

Predicting auxiliary selection in Italo-Romance dialects: A spatial modelling approach

Pavel Štichauer¹ & Fabio Ripamonti²

¹Charles University, Prague ²University of South Bohemia in České Budějovice
pavel.stichauer@ff.cuni.cz

1 Background

- This poster investigates **auxiliary selection in Italo-Romance dialects**, focusing on **mixed paradigms**, i.e. systems that alternate between *have* (H) and *be* (E) within a single perfective paradigm (cf., among others, Bentley & Eythórsson, 2001; D'Alessandro & Roberts, 2010; Ledgeway, 2019; Loporcaro, 2007, 2014).
- These systems disrupt the expected uniformity of auxiliary distribution within and across TAM subparadigms, yielding varied splits whose distribution calls for systematic quantitative analysis.

2 Research aim

- The study provides the **first probabilistic account** of auxiliary selection in Italo-Romance mixed paradigms, integrating **linguistic** and **spatial** predictors within a unified modelling framework.
- It asks which factors most robustly condition auxiliary choice, and to what extent spatial structure improves the modelling of dialectal variation.

3 Data

- The analysis is based on the **MIXPAR** database (Štichauer & Ripamonti, 2025), which currently contains more than **3,400 annotated forms** from **192 dialect locations**.
- Each record is annotated for **Cell** (1SG–3PL), **TAM**, **Verb Class**, **Verb Construction**, **AUX**, and geographic metadata including **Region**, **Place**, and **GPS coordinates**.
- This structure supports systematic comparison across paradigms, verb classes, TAM values, and dialect areas.

4 Descriptive patterns

- Descriptive statistics show that **Cell**, i.e. the concrete person-number combination, is the strongest factor in auxiliary choice.
- Third-person forms, especially **3PL**, strongly favour *have*, whereas first- and second-person forms more often select *be*.
- The most frequent non-uniform pattern is **EEHEEH**, which recurs especially in central and southern Italy.
- Local Moran's *I* confirms that EEHEEH is **spatially clustered**, rather than randomly distributed.

5 Modelling approach

- Auxiliary choice was modelled as a **binary outcome** (*be* vs. *have*) using **Bayesian mixed-effects models** in *brms* (Bürkner, 2018).
- The models included **Cell**, **TAM**, and **Verb Class** as fixed effects, and **Verb Construction** as a random effect.
- A baseline model was compared with spatially informed models incorporating a **Gaussian process smoother** over geographic coordinates.
- To address **spatial confounding**, we compared **RSR** (Hodges & Reich, 2010), **DSR** (Wiecha et al., 2025), and **Spatial+** (Dupont et al., 2022).
- Spatial+** provided the best fit and captured fine-grained **local microvariation**.

6 Main results

- Among the fixed effects, **Cell** emerges as the dominant predictor: third-person forms, especially **3PL** ($\beta \approx 2.45$) and **3SG** ($\beta \approx 1.73$), are robustly associated with *have* (H).

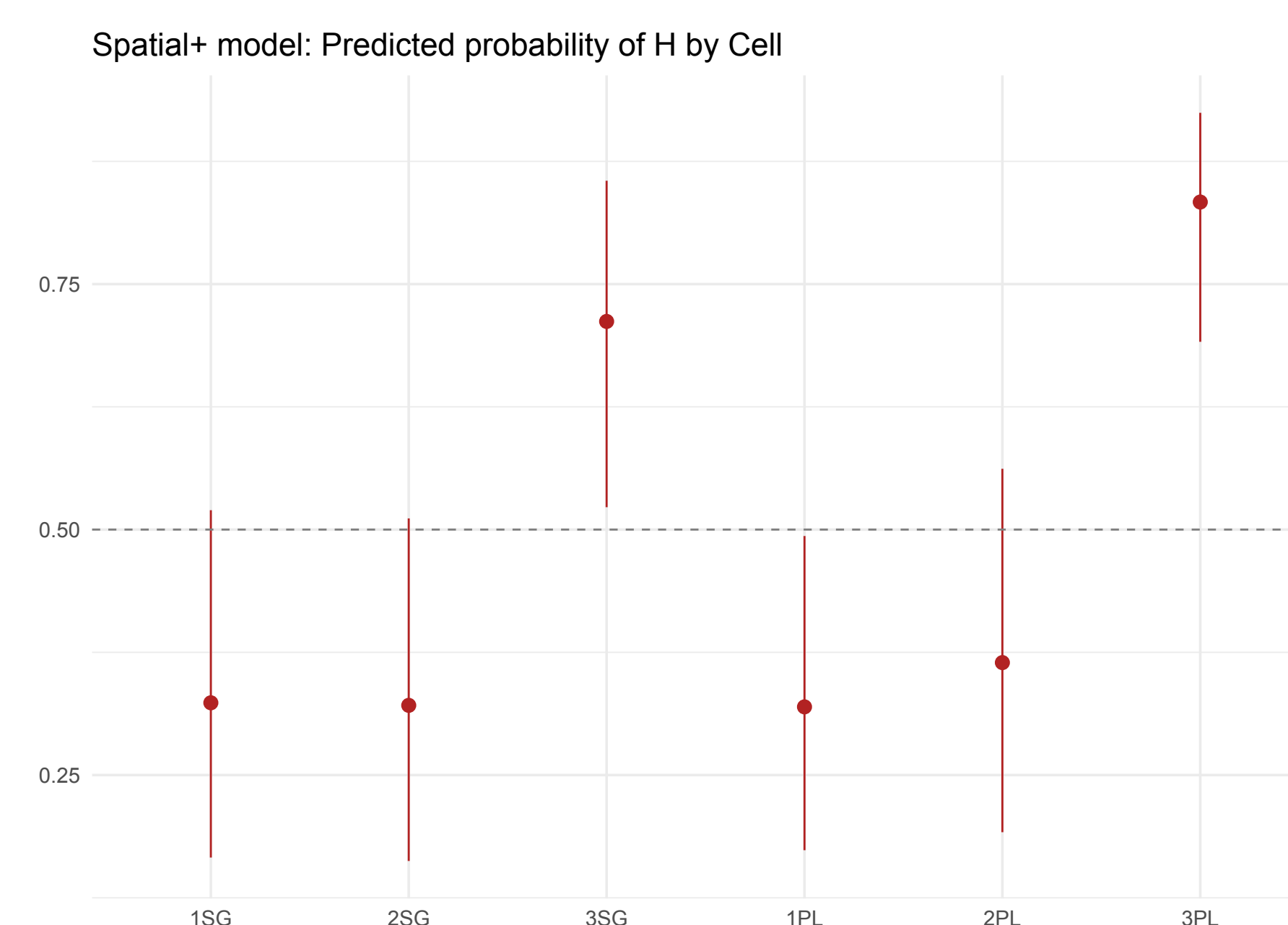


Figure 1: Conditional effects of Cell on the probability of selecting *have* (H) in the Spatial+ model.

- TAM** also contributes, with the **pluperfect** and **pluperfect subjunctive** showing a stronger tendency towards H.
- By contrast, **Verb Class** shows weaker and largely overlapping effects, suggesting that argument-structure factors remain relevant but are less robust than person-based asymmetries.
- Random effects for individual verbs are comparatively small, indicating that **lexical idiosyncrasy** plays a more limited role than grammatical and areal conditioning.

7 Conclusions

- Person-based asymmetry** is the most stable conditioning factor of auxiliary alternation in Italo-Romance mixed paradigms.
- The mixed pattern **EEHEEH** emerges as the quantitatively dominant and most robustly supported non-uniform configuration.
- Future work must explain why this split is favoured – possibly in connection with a participant/non-participant asymmetry – while other logically possible patterns remain rare.

References

- Bentley, D., & Eythórsson, T. (2001). Alternation according to person in italo-romance. In L. J. Brinton (Ed.), *Historical linguistics 1999. selected papers from the 14th international conference on historical linguistics, vancouver, 9-13 august 1999* (pp. 63–74). Benjamins.
- Bürkner, P.-C. (2018). Advanced Bayesian multilevel modelling with the R package brms. *The R Journal*, 10(1), 395–411.
- D'Alessandro, R., & Roberts, I. (2010). Past participle agreement: Split auxiliary selection and the null-subject parameter. *Natural Language and Linguistic Theory*, 28(1), 41–72.
- Dupont, E., Wood, S. N., & Augustin, N. H. (2022). Spatial+: A novel approach to spatial confounding. *Biometrics*, 78(4), 1279–1290.
- Hodges, J. S., & Reich, B. J. (2010). Adding spatially correlated errors can mess up the fixed effect you love. *The American Statistician*, 64(4), 325–334.
- Ledgeway, A. (2019). Parameters in the development of romance perfective auxiliary selection. In M. Robbeets & W. Bisang (Eds.), *Historical linguistics 2015* (pp. 344–384). John Benjamins Publishing Company.
- Loporcaro, M. (2007). On triple auxiliation in Romance. *Linguistics*, 45(1), 173–222.
- Loporcaro, M. (2014). