The effect of inferred explanations in a Bayesian theory of pronominal reference

Andrew Kehler (University of California San Diego) & Hannah Rohde (University of Edinburgh)
akehler@ucsd.edu

Background and Study Kehler & Rohde (2013) posit a Bayesian analysis of pronoun use whereby biases towards referents of pronouns ($P(\textit{referent}|\textit{pronoun})$) are determined by combining the prior probability that a referent will get mentioned next ('next-mention' biases; $P(\textit{referent})$) and the likelihood that a pronoun will be used to mention that referent ($P(\textit{pronoun}|\textit{referent})$). Crucially, the factors that condition these terms are different: Next-mention biases are determined primarily by semantically-driven factors (e.g., coherence relations), whereas the production bias is sensitive primarily to information structure and grammatical role (e.g., favoring a greater rate of pronominalizing mentions of subject referents compared to other roles; Rohde 2008, Fukumura & van Gompel 2010, Rohde & Kehler 2013).

We examine the model using data from a passage completion task. The design employs a manipulation that utilizes the fact that relative clauses (RCs) attached to direct objects can be inferred to provide explanations of the matrix event (Rohde et al 2011). RC type is manipulated as in (1a-b):

(1) a. The boss fired the employee who was embezzling money.
    b. The boss fired the employee who was hired early last year.

Although not entailed, (1a) invites the inference that the employee was fired because of the embezzling. Crucially, this inference is not necessary to make the sentence felicitous; (1b) is fine without inferring an analogous causal link between the firing and the hiring. Participants (n=17) were given a context sentence from stimulus sets including alternations like (1a-b) and asked to write a follow-on sentence (24 stimulus sets interleaved with 36 fillers). All stimuli used object-biased implicit-causality (IC2) verbs in the matrix. The continuations were annotated for coherence relation (explanation or other), next-mention (whether the matrix subject referent in the continuation was the subject of the context sentence, the object, or something else), and referential form (pronoun or other). Outcomes were modelled using mixed-effects logistic regression with maximal random effects structure when supported by the data.

Hypotheses and Results Accounts that appeal primarily to surface-level characteristics of the context (e.g. first-mention, subject assignment, grammatical role parallelism) find little to distinguish (1a-b). The Bayesian analysis does predict a difference, however, based on an interconnected sequence of referential and coherence-driven interdependencies. First, it predicts that participants will write fewer explanation continuations in (1a) than (1b), since the RC in (1a) already provides a cause (Simner & Pickering 2005; Kehler et al. 2008; Bott & Solstad 2012). This prediction was confirmed. Second, this difference is predicted to yield a difference in next-mention biases: Since IC2 verbs impute causality to the object, a greater number of explanation continuations for (1b) should lead to a greater number of next-mentions of the object as well. This was also confirmed. Third, the analysis predicts that pronoun production should be unaffected by semantically-driven factors, instead being affected only by grammatical role as discussed above. A model that crossed RC type and grammatical role showed only an effect of grammatical role. Finally, the RC manipulation is expected to affect which referent a pronoun refers to, since $P(\textit{referent}|\textit{pronoun})$ is determined in part by next-mention expectations. Analysis confirms an effect of RC type in the pronoun-only subset of the data. Since this probability is also determined in part by production biases, an effect of grammatical role favoring subjects is also predicted. This was likewise confirmed in a comparison of pronoun vs. other next-mentions.

Conclusion Biases towards referents of pronouns are sensitive to whether or not an implicit explanation can be inferred from context whereas production biases are not, revealing precisely the asymmetry predicted by the Bayesian analysis. We are currently running a second experiment with more participants to further establish the lack of effect of RC type on production biases, and with pronoun prompts in addition to full stop prompts to further establish the effects on pronominal reference.