Writing does not impact the evolutionary dynamics of syntax

Carlo Y. Meloni, Chundra A. Cathcart, Jessica K. Ivani, Taras Zakharko, Guanghao You & Balthasar Bickel

The Institute for the Interdisciplinary Study of Language Evolution

Writing Effects

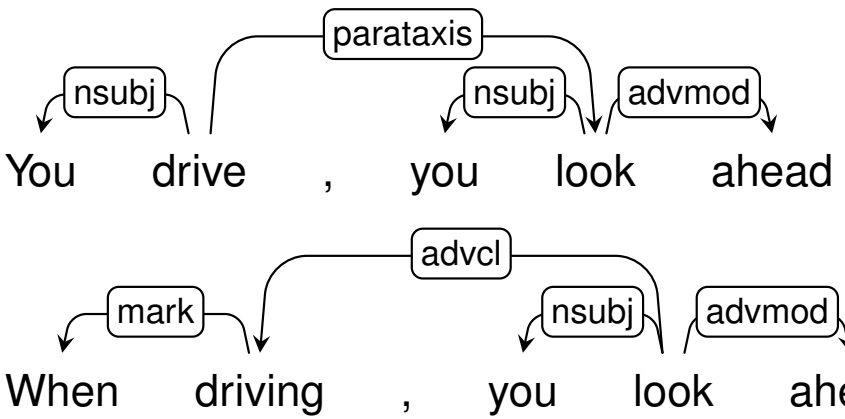
- Writing acquisition significantly impacts language processing (Dehaene et al. 2010; Cilibrasi, Adani, and Tsimpli 2019).
- However, its influence on the evolution of syntax remains unclear.
- Common hypothesis: writing fosters greater hierarchy and “syntactic complexity” (Delbrück 1900; Small 1924; Dąbrowska 2015).

Potential Effects on Language

- Writing may impact language on two levels:
  - **Language use**: Variation in syntactic usage across different communication mediums (synchronic).
  - **Language grammar**: Changes in the grammatical structure of a language (diachronic).

Assessing Language Use

– Universal Dependencies (UD) data used to quantify “syntactic complexity”:



– Metrics: total number of clauses per sentence, maximum clause depth (“maximum clausal path”).

Data

Catena Len.	Genre	Language	...
0	spoken	Abaza	...
2	spoken	Abaza	...
⋮	⋮	⋮	...
1	fiction	Czech	...
⋮	⋮	⋮	...
3	wiki	Chinese	...
⋮	⋮	⋮	...

Writing does not impact the evolutionary dynamics of syntax

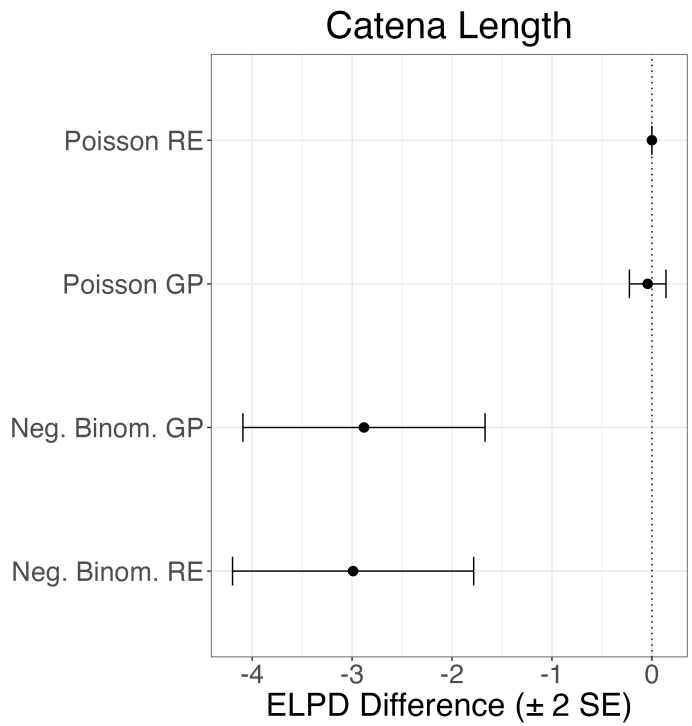
- We contrasted three genres: Spoken, Fiction and Wikipedia.
  - Spoken: 31,277 sentences.
  - Fiction: 35, 103 sentences.
  - Wiki: 33,454 sentences.

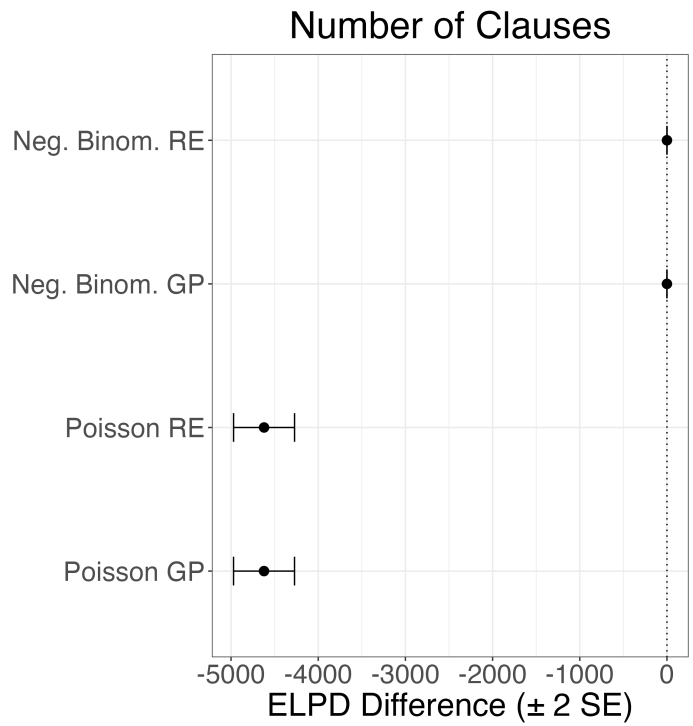
Modeling Approach

```
catena_len ~ 1 + genre + (1|treebank) + (1|area) +
             (1|gr(phylo, cov = phylo))
```

- We controlled for phylogenetic, areal and dataset-specific effects ("treebank").
- Employed regression models:
  - Poisson vs. Negative Binomial (accounting for over-dispersion).
  - Controlled for areal effect via Gaussian Process (GP) and Random Effects (RE) over micro-areas (Bickel et al. 2023).

