Lectal Variants as Linear Predictors: A Case Study

Simon Masloch, Ruhr-Universität Bochum
simon.masloch@rub.de

25 February, 2026. Workshop *More than just noise – Detecting patterns in acceptability judgment data* at the 48th Annual Conference of the German Linguistic Society (DGfS), University of Trier

Abbreviations **BF**: Bayes Factor, **EO**: experiencer-object, **O(bj)**: object **OS**: object before subject, **S(ubj)**: subject, **SO**: subject before object, **WCO**: weak crossover, **ZOIB**: zero-one-inflated Beta
$\hat{\cdot}$: mean of posterior as estimated in model

Scripts The scripts for the analyses presented in this talk are available from *https://osf.io/zsxdh*

1 Introduction

- Participants in linguistic experiments differ.
- Differences may be interesting, e.g.,
  - … in themselves: In the most extreme case one may need different grammars for the phenomenon at hand and a plausible account of the differences between them.
  - … because properties of the items not directly under investigation may lead to a rejection by some participants, obscuring the patterns one is mainly interested in.
- Proposal: Components for statistical models in which potential differences between participants are directly reflected in assuming that fundamentally different linear predictors are relevant for participants belonging to different groups.
- Purposes:
  - Get a clearer picture of patterns where some participants may reject items for independent reasons.
  - Measure share of participants with different grammar (or at least acceptability pattern)

2 Binding Data

2.1 Example Data: (Masloch & Poppek & Kiss 2025: (MPK))

- The main example data for this talk will come from (Masloch & Poppek & Kiss 2025, MPK)
- Acceptability judgment study on reflexive binding into the subjects of German EXPERIENCER-OBJECT (EO) verbs. A test item in its two ordering conditions:

(3) (Masloch & Poppek & Kiss 2025: 206, without judgment)

a. Es ist offensichtlich, dass [das Gerücht [über sich$_1$]]$_S$ [den Professor$_1$]$_O$ genervt hat.
it is obvious that the.NOM rumour about REFL the.ACC professor annoyed has

b. Es ist offensichtlich, dass [den Professor$_1$]$_O$ [das Gerücht [über sich$_1$]]$_S$ genervt hat.
'It is obvious that the rumour about himself annoyed the professor.'

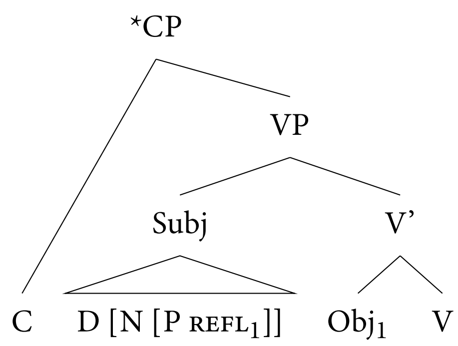
- 2 × 2 Acceptability / "naturalness" rating study (5-point scale)
  - Order: subject before object (SO), object before subject (OS)
  - Case: of object. Accusative or dative.
- Every item in both order conditions, but participants saw one only.
- Each items contains either dative or accusative EO verb, participants see both kinds of verbs.
- Materials:
  - 8 test items containing accusative-object EO verb, 8 containing dative EO verb
  - 64 (related and unrelated) filler items, of which 6 calibration, 16 control, 10 attention

Hypotheses and Predictions MPK:

- Reflexive embedded in subject, object only possible antecedent.

(4) Main Hypothesis MPK (p. 209)
In the German midfield, the object of an experiencer-object verb cannot bind a reflexive embedded in a subject preceding it.

(5)



- Depending on views on German clausal syntax, the syntax of EO verbs and reflexive binding, various different predictions possible, but not relevant for this talk (see MPK on that).

Results MPK:

- 90 participants, 48 surveys entered analysis
- Empirical distribution in figure 1

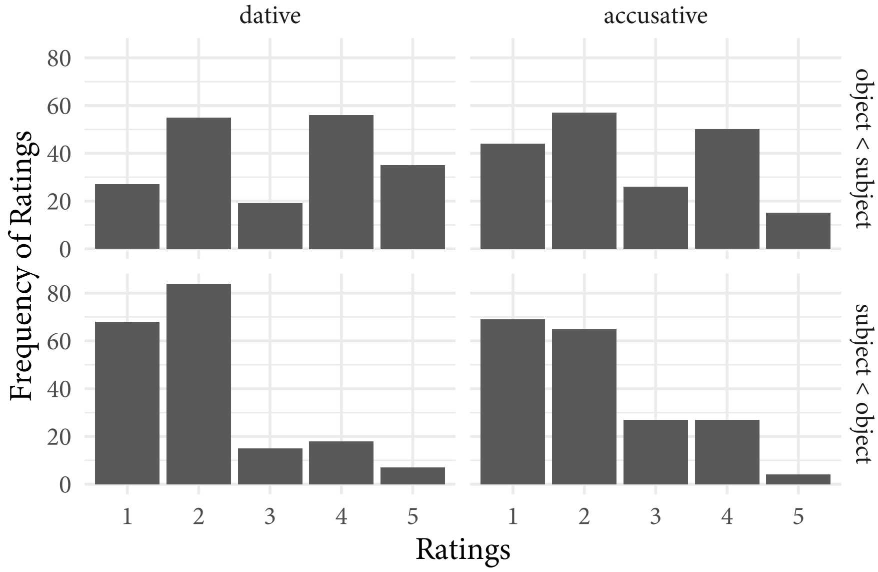


Figure 1: Distribution of responses for the binding study from MPK, reproduced from MPK, p. 211. Choice options had natural language labels, 1 stands for *vollkommen unnatürlich* 'completely unnatural' here, 5 for *vollkommen natürlich* 'completely natural'.

2.2 The Standard Model

In MPK, we mainly used a Bayesian cumulative generalised linear mixed model with logit link and flexible thresholds. R / Brms formula:

```
ANSWER ~ case * order + (1 + case * order | participant) + (1 + order | item)
```

- Factors sum-coded: *dative* and *OS* as 1, *accusative* and *SO* as −1:
  - β_{Order}: overall effect of Order
  - β_{Case}: overall difference between accusative and dative
  - $\beta_{\text{Order}\times\text{Case}}$: positive value would correspond to preference for normal order irrespective of other factors including binding constraints (see MPK)
- Bayesian Model: Parameters are random variables ⇒ One can talk about the credibility of different values
- Prior reflects how likely the parameter values are considered to be before taking data into account
- Posterior captures updated beliefs after seeing the data. Described here by estimate ($\hat{\cdot}$) and 95 % credible interval.

Case \ Order	*OS*	*SO*
dat	$\beta_{\text{Order}} + \beta_{\text{Case}} + \beta_{\text{Order}\times\text{Case}}$	$-\beta_{\text{Order}} + \beta_{\text{Case}} - \beta_{\text{Order}\times\text{Case}}$
acc	$\beta_{\text{Order}} - \beta_{\text{Case}} - \beta_{\text{Order}\times\text{Case}}$	$-\beta_{\text{Order}} - \beta_{\text{Case}} + \beta_{\text{Order}\times\text{Case}}$

Figure 2: Conditions binding study MPK with linear predictors for them given the sum-coding used in M_{standard}

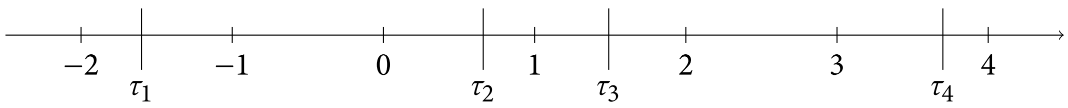


Figure 3: Thresholds partition a latent continuous variable in a cumulative model. The values for the τs here are the estimates from $\mathrm{M}_{\mathrm{standard}}$.

- In equations 1–3, $\mathrm{M}_{\mathrm{standard}}$ is provided in a different notation. n: data point number, $part[n]$, $item[n]$: participant / item of data point. In the following, especially 2 will be relevant. For the following models, I will concentrate on the conceptually relevant parts.

$$
\begin{aligned}
\textsc{Rating}_n &\sim Categorical\,(\xi_{n,1}, \dots, \xi_{n,5}) \\
\xi_{n,i} &= \begin{cases} logit^{-1}(\tau_i - \eta_n) & \text{if } i = 1 \\ logit^{-1}(\tau_i - \eta_n) - logit^{-1}(\tau_{i-1} - \eta_n) & \text{if } 1 < i < 5 \\ 1 - logit^{-1}(\tau_{i-1} - \eta_n) & \text{if } i = 5 \end{cases}
\end{aligned} \tag{1}
$$

$$
\begin{aligned}
\eta_n =& u_{part[n],1} + w_{item[n],1} + \\
& \textsc{order}_n \times (\beta_{\textsc{order}} + u_{part[n],2} + w_{item[n],2}) + \\
& \textsc{case}_n \times (\beta_{\textsc{case}} + u_{part[n],3}) + \\
& \textsc{case}_n \times \textsc{order}_n \times (\beta_{\textsc{case:order}} + u_{part[n],4})
\end{aligned} \tag{2}
$$

$$
\begin{aligned}
\tau_1 &\sim \mathcal{N}(-3, 7.5), \tau_2 \sim \mathcal{N}(-1, 7.5), \tau_3 \sim \mathcal{N}(1, 7.5), \tau_4 \sim \mathcal{N}(3, 7.5) \\
\beta_{\textsc{order}}, \beta_{\textsc{case}}, \beta_{\textsc{case:order}} &\sim \mathcal{N}(0, 4) \\
\begin{pmatrix} w_{i,1} \\ w_{i,2} \end{pmatrix} &\sim \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \Sigma_w\right) \\
\Sigma_w &= \begin{pmatrix} \sigma^2_{w_1} & \rho_w \sigma_{w_1} \sigma_{w_2} \\ \rho_w \sigma_{w_1} \sigma_{w_2} & \sigma^2_{w_2} \end{pmatrix} \\
\sigma_{w_1}, \sigma_{w_2} &\sim Student\_t_+(3, 0, 2.5) \\
\begin{bmatrix} 1 & \rho_w \\ \rho_w & 1 \end{bmatrix} &\sim LKJcorr(1) \begin{pmatrix} u_{i,1} \\ u_{i,2} \\ u_{i,3} \\ u_{i,4} \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \Sigma_u\right) \\
\Sigma_u &= \begin{pmatrix} \sigma^2_{u_1} & \rho_{u1,u2}\sigma_{u_1}\sigma_{u_2} & \rho_{u1,u3}\sigma_{u_1}\sigma_{u_3} & \rho_{u1,u4}\sigma_{u_1}\sigma_{u_4} \\ \rho_{u1,u2}\sigma_{u_1}\sigma_{u_2} & \sigma^2_{u_2} & \rho_{u2,u3}\sigma_{u_2}\sigma_{u_3} & \rho_{u2,u4}\sigma_{u_2}\sigma_{u_4} \\ \rho_{u1,u3}\sigma_{u_1}\sigma_{u_3} & \rho_{u2,u3}\sigma_{u_2}\sigma_{u_3} & \sigma^2_{u_3} & \rho_{u3,u4}\sigma_{u_3}\sigma_{u_4} \\ \rho_{u1,u4}\sigma_{u_1}\sigma_{u_4} & \rho_{u2,u4}\sigma_{u_2}\sigma_{u_4} & \rho_{u3,u4}\sigma_{u_3}\sigma_{u_4} & \sigma^2_{u_4} \end{pmatrix} \\
\sigma_{u_1}, \sigma_{u_2}, \sigma_{u_3}, \sigma_{u_4} &\sim Student\_t_+(3, 0, 2.5) \\
\begin{bmatrix} 1 & \rho_{u1,u2} & \rho_{u1,u3} & \rho_{u1,u4} \\ \rho_{u1,u2} & 1 & \rho_{u2,u3} & \rho_{u2,u4} \\ \rho_{u1,u3} & \rho_{u2,u3} & 1 & \rho_{u3,u4} \\ \rho_{u1,u4} & \rho_{u2,u4} & \rho_{u3,u4} & 1 \end{bmatrix} &\sim LKJcorr(1)
\end{aligned} \tag{3}
$$

2.3 Results M_{standard}

Model fit with Stan (Stan Development Team 2024) via brms (Bürkner 2017) in R (R Core Team 2023) and sensitivity analyses etc. conducted in MPK. Now re-implemented in PyMC (Martin & Kumar & Lao 2021) for comparison with other models.

