

## Does Gender Bias Affect the Interpretation of English Sloppy Ellipsis?

Dennis Ryan Storoshenko, University of Calgary & Jesse Weir, Simon Fraser University

**Issue** English reflexives are well-known to be sloppy under VP ellipsis:

(1) Tom hurt himself, and Josh did too.

(1) typically only means that Josh also hurt himself, not allowing the strict reading where Josh hurt Tom. Following Kratzer (2009), the object position is saturated by a variable without  $\phi$ -features, predicting that the sloppy reading should survive a  $\phi$ -mismatch of gender:

(2) Tom hurt himself, and Hannah did too.

We use a contextual felicity task to test this. However, we ask further questions regarding gender. Specifically, will third person reflexives opposing on a gender binary behave differently, and will this interact with the gender identities of the participant pool?

**Study Design** Participants read a short context, and are asked to rate on a 7pt Lickert scale how well a target sentence matches the context. There are two levels of context (*strict* vs *sloppy*). There are also four types of target sentence, as described in the table below. This yields a 2x2x2 design for trial items, with each participant seeing six trials from each condition in a Latin-square design.

	$\phi$ -Match	$\phi$ -Mismatch
<i>herself</i>	Rose hurt herself and Hannah did too	Rose hurt herself and Josh did too
<i>himself</i>	Tom hurt himself and Josh did too	Tom hurt himself and Hannah did too

Each participant sees 48 trial items, intermixed with 60 distractor items. All trial item verbs were previously tested in a norming study with explicitly gendered agents (e.g. “the boy”, “the mother”) to ensure no gender preference exists for a given verb. Participant gender identity is also a controlled independent variable. Recruiting was restricted to participants of binary gender to match with the  $\phi$ -features of the trial items, making the final design 2x2x2x2. Using Prolific Academic, 24 Female and 24 Male identifying participants (all native English speakers) were recruited. The study was completed online using PsychoPy3 (Pierce et al., 2019).

**Results** Ratings are analyzed using lme4 and lmerTest (Bates et al., 2015; Kuznetsova et al., 2017). We find that across all trials, sloppy conditions have significantly higher means (5.69) than strict ones (3.84).  $\phi$ -feature match vs mismatch (overall means 4.77 and 4.76, respectively) is not a significant predictor of ratings. Within the sloppy conditions, the anaphors behave differently from each other. There is a significant interaction between participant gender identity and the anaphor’s  $\phi$ -features: female participant ratings in sloppy *himself* trials have a mean of 6.04, while male participants have a mean of 5.39. Anaphor *herself* shows no such interaction in sloppy conditions, and there are no parallel interactions in strict conditions. Our best-fitting model is given in (3).

(3) Rating  $\sim$  AnaphorGender \* ParticipantGender \* Context +  
 (1 + Context | ParticipantID) + (1 + ParticipantGender | ItemID)

**Discussion** Finding (i), that sloppy readings are preferred over strict ones is not surprising. Finding (ii), the different results for *herself* vs *himself* in sloppy contexts is surprising as it seems to challenge the Kratzerian analysis of reflexives as underlyingly featureless variables. However, Finding (iii), the difference in participant gender, introduces a new complication. If participant gender were not the locus of the difference between *herself* and *himself*, we might question Kratzer’s analysis, but as the participants’ own gender identities are influencing the ratings, we are more likely detecting a performance effect. The methodological implications remain profound: studies

targeting features not normally treated as sociolinguistic, such as the acceptance of sloppy readings under ellipsis, must also pay close attention to participant demographics and trial item content. Our results show that interaction between these factors can significantly shift ratings.

## References

- Bates, Douglas, Martin Maechler, Ben Bolker, and Steve Walker. 2015. Fitting linear mixed-effects models using *lme4*. *Journal of Statistical Software* 67:1–48.
- Kratzer, Angelika. 2009. Making a pronoun: Fake indexicals as windows into the properties of pronouns. *Linguistic Inquiry* 40:187–237.
- Kuznetsova, Alexandra, Per B. Brockhoff, and Rune H.B. Christensen. 2017. lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software* 82:1–28.
- Pierce, Jonathan, Jeremy R. Gray, Sol Simpson, Michael MacAskill, Richard Höchenberger, Hiroyuki Sogo, Erik Kastman, and Jonas Kristoffer Lindelov. 2019. PsychoPy2: Experiments in behavior made easy. *Behavior Research Methods* 51:195–203.