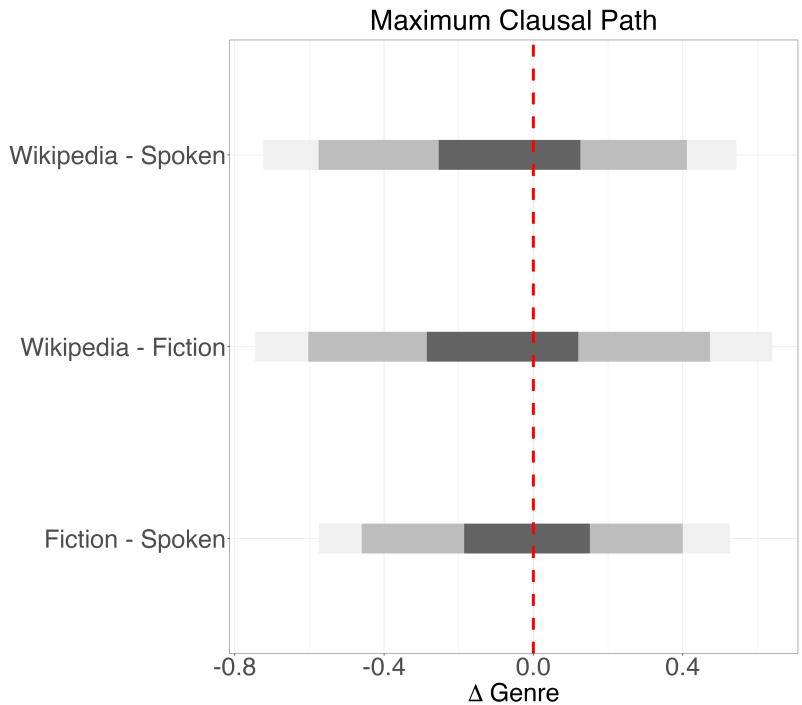
Model Comparison

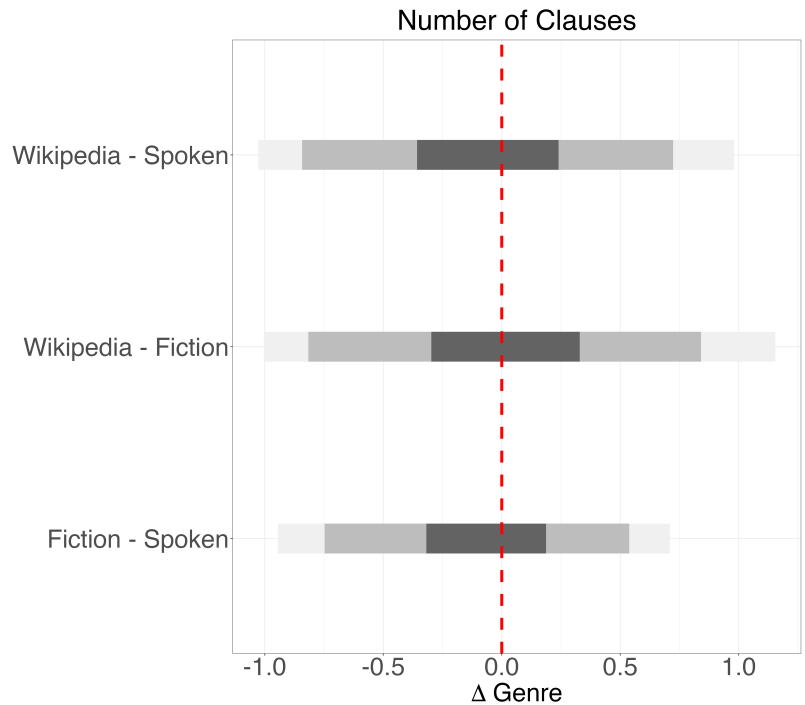




Results

- Marginal effects show no significant effect of genre on either measure of “syntactic complexity”.
- No detectable phylogenetic or areal influences.





Assessing Language grammar

– Examined 763 clause-combining constructions across 59 languages.

Data

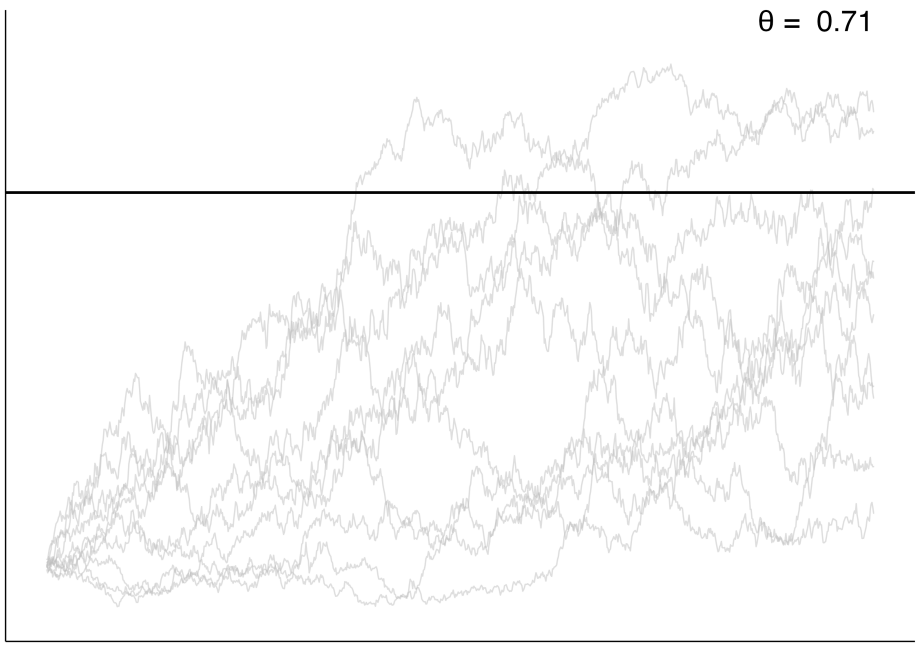
– Structured according to van Gijn, Galucio, and Nogueira (2015).

Construction	Asymmetry	Language	Value
C_{A1}	SubjAgrLim	Albanian	0
C_{F1}	SubjAgrLim	French	1
⋮	⋮	⋮	⋮
C_{T2}	ObjAgrLim	Tocharian	0
⋮	⋮	⋮	⋮
C_{G3}	OrderLim	Gothic	1
⋮	⋮	⋮	⋮

– Asymmetry values used to calculate the probability of a binomial process ($\mathrm{B}(n_{success}, p_{asymmetry})$).

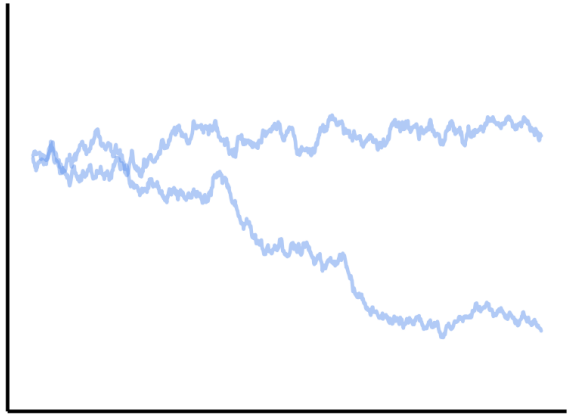
Phylogenetic Modeling

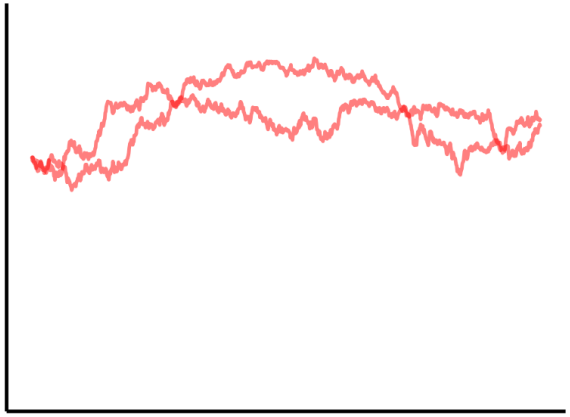
$$\underline{\text{SDE}}: \underbrace{d\overbrace{X_t}^{\text{trait value}}}_{} = \underbrace{\alpha}_{\text{strength}}(\overbrace{\theta}^{\text{optimum}} - X_t)dt + \underbrace{\sigma}_{\text{rate}} dW_t$$



Writing does not impact the evolutionary dynamics of syntax

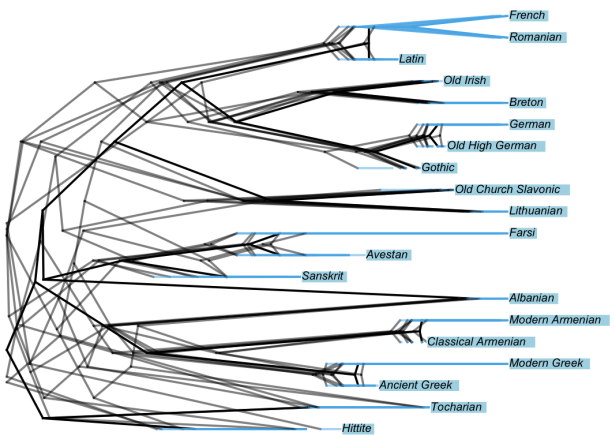
– Each regime has different θ, α and σ parameters.

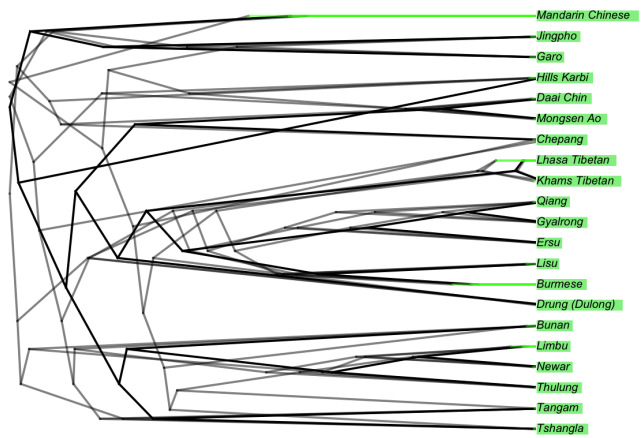


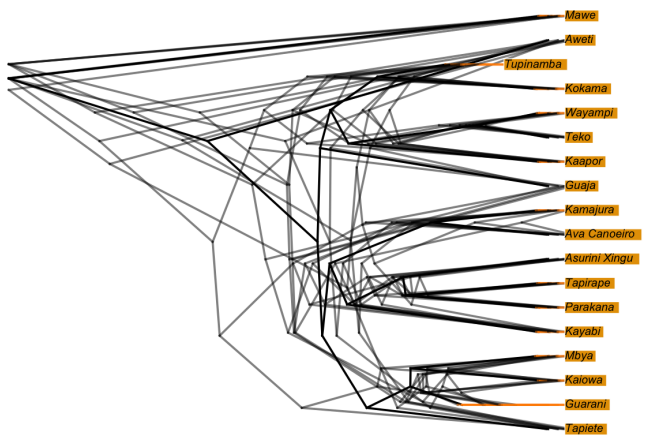


Writing does not impact the evolutionary dynamics of syntax

– We compared two regimes, “writing” (colored branches) vs. “non-writing” (black branches).

