Pronominalization and expectations for re-mention: Evidence from benefactives

Jet Hoek (The University of Edinburgh), Andrew Kehler (University of California, San Diego), & Hannah Rohde (The University of Edinburgh)

Coreference provides a window into speakers' inferences and expectations about relationships that hold across sentences. Different approaches to coreference place different emphasis on the roles of meaning (Winograd 1972; Hobbs 1979) and form (Grosz et al. 1995)—two components which are combined in the Bayesian Model put forward by Kehler et al. (2008). The Bayesian Model, in its strong form, posits the independence of a referent's predictability for re-mention and its likelihood of being mentioned with a pronoun. However, evidence regarding this independence is mixed. Our goal is to use a new context type to test (i) whether predictability influences pronominalization and (ii) whether Bayes' Rule captures the relationship between pronoun interpretation and production.

Models of coreference and independence predictions

Models of pronoun interpretation typically appeal to a notion of salience, but they do so in different ways. According to the Mirror Model (Rohde & Kehler, 2014), listeners base their interpretation decisions on their estimates of the speaker's likelihood to use a pronoun to mention particular referents (Ariel 1990; Gundel et al. 1993). Salience also plays a role in the Expectancy Model (Arnold 2001), whereby listeners' expectations about who will be mentioned next determines their interpretation of a subsequent pronoun. The Bayesian Model incorporates both of these components—an expectation about which referent will be re-mentioned (the *prior*) and an estimate of how likely a speaker is to use a pronoun when re-mentioning a particular referent (the *likelihood*).

(1) $p(referent \mid pronoun)_{\text{INTERPRETATION}} \sim p(referent)_{\text{PRIOR}} * p(pronoun \mid referent)_{\text{LIKELIHOOD}}$

The Bayesian Model is successful in capturing a well-known asymmetry in story continuation results involving items like (2a-b) (Stevenson et al. 1994), whereby a pronoun is preferentially interpreted to refer to one referent (for (2a), *He*→NP2 Bob) but is produced at much higher rates for the other referent (in (2b), pronominalization of NP1 John > pronominalization of NP2 Bob).

(2) a. John scolded Bob. He _____ [pronoun-prompt condition]
b. John scolded Bob. _____ [full-stop condition]

According to (1), pronoun interpretation in (2a) reflects two things that can be measured in (2b): the prior probability that a referent will be re-mentioned (NP2 Bob is favored as the causally implicated referent following *scold*) and the likelihood that a pronoun will be produced (NP1 John is the subject, subjects are often topics, and pronouns are the preferred form for re-mentioning topics in English). When experimental data is used to estimate the probabilities in (1), participants' observed pronoun interpretation behaviour (from items like (2a)) is strongly correlated with the Bayes-derived interpretation values that are computed by estimating the prior and likelihood (from items like (2b)).

In its strong from, the Bayesian Model separates the discourse features that influence the prior and the likelihood: Features related to meaning drive the prior whereas features related to topicality drive the likelihood. This posited independence means that the likelihood of pronominalization for a referent is independent of its prior for re-mention (Rohde & Kehler 2014; Fukumura & van Gompel 2010). However, some recent work shows the likelihood of pronominalization increasing for referents with higher priors (Rosa & Arnold, 2017), raising the possibility of non-independence between the features that are relevant to these two components. Given that the evidence for independence comes primarily from studies with implicit causality verbs (like (2)) whereas non-independence has been shown with transfer-of-possession verbs, we consider a new context type.

Story continuation study with benefactives

A story continuation experiment (N=83) varied prompt type (pronoun vs. full-stop) to test participants' pronoun interpretations (3a), re-mention preferences (3b), and pronominalization rates (3b) in contexts containing a benefactive sentence frame with three event participants. We counterbalanced which potential referents were gender-matched (NP1&NP2, NP1&NP3, NP2&NP3).

(3) a. Adam scolded Diana for Russell. He _____ [pronoun-prompt condition]
b. Adam scolded Diana for Russell. _____ [full-stop condition]

We replicate two previously-established patterns. First, the pronoun prompt yields more NP1 continuations (β=1.52, p<.001; compare Figures 1 & 2). Second, grammatical role influences pronominalization: the subject referent is preferentially re-mentioned with a pronoun (Figure 3). For question (i) on predictability~pronominalization independence, compare Figures 1 and 3. The re-mention rates of NP1 and NP2 do not differ (β=0.22, p=.53) but their pronominalization rates do (β=-3.26, p<.001); conversely, the re-mention rates of NP2 and NP3 differ (β=1.12, p<.001) but their pronominalization rates do not (β=0.19, p<.42). We thus find no evidence of any dependence between predictability and pronominalization.

For question (ii), we are interested in which model yields the best correlations with the observed pronoun interpretation behavior. We used the full-stop continuations to calculate Bayes-derived estimates of *p(referent | pronoun)* via the prior *p(referent)* and likelihood *p(pronoun|referent)*, per (2), as well as estimates for the Expectancy Model (prior alone) and the Mirror Model (likelihood alone; normalized), following Rohde and Kehler (2014). As in earlier work, the Bayesian Model's correlation with observed pronoun interpretation (R^2_{item}=.13, p<.001; $R^2_{participant}$=.060, p<.01) is stronger than that of the Expectancy model (R^2_{item}=-.004, p=.65; $R^2_{participant}$=-.002 , p=.59). In contrast, however, the Mirror model (R^2_{item}=.20, p<.001; $R^2_{participant}$=.063, p<.001) provided the best fit to the observed data. When comparing pronoun interpretation (Figure 2) and production (Figure 3), the patterns are indeed very similar.

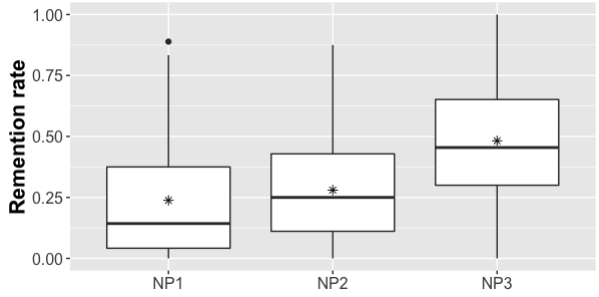

Figure 1. Re-mention rates in the full-stop condition (sums to 1 across referents); asterisks show participant means.

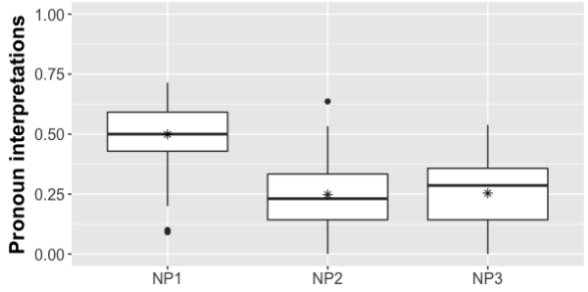

Figure 2. Pronoun interpretation in the pronoun-prompt condition (sums to 1 across referents)

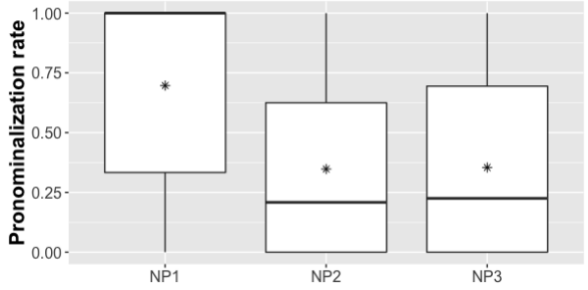

Figure 3. Pronoun rates by referent in full-stop condition.

In sum, our results replicate findings that are best explained with the Bayesian Model, including the independence of predictability and pronominalization. However, the Mirror Model is shown to provide the best fit for the data. We are planning two follow-up studies to determine whether the difference between our results and those of previous work has more to do with the construction type or with the greater number of event participants.

References

Arnold (2001). *Discourse Processes*,137-162.

Ariel (1990). *Accessing noun phrase antecedents*. London: Routledge.

Fukumura & van Gompel (2010). *Jrnl of Memory and Language*, 52-66.

Grosz et al. (1995). *Computational Linguistics* 21, 203–225.

Gundel et al. (1993). *Language*, 274–307.

Hobbs (1979). *Cognitive Science*, 67–90.

Kehler et al. (2008). *Jrnl of Semantics*, 1-44.

Rohde & Kehler (2014). *Language, Cognition, and Neuroscience*, 1-16.

Rosa & Arnold (2017). *Jrnl of Memory and Language*, 43-60.

Stevenson et al. (1994). *Language and Cognitive Processes*, 519–548.

Winograd (1972). *Natural language understanding.* NY: Academic Press.