0.025	0.138	0.0007
RSA	0.674	0.006	0.048	0.0002

RESULTS

- Item-by-item quantitative analysis shows that **both SBM and RSA can make accurate predictions** for the resolution of ambiguous pronoun ziji.
- SBM **overestimates** the non-local antecedent choices and **underestimates** the local antecedent choices.

DISCUSSIONS

A multilevel recursive reasoning between listeners and speakers coupled with the **explicitly encoded frequency** of pronouns could improve Bayesian model predictions.

CONCLUSION

This study first shows interpretation biases for the within-sentence reflexive differ across items. We then provide a support case for Bayesian models of these non-uniform biases for within-sentence reflexive.

REFERENCES

- [1] Schulz, M., Burnett, H., & Hemforth, B. (2021). A Journal of General Linguistics. [2] Zhan, M., Levy, R., Kehler, A. (2020). PLOS ONE. [3] Kehler, A, & Rohde, H. (2013). Theoretical Linguistics. [4] Frank, M. C., & Goodman, N. D. (2012). Science.



Scan for more info.
Contact at
fz227@cornell.edu