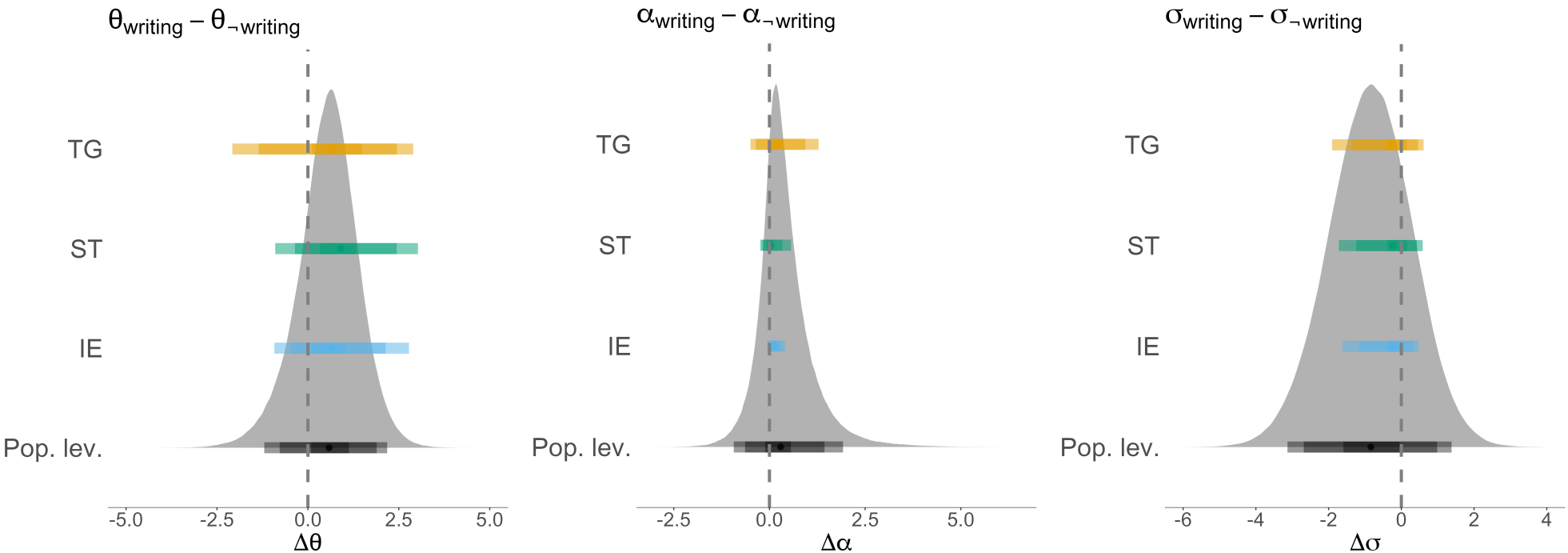






Writing does not impact the evolutionary dynamics of syntax

– No significant differences in parameter values (θ, α, σ) between writing and non-writing regimes.



Conclusions

- No evidence supporting the impact of writing on hierarchy degree, both in terms of language use and grammar.
  - No differences found across genres.
  - No differences in the parameter values across the two regimes.
- Decreased σ and increased α in IE (and to a lesser extent, ST and TG) may indicate a trend toward normativization in the writing regime.
  - Differences among families may reflect distinct cultural traditions of writing.

Thanks for your attention!

Questions?

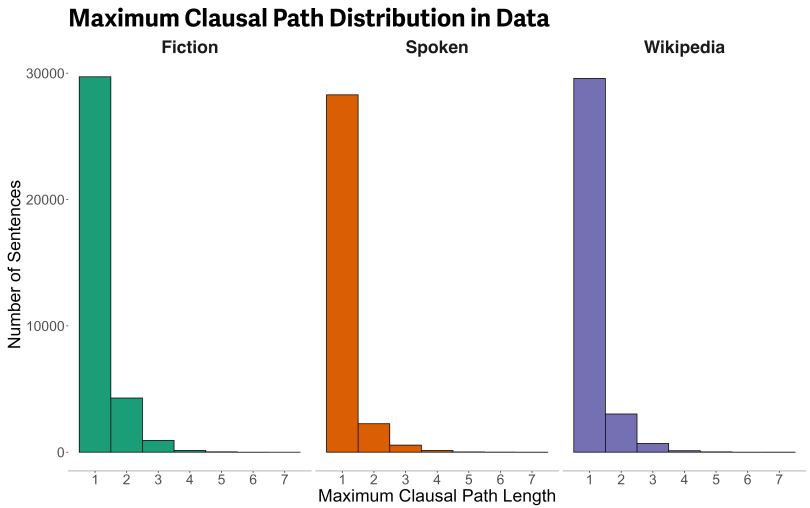
References I

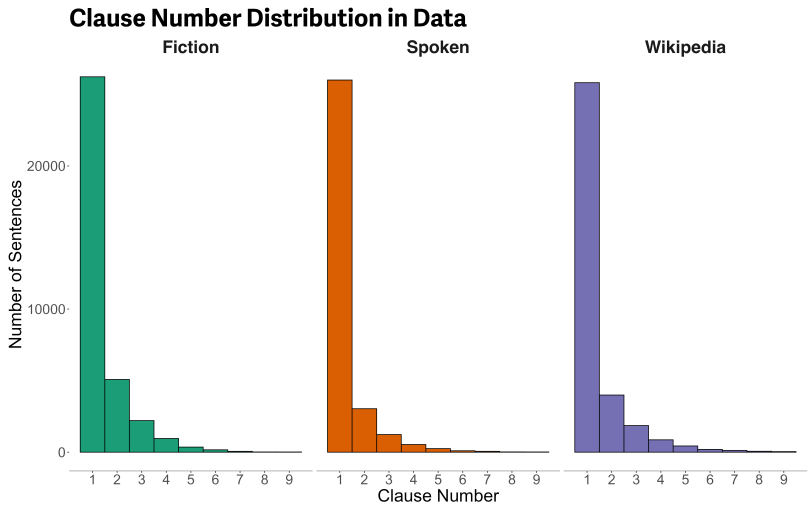
- Bickel, Balthasar et al. (2023). *The AUTOTYP database (v1.1.1)*. DOI: 10.5281/zenodo.7976754. URL: https://doi.org/10.5281/zenodo.7976754.
- Cilibrasi, Luca, Flavia Adani, and Ianthi Tsimpli (2019). "Reading as a Predictor of Complex Syntax. The Case of Relative Clauses". In: *Frontiers in Psychology* 10. ISSN: 1664-1078. DOI: 10.3389/fpsyg.2019.01450. URL: https://www.frontiersin.org/journals/psychology/articles/10.3389/fpsyg.2019.01450.
- Dąbrowska, Ewa (2015). "Language in the mind and in the community". In: *Change of paradigms–New paradoxes*, pp. 221–235.
- Dehaene, Stanislas et al. (2010). "How Learning to Read Changes the Cortical Networks for Vision and Language". In: *Science* 330.6009, pp. 1359–1364. DOI: 10.1126/science.1194140. eprint: https://www.science.org/doi/pdf/10.1126/science.1194140. URL: https://www.science.org/doi/abs/10.1126/science.1194140.

References II

- Delbrück, Berthold (1900). *Vergleichende syntax der indogermanischen sprachen*. Walter de Gruyter GmbH & Co KG.
- Small, George William (1924). *The Comparison of Inequality: The Semantics and Syntax of the Comparative Particle in English*. Greifswald, p. 173.
- van Gijn, Rik, Ana Vilacy Galucio, and Antonia Fernanda Nogueira (2015). "Subordination strategies in Tupian languages". In: *Boletim do Museu Paraense Emilio Goeldi. Ciencias Humanas* 10.2, pp. 297–324. ISSN: 2178-2547. DOI: 10.1590/1981-81222015000200006. (Visited on 01/22/2024).

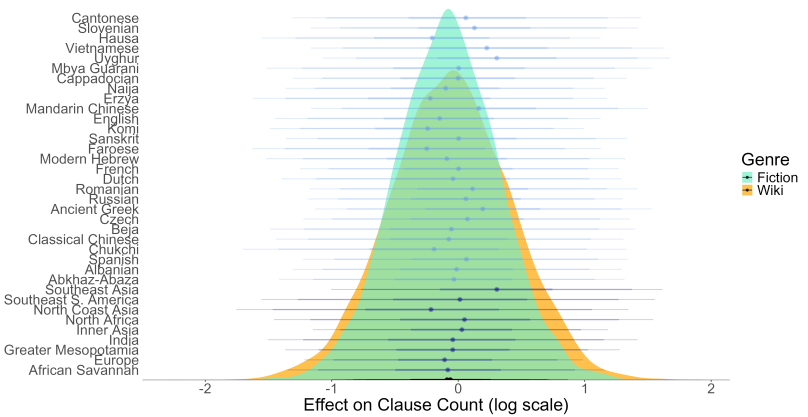
Writing does not impact the evolutionary dynamics of syntax



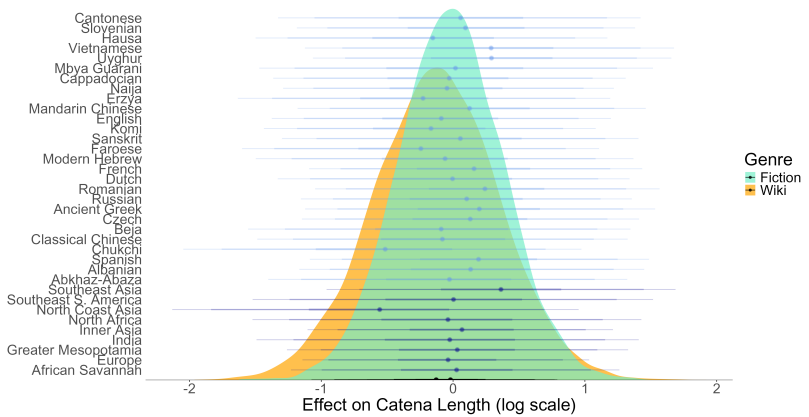


Writing does not impact the evolutionary dynamics of syntax

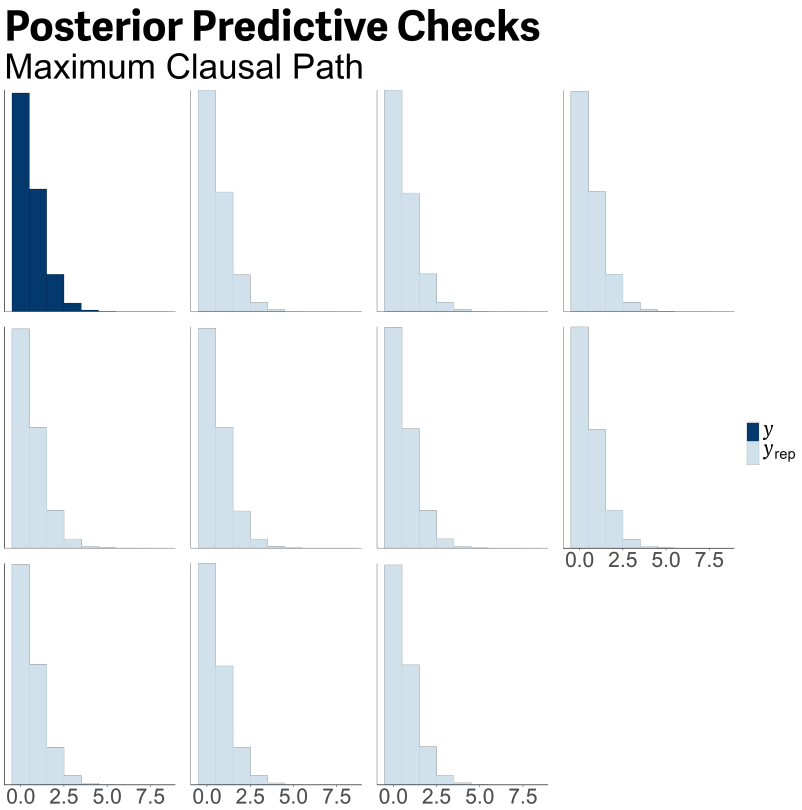
Clause Number
Predicted Differences from Spoken Genre