- In MPK, we used Bayes Factors (BFs) for hypothesis testing: BF: $\frac{P(Data\,|\,Model_1)}{P(Data\,|\,Model_2)}$
  Summary of evidence provided by data for one model over the other.
- BF_{10}: BF of model containing effect vs. model where it is set to 0.
- Values > 3 moderate evidence, > 10 strong evidence (Jeffreys 1939)
- Population-level effects:
  - $\hat{\beta}_{\text{ORDER}} = 0.79\ [0.47, 1.12]$, $BF_{10} = 126.7$: *Ratings are better if binding not backward*
  - $\hat{\beta}_{\text{CASE}} = 0.16\ [-0.24, 0.57]$, $BF_{10} = 0.074$: Evidence against effect of CASE.
  - $\hat{\beta}_{\text{ORDER}\times\text{CASE}} = 0.23\ [-0.06, 0.52]$, $BF_{10} = 0.138$. Evidence against interaction effect.
- *Participants differ strongly*: intercepts: $\hat{\sigma}_{u_1} = 1.57\ [1.22, 2]$ (1.57 is large: Two SDs effectively span the whole scale), ORDER: $\hat{\sigma}_{u_2} = 0.5\ [0.26, 0.75]$,
- For comparison: $\hat{\tau}_1 = -1.6\ [-2.23, -0.98]$, $\hat{\tau}_2 = 0.66\ [0.05, 1.28]$, $\hat{\tau}_3 = 1.49\ [0.87, 2.12]$, $\hat{\tau}_4 = 3.74\ [3.06, 4.46]$

2.4 Unattractive Aspects of M_{standard}, Problems N P refl

- Variation between participants assumed effectively spans the whole scale
- Even in the condition we would expect to be grammatical, estimate is mediocre
- Explorative investigation in MPK shows: Participants with individual intercepts which would mean that they must consider everything bad do *not* reject all filler items
- Possible reason: Test items contained [D [N [P REFL]]] structures (6), which are problematic in that they often do not sound very natural.

(6) das Gerücht über sich
    the rumour about REFL

- We took great care to use only N P REFL sequences sounding natural.
- Still: May be a problem and some of the participants rejecting all test items reject the four fillers containing N P REFL sequences mostly accepted by the others, too (e.g. (7))!

(7) Maryam hat behauptet, dass sie eine Tante von sich in Venedig getroffen hat.
    Maryam has claimed that she a aunt of REFL in Venice met has
    'Maryam claimed that she met an aunt of her's in Venice.'

2.5 Idea Alternative Model

- What if there is lectal variation and some speakers simply reject [N [P REFL]] altogether?
- M_{standard} has means to account for this via the the group-level effects (⇒ Large estimates for their SDs!)

- But we can build models that reflect this assumption directly: Each participant may either belong to a group rejecting all test items or to the "normal" group.
- Expected advantage: Responses of participants rejecting everything do not influence main parameters of interest anymore.

2.6 Lectal Variants as Linear Predictors

Straightforward implentation:

- π: Overall rate of participants who consider all test items unnatural. We may assume an uninformative prior like $Beta\,(1, 1)$
- $b \sim Bernoulli\,(\pi)$: vector of Booleans, indicating for each participant which group they belong to
- $\eta_n = \begin{cases} \eta_{1,n} & \text{if } b_{part[n]} = 1 \\ \eta_{2,n} & \text{if } b_{part[n]} = 0 \end{cases}$
- $\eta_{1,n} = u_{part[n],1} + w_{item[n],1} + \alpha_{bad}$ (to be revised immediately)
- $\eta_{2,n} = u_{part[n],1} + w_{item[n],1} +$ (to be revised immediately)
$$\begin{aligned} &\textsc{order}_n \times (\beta_{\textsc{order}} + u_{part[n],2} + w_{item[n],2}) + \\ &\textsc{case}_n \times (\beta_{\textsc{case}} + u_{part[n],3}) + \\ &\textsc{case}_n \times \textsc{order}_n \times (\beta_{\textsc{case:order}} + u_{part[n],4}) \end{aligned}$$

Including Filler Item Data:

- The information that not everything is rejected is contained in the filler items.
- Only data from fillers in narrow sense included (no attention and control items to avoid double use as they were used to in eligibility checks, no calibration as participants may have needed them to familiarise themselves with the setting), no medium acceptability fillers
- Data for filler items treatment-coded (1 if (un)acceptable filler, 0 for the other and test items). Other effects 0 for fillers. Result: ORDER, CASE, interaction 0 for fillers, so the corresponding summands become irrelevant
- We can then assume the following as η_2, the linear predictor for the "normal" participants:

$$\begin{aligned} u_{part[n],1} + w_{item[n],1} + &\textsc{order}_n \times (\beta_{\textsc{order}} + u_{part[n],2} + w_{item[n],2}) + \\ &\textsc{case}_n \times (\beta_{\textsc{case}} + u_{part[n],3}) + \\ &\textsc{case}_n \times \textsc{order}_n \times (\beta_{\textsc{case:order}} + u_{part[n],4}) + \\ &\textsc{accept}_n \times \beta_{\textsc{accept}} + \textsc{reject}_n \times \beta_{\textsc{reject}} \end{aligned} \tag{4}$$

- Priors: $\beta_{\textsc{accept}} \sim \mathcal{N}(3, 1.5)$, $\beta_{\textsc{reject}} \sim \mathcal{N}(-3, 1.5)$
- If test items are really ungrammatical for these participants, they should be as bad as unacceptable fillers, so $\beta_{\textsc{reject}}$ will be relevant for them. η_1 could then be:

$$u_{part[n],1} + w_{item[n],1} + \beta_{\textsc{reject}} + \textsc{accept}_n \times (\beta_{\textsc{accept}} - \beta_{\textsc{reject}}) \tag{5}$$

- For acceptable fillers this effect has to be subtracted again.

I will call this model M_{diff}.

2.7 Results $\mathbf{M_{diff}}$

M_{diff} was fit with PyMC (Martin & Kumar & Lao 2021) (4 chains, 30 000 draws per chain, using default Metropolis-within-Gibbs sampler for b, the default No-U-Turn sampler for the other parameters)

- Estimates for the thresholds are a bit smaller than in $\mathrm{M}_{\mathrm{standard}}$ overall: $\hat{\tau}_1 = -1.82\ [-2.33, -1.32]$, $\hat{\tau}_2 = 0.32\ [-0.16, 0.8]$, $\hat{\tau}_3 = 1.06\ [0.58, 1.56]$, $\hat{\tau}_4 = 2.95\ [2.43, 3.47]$
- $\hat{\beta}_{\text{ORDER}} = 0.8\ [0.58, 1.04]$, $\mathrm{BF}_{10} = 60.77$
- $\hat{\beta}_{\text{CASE}} = -0.06\ [-0.39, 0.26]$, $\mathrm{BF}_{10} = 0.04$
- $\hat{\beta}_{\text{CASE:ORDER}} = 0.26\ [0.1, 0.43]$, $\mathrm{BF}_{10} = 1.77$
- $\hat{\beta}_{\text{REJECT}} = -1.9\ [-2.46, -1.35]$, $\mathrm{BF}_{10} = 82.54$, $\hat{\beta}_{\text{ACCEPT}} = 4.02\ [3.3, 4.76]$, $\mathrm{BF}_{10} = 79.55$
- $\hat{\pi} = 0.2\ [0.06, 0.36]$ (the posterior has a bit positive skew, but not much) ⇒ estimate is that 20 % of the participants will belong to the group rejecting all test items.
- $\hat{\sigma}_{u_1} = 0.9\ [0.65, 1.15]$. Difference may not look that big at first glance but with $\sigma_{u_1} = 0.9$ 80 % of the participants have individual intercepts between −1.15 and 1.15, while with $\sigma_{u_1} = 1.5$ 80 % of the participants have individual intercepts between −1.92 and 1.92. The latter would mean that individual differences within the bulk of participants effectively span the whole scale.

2.8 Learnings Binding

- For some (ca. 20 %) of the participants, the factors tested are irrelevant as they reject all test items
- Even if this is accounted for, the test items in the OS condition are not rated as fully natural
- But they are considerably more natural than the test items in the SO condition and unacceptable fillers, and we can be very sure about that.
- The fact that the test items are not fully natural even when the behaviour of the participants who reject everything is accounted for needs to be explained.
- Arguably account of acceptability of N P REFL in diverse settings needed, but this goes beyond this talk.

3 Crossover Violations

The analysis of the data by (Howitt & Scontras & Polinsky 2025) here strongly builds on a re-analysis of that data by Tibor Kiss.

3.1 Crossover Violations English Relative Clauses

- Longstanding conflicting judgments about the existence of weak crossover (WCO) effects in English restricted relative clauses in examples like (8a) (Howitt & Scontras & Polinsky 2025).

(8) (adapted from Howitt & Scontras & Polinsky 2025: 5)
  a. The plane$_1$ which$_1$ its$_1$ pilot t_1 flew around the country avoided turbulence.
  b. The plane$_1$ which$_1$ the pilot flew t_1 around the country avoided turbulence.

