Interpretation of null pronouns in Mandarin Chinese does not follow a Bayesian model

Suet Ying Lam and Heeju Hwang (The University of Hong Kong)
lsy317@connect.hku.hk

INTRODUCTION There are at least three ways to model a speaker's interpretation of a pronoun. The Mirror Model (MM) argues that the interpretation bias of a pronoun toward a referent is proportional to the likelihood that a pronoun is used to refer to that referent (production bias). The Expectancy Hypothesis (EH, e.g., Arnold, 2001) suggests that the interpretation bias of a pronoun toward a referent is correlated with the likelihood that the referent is re-mentioned regardless of its referential form (next-mention bias). A Bayesian Model (BM, e.g., Kehler et al., 2008) proposes that pronoun interpretation is determined by both the production bias and the next-mention bias. Previous work suggests that BM best explains the interpretations of English pronouns (Rhode & Kehler, 2014), Chinese overt pronouns (Zhan et al., 2020) and German personal pronouns (Patterson et al., 2020). The current study tests the validity of the three models on Mandarin null pronouns. Zhan et al. (2020) assume that the interpretation of Mandarin null pronouns would follow BM just like overt pronouns, given that both are subject-biased. Yet study suggests that Mandarin null pronouns exhibit a much stronger bias toward the subject than overt pronouns (Zhang, 2018). This raises the possibility that the interpretation of null pronouns is less sensitive to the semantically-driven biases such as the next-mention bias and may not be best explained by the models that incorporate the next-mention bias as a predictor of pronoun interpretation, i.e., EH and BM.

EXPERIMENTS We conducted two story-continuation experiments. Exp. 1 aims to replicate previous findings on overt pronouns. Exp. 2 assesses the validity of the models on null pronouns. We included both subject (N1)- and object (N2)-biased verbs to investigate the effect of the next-mention bias, and both implicit causality (IC) and transfer-of-possession (TOP) verbs to examine if the best model generalizes across verb types. We controlled coherence relations by using 'because' for IC verbs and 'so' for TOP verbs to maximize our chance of detecting a potential effect of the next-mention bias. Each experiment contained two versions of prompts: free prompts (to measure the next-mention bias and the production bias) and pronoun prompts (to measure the interpretation bias). We indicated the presence of null pronoun with a verb 'want to/think' in Exp. 2. The below are example stimuli using (1) N1-/N2-biased IC verbs and (2) N1-/N2-biased TOP verbs.

(1) 小玲吓到了/害怕嘉怡，因为 (free)… /因为她 (overt)… /因为想 (null)…
Xiaoling frightened/fears Jiayi, because… /because she… /because Ø wants to/think…

(2) 立强从小刚那里收到了/向小刚寄了一个包裹，所以 (free)… /所以他 (overt)… /所以想 (null)…
Liqiang received/sent a package from/to Xiaogang, so… /so he… /so Ø wants to/think…

MODEL EVALUATION Following Zhan et al. (2020), we compared the predicted data against the observed data on an item-by-item basis, using R^2 (correlation between the predicted and the observed data), and MSE/ACE (prediction error compared to the observed data). Larger R^2 and smaller MSE/ACE imply better performance. Sometimes pronouns did not occur in an item at all, so we used additive smoothing to avoid zero-probability estimates (see Appendix B).

RESULTS For overt pronouns, the mixed effect logistic regression models showed that the interpretation bias was sensitive to both the next-mention bias and the production bias, consistent with BM. Fig. 1 also shows that BM works the best for overt pronouns, whereas EH underestimates the N1-bias and MM overestimates it. In terms of statistical metrics (see Table 1), although EH has a higher R^2, BM has a much smaller prediction error. For null pronouns, however, the next-mention bias affected only TOP but not IC verbs. As can be seen in Fig. 2, the interpretation of null pronouns is strongly N1-biased compared to overt pronouns. Although BM outperforms EH and MM in statistical metrics, it systematically underestimates the N1-bias. Our results suggest that the existing models do not accurately capture the interpretation of null pronouns, and BM may only apply to overt pronouns across languages.