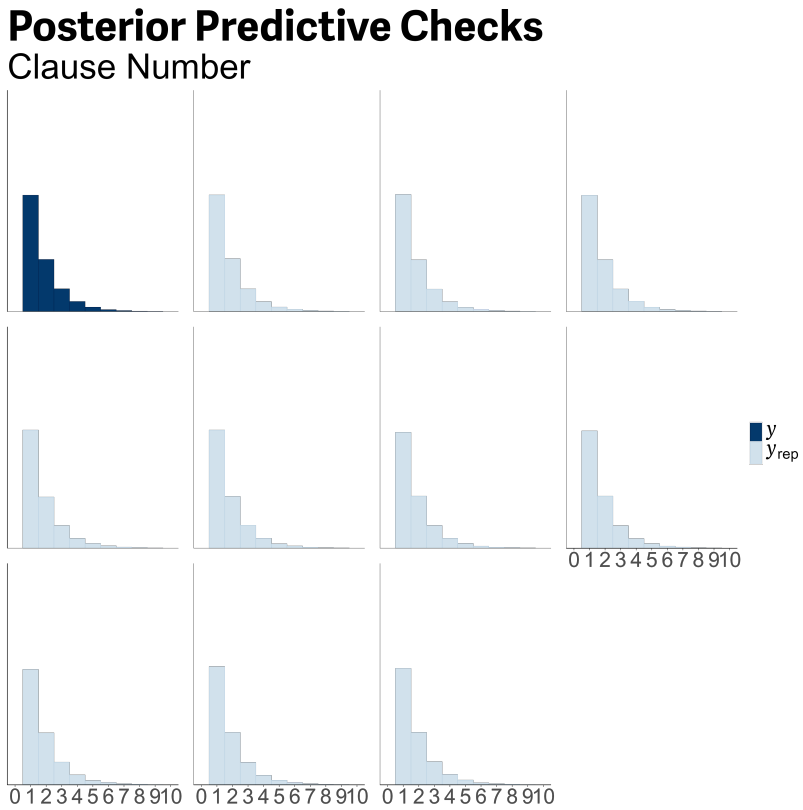


Maximum Clausal Path
Predicted Differences from Spoken Genre

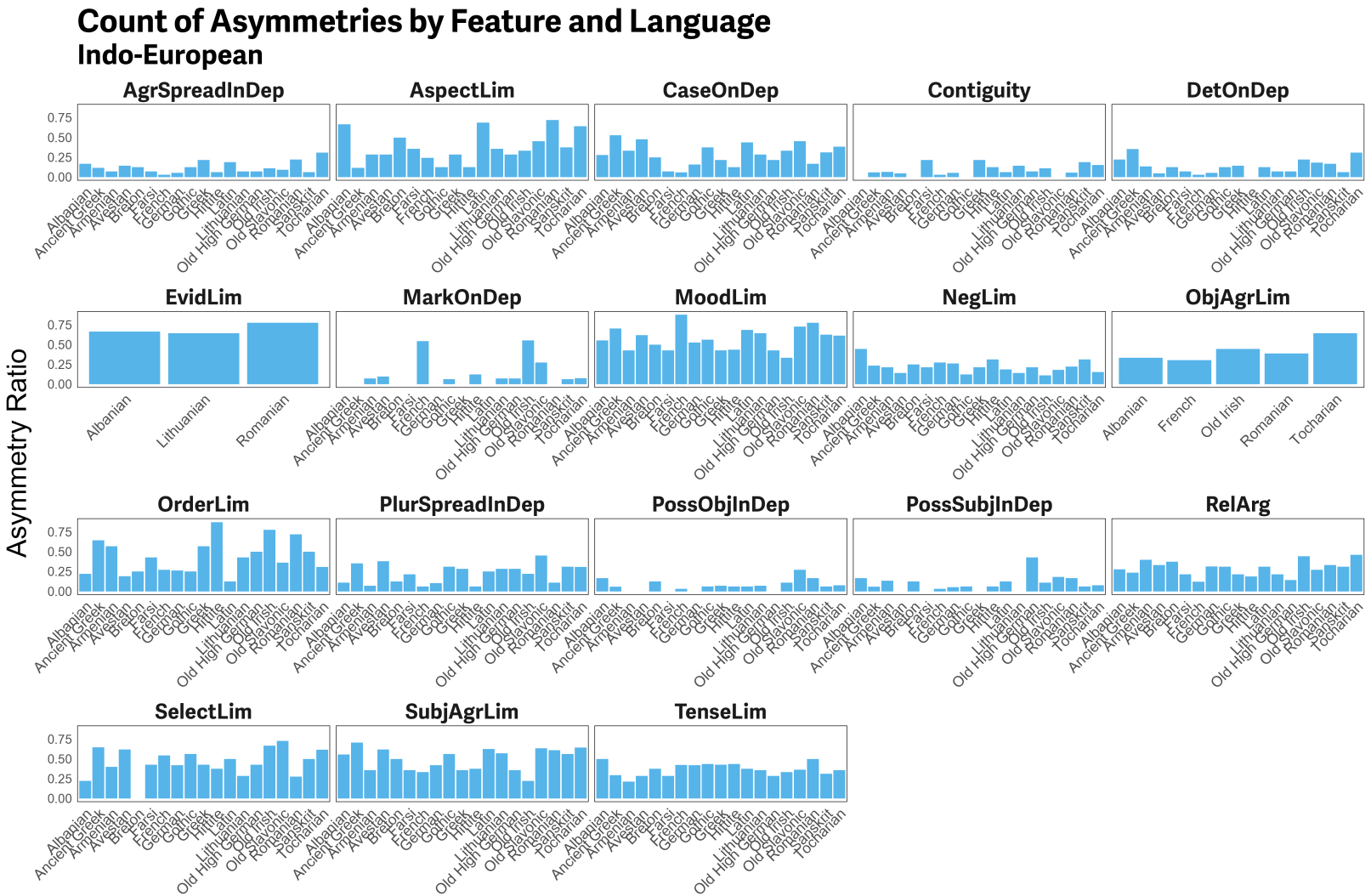


Writing does not impact the