- Howitt & Scontras & Polinsky (2025) conducted a slider-scale acceptability judgment study on the phenomenon:

  - 24 test items (2 of which excluded), 6 fillers only for attention checks
  - 3 factors: WCO (yes or no), ANIMACY (of relative clause head), DETERMINER (R-expression or quantificational), but ANIMACY and DETERMINER turned out not to have an effect
  - data from 78 participants considered in analysis
- They checked for differences between participants (not their main interest) by looking for bimodality in a histogram. In their view there seems to be none (but hard to tell in my opinion).

3.2 Crossover Violations: Lectal Variants

- Some reason to assume that there could be different groups of participants: WCO relevant for some but not others.
- No (non-attention-check) filler data that could be used.
- I ran linear models as well as (for this kind of data arguably more appropriate) Zero-one-inflated Beta (ZOIB, see i.a. Ospina & Ferrari 2010; Liu & Eugenio 2018) models on the data with latent discrete variables governing the choice of linear predictor as before.
- Linear predictors for the linear model (ZOIB model analogous, same formula for mean, φ, α, γ):

$$\eta_1 = \alpha + u_{part[n],1} + w_{item[n],1} + \textsc{wco}_n \times (\beta_{\textsc{wco}} + u_{part[n],2} + w_{item[n],2}) \tag{6}$$

$$\eta_2 = \alpha + u_{part[n],1} + w_{item[n],1} \tag{7}$$

- $\pi \sim Beta\,(1,1)$
- $b \sim Bernoulli\,(\pi)$
- $\eta_n = \begin{cases} \eta_{1,n} & \text{if } b_{part[n]} = 1 \\ \eta_{2,n} & \text{if } b_{part[n]} = 0 \end{cases}$

Results:

- Parameters for population-level effects and participants' group-level effect standard deviations in the models incorporating participant differences do not differ substantially from the corresponding standard models, though they are a bit more pronounced for the ZOIB model
- $\hat{\pi} = 0.71$ [0.51, 0.92] (linear model), $\hat{\pi} = 0.67$ [0.51, 0.84] (ZOIB), indicating that there is no WCO effect in this setting for ca. ⅓ of the participants

4 Conclusion

- I discussed a way to handle assumed differences between participants in statistical models via different linear predictors and latent discrete variables governing group membership.
- Such models would probably not be the first or only ones one uses, but arguably they can help us getting a better understanding of the data and identifying linguistically relevant patterns:
  - In the binding study, they showed that the test items in the arguably grammatical condition are not fully acceptable even when it is taken into account that some participants reject them for presumably independent reasons.
  - For WCO in English relative clauses the models based on data by Howitt & Scontras & Polinsky (2025) suggest that a significant share of speakers does not get the effect, which is in line with the patterns in the theoretical literature and may itself need an explanation.

References

• Bürkner, Paul-Christian. 2017. Brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software* 80(1). *https://doi.org/10.18637/jss.v080.i01*. • Howitt, Katherine & Scontras, Gregory & Polinsky, Maria. 2025. English restrictive relative clauses are subject to crossover violations. *Linguistic Inquiry*. *https://doi.org/10.1162/LING.a.541*. • Jeffreys, Harold. 1939. *Theory of probability*. Oxford: Clarendon Press. • Liu, Fang & Eugenio, Evercita C. 2018. A review and comparison of Bayesian and likelihood-based inferences in beta regression and zero-or-one-inflated beta regression. *Statistical Methods in Medical Research* 27(4). 1024–1044. *https://doi.org/10.1177/0962280216650699*. • Martin, Osvaldo A. & Kumar, Ravin & Lao, Junpeng. 2021. *Bayesian Modeling and Computation in Python*. Boca Raton: CRC Press. *https://doi.org/10.1201/9781003019169*. • Masloch, Simon & Poppek, Johanna M. & Kiss, Tibor. 2025. On the (im-)possibility of reflexive binding into the subject of German experiencer-object verbs. In Bîlbîie, Gabriela & Schaden, Gerhard (eds.), *Empirical issues in syntax and semantics. Selected papers from CSSP 2023*, 197–222. Berlin: Language Science Press. *https://doi.org/10.5281/zenodo.15450442*. • Ospina, Raydonal & Ferrari, Silvia L. P. 2010. Inflated beta distributions. *Statistical Papers* 51(1). 111–126. *https://doi.org/10.1007/s00362-008-0125-4*. • R Core Team. 2023. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. *https://www.R-project.org/*. • Stan Development Team. 2024. *Stan User's Guide*. Version 2.35. *https://mc-stan.org/docs/2_35/stan-users-guide/*.