A. Quantitative models used:

- Bayesian: $P(\text{referent}|\text{pronoun}) = \dfrac{P(\text{pronoun}|\text{referent})\ P(\text{referent})}{\sum_{\text{referent}\in\text{referents}} P(\text{pronoun}|\text{referent})\ P(\text{referent})}$
- Mirror: $P(\text{referent}|\text{pronoun}) \leftarrow \dfrac{P(\text{pronoun}|\text{referent})}{\sum_{\text{referent}\in\text{referents}} P(\text{pronoun}|\text{referent})}$
- Expectancy: $P(\text{referent}|\text{pronoun}) \leftarrow P(\text{referent})$

B. Additive Smoothing:

$$\widehat{P}(NP_j)=\frac{\text{Count}(NP_j)+3}{\text{Count}(NP_1)+\text{Count}(NP_2)+2\times 3} \qquad \widehat{P}(\text{pronoun}|NP_j)=\frac{\text{Count}(NP_j \wedge \text{pronoun})+1}{\text{Count}(NP_j)+3}$$

C. Item-by-item quantitative model evaluation collapsing over IC and TOP

Figure 1: **Overt pronoun**

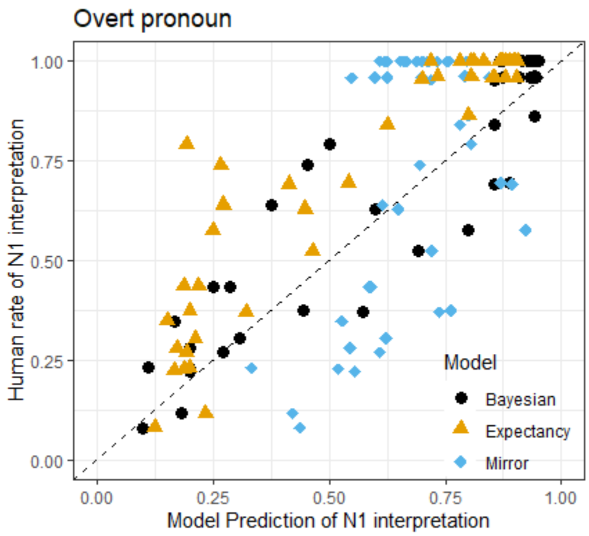

Figure 2: **Null pronoun**

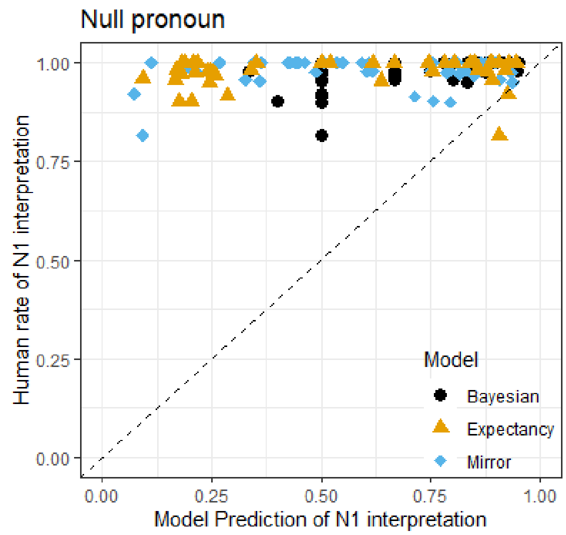

Table 1. Statistical metrics of model evaluation

IC Overt	**BM**	**EM**	**MM**	TOP Overt	**BM**	**EM**	**MM**
R^2	0.950***	**0.952***	0.491***	R^2	0.585***	**0.772***	0.080
MSE	**0.009**	0.016	0.085	MSE	**0.019**	0.060	0.068
ACE	0.804	0.888	**0.521**	ACE	**0.253**	0.575	0.309
IC Null	**BM**	**EM**	**MM**	TOP Null	**BM**	**EM**	**MM**
R^2	**0.253***	0.004	0.045	R^2	**0.488***	0.383**	0.041
MSE	**0.114**	0.297	0.262	MSE	**0.067**	0.197	0.137
ACE	**0.445**	0.858	0.820	ACE	**0.277**	0.606	0.452

*****:** $p < .001$; ****:** $p < .01$; ***:** $p < .05$;

References [1] Arnold (2001). The Effect of Thematic Roles on Pronoun Use and Frequency of Reference Continuation. *Discourse Processes, 31*(2), 137-162. **[2]** Kehler et al. (2008). Coherence and Coreference Revisited. *Journal of Semantics*, *25*(1), 1-44. **[3]** Rohde & Kehler (2014). Grammatical and information-structural influences on pronoun production. *Language, Cognition and Neuroscience, 29*(8), 912-927. **[4]** Patterson et al. (2020). A Bayesian approach to modelling German personal and demonstrative pronouns. Poster presented at the 33rd annual CUNY Human Sentence Processing Conference. **[5]** Zhan et al. (2020). Pronoun interpretation in Mandarin Chinese follows principles of Bayesian inference. *Plos One,15*(8), e0237012. **[6]** Zhang (2018). Interpretational biases and processing of null and overt pronouns in Chinese. Unpublished doctoral dissertation, Konkuk University.