evolutionary dynamics of syntax

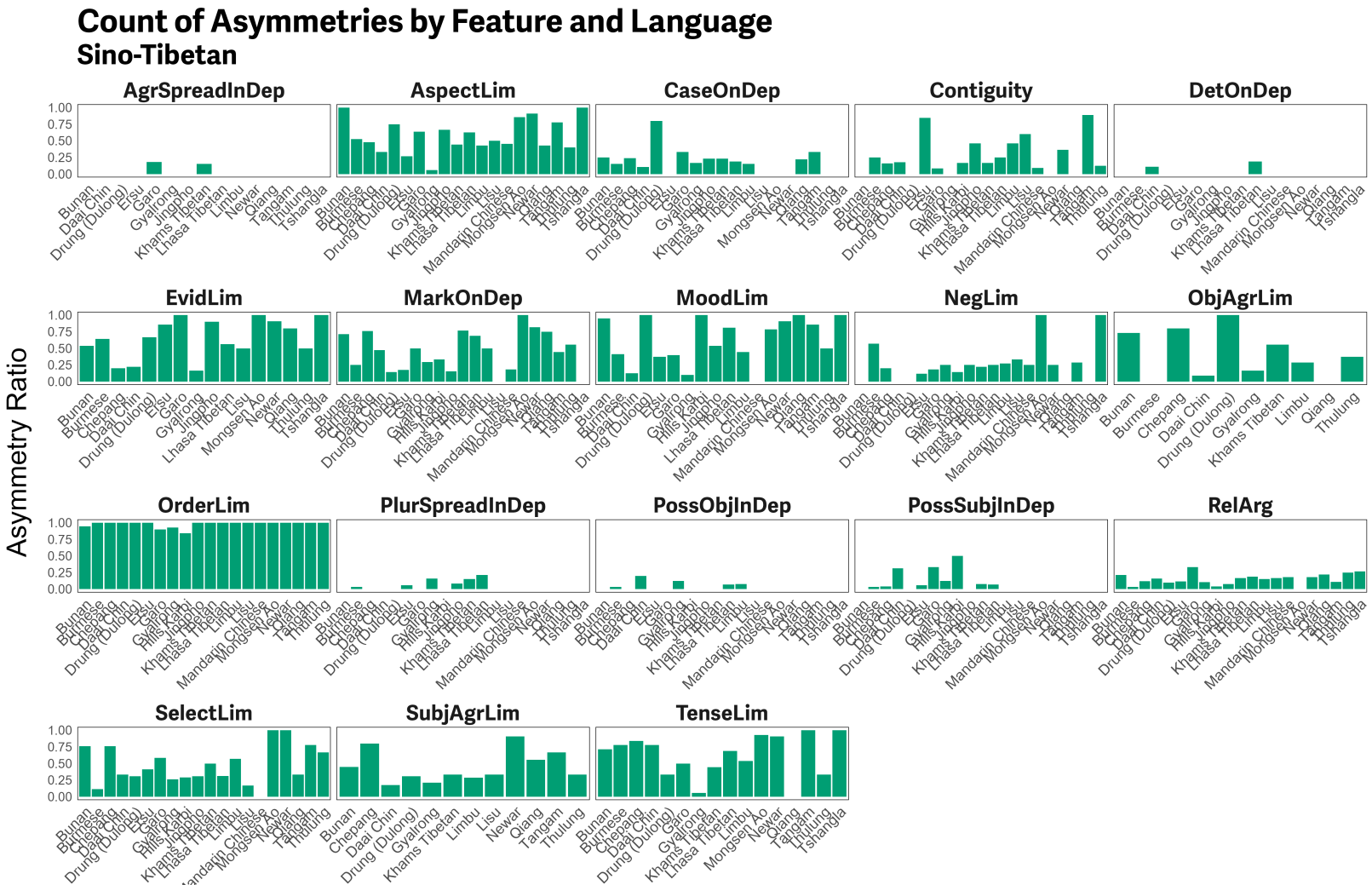




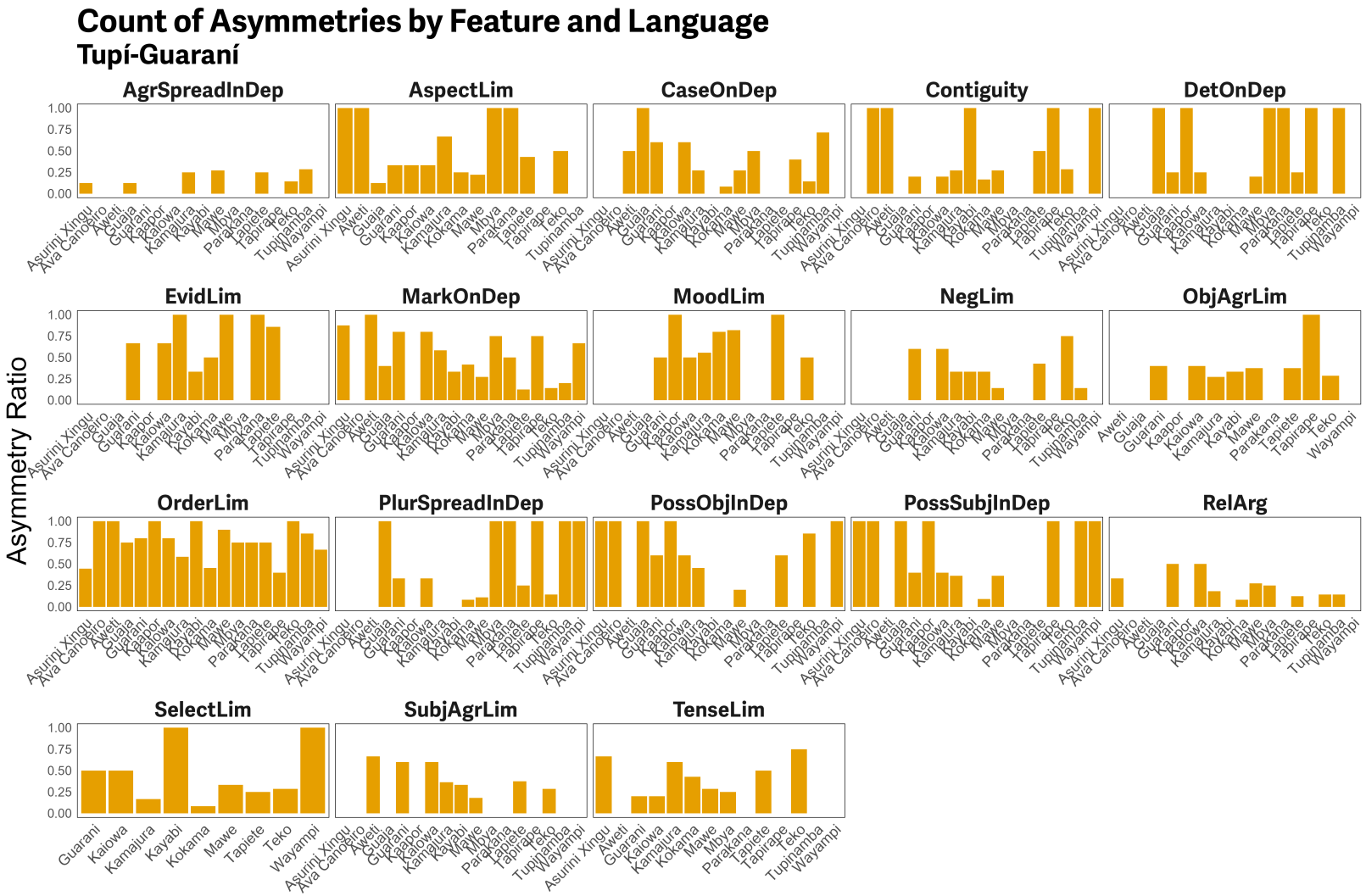
Writing does not impact the evolutionary dynamics of syntax



