The Effect of Inferred Explanations in a Bayesian Theory of Pronominal Reference

Andrew Kehler and Hannah Rohde UCSD and University of Edinburgh akehler@ucsd.edu; hannah.rohde@ed.ac.uk

UC San Diego

Background

- Bayesian Pronoun Interpretation (Kehler et al. 2008; Kehler & Rohde 2013, Rohde & Kehler 2014):

$$P(\textit{referent} \mid \textit{pronoun}) = \frac{P(\textit{pronoun} \mid \textit{referent})P(\textit{referent})}{\sum_{\textit{referent} \in \textit{referents}} P(\textit{pronoun} \mid \textit{referent})\ P(\textit{referent})}$$

- Two terms in numerator are conditioned on different factors:
  - Production bias $P(\textit{pronoun} \mid \textit{referent})$: topichood (often manifested as an effect of grammatical role)
  - Next-mention bias $P(\textit{referent})$: semantic factors, e.g. coherence relations:

    The boss fired the employee.
    → He was always late. [Explanation]
    → He re-advertised the position. [Occasion]

Experiment: Design

- Participants (n=40) completed passages containing object-biased IC verbs on Mechanical Turk
- 2x2 (RC type x prompt type); 24 stimulus sets and 36 fillers
- Clip art indicated gender (always same for both event participants)

a. The boss fired the employee who was hired in 2002.
_____
[NoExplanationRC, FreePrompt]

b. The boss fired the employee who was embezzling money.
_____
[ExplanationRC, FreePrompt]

c. The boss fired the employee who was hired in 2002.
He _____
[NoExplanationRC, PronounPrompt]

d. The boss fired the employee who was embezzling money.
He _____
[ExplanationRC, PronounPrompt]

- Analyze:
  - Coherence relations (Explanation or Other)
  - Next-mentioned referent (Subject or Object)
  - Form of reference in FreePrompt condition (Pronoun or Other)

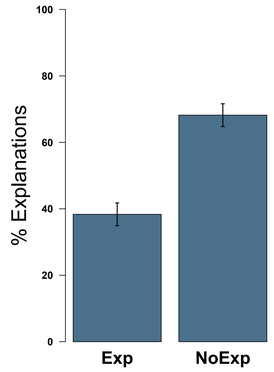
Acknowledgments

Presented at the 27th CUNY meeting at Ohio State. We thank Melodie Yen and Ksenia Kozhukhovskaya for their work in annotating the data.

Prediction 1: Coherence Relations

- Predict a greater percentage of Explanation relations in NoExplanationRC condition than ExplanationRC condition

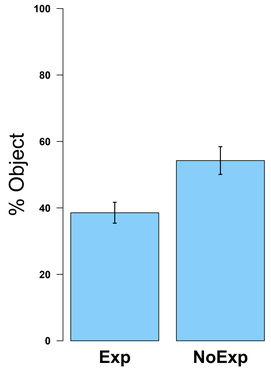
Confirmed (β=2.06; p<.001)



Prediction 2: Next-Mention Biases

- For FreePrompt condition, predict a greater percentage of Next Mentions of Object in NoExplanationRC condition than ExplanationRC condition

Confirmed (β=.720; p<.05)



Prediction 3: Rate of Pronominalization

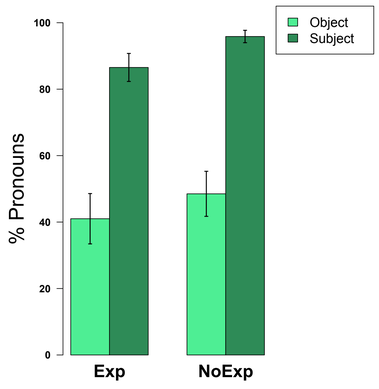
- Predict an effect of grammatical role on pronominalization rate (favoring subjects; FreePrompt condition)

Confirmed (β=4.11; p<.001)

- But no interaction with RC condition

Confirmed: (β=0.12; p=.92)

- Marginal effect of RC condition (β=0.94; p=.078)



Predictions 4 & 5: Pronoun Interpretation

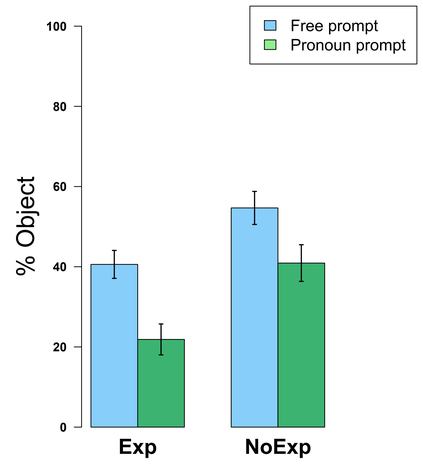
- Predict a greater percentage of object mentions in the NoExplanationRC condition than ExplanationRC condition...

Confirmed (β=1.17; p<.005)

- ...and in the FreePrompt condition than the PronounPrompt condition

Confirmed (β=-1.27; p<.001)

- Marginal interaction (β=0.85; p=.078)
- Effect in PronounPrompt subset only (β=1.46; p<.005)



Comparison with Competing Models

Comparison of Actual Rates of Pronominal Reference to Object (PronounPrompt condition) to the predicted rates for three competing models (using estimates from FreePrompt condition)

	Actual	Bayesian	Mirror	Expectancy
ExplRC	.215	.229	.321	.385
NoExplRC	.410	.373	.334	.542

Conclusions

- Predictions of the Bayesian analysis confirmed:
  - Information inferred from an RC affects interpretation biases in structurally similar passages
  - This information does not affect pronoun production biases, revealing an asymmetry between production and interpretation

Selected References

[1] Jennifer Arnold. 2001. The Effect of Thematic Roles on Pronoun Use and Frequency of Reference Continuation. *Discourse Processes* 21:2, pp. 137-162.
[2] Oliver Bott and Torgrim Solstad. 2012. The mechanics of causal interpretation: Explaining the implicit causality bias. Poster presented at CUNY 25.
[3] Kumiko Fukumura and Roger P. G. van Gompel. 2010. Choosing anaphoric expressions: Do people take into account likelihood of reference? *Journal of Memory and Language* 62, pp. 52-66.
[4] Andrew Kehler, Laura Kertz, Hannah Rohde, and Jeffrey L. Elman. 2008. Coherence and Coreference Revisited. *Journal of Semantics* 25:1, pp. 1-44.
[5] Andrew Kehler and Hannah Rohde. 2013. A Probabilistic Reconciliation of Coherence-Driven and Centering-Driven Theories of Pronoun Interpretation. *Theoretical Linguistics* 39:1-2.
[6] Hannah Rohde. 2008. Coherence-Driven Effects in Sentence and Discourse Processing. PhD Thesis, UC San Diego.
[7] Hannah Rohde and Andrew Kehler. 2014. Grammatical and Information-Structural Influences on Pronoun Production. *Language, Cognition and Neuroscience*, in press.
[8] Julia Simner and Martin J. Pickering. 2005. Planning causes and consequences in discourse. *Journal of Memory and Language* 52, pp. 226-239.