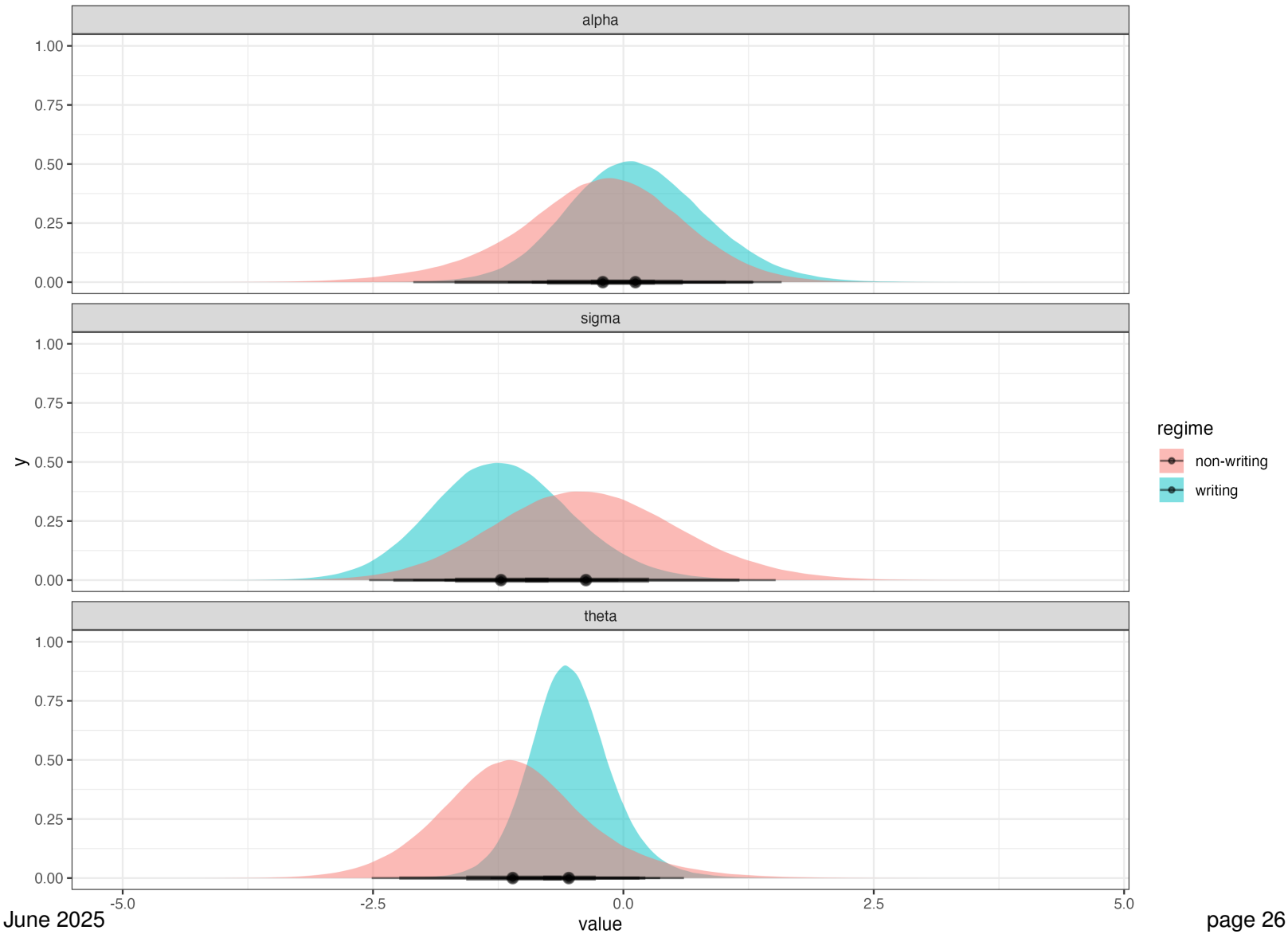
Writing does not impact the evolutionary dynamics of syntax



Writing does not impact the evolutionary dynamics of syntax



Writing does not impact the evolutionary dynamics of syntax



Writing does not impact the evolutionary dynamics of syntax

Parameter	Median	HPDI (95%)	$\Pr(\Delta\beta) > 0$
θ_μ	0.58	$[-1.19, 2.18]$	76.62%
θ_{IE}	0.66	$[-0.95, 2.78]$	84.46%
θ_{ST}	0.90	$[-0.90, 3.08]$	89.45%
θ_{TG}	0.69	$[-2.13, 2.91]$	73.08%
α_μ	0.30	$[-0.93, 1.92]$	74.44%
α_{IE}	0.14	$[-0.06, 0.41]$	95.42%
α_{ST}	0.04	$[-0.24, 0.56]$	63.12%
α_{TG}	0.21	$[-0.50, 1.28]$	77.34%
σ_μ	-0.84	$[-3.14, 1.37]$	23.02%
σ_{IE}	-0.25	$[-1.62, 0.47]$	20.29%
σ_{ST}	-0.22	$[-1.72, 0.58]$	26.75%
σ_{TG}	-0.26	$[-1.90, 0.62]$	25.47%

Differences in Ornstein-Uhlenbeck process parameter values between writing and non-writing regimes. All values in the table represent the difference between the writing and non-writing regimes, calculated as $\beta_{\text{writing}} - \beta_{\neg\text{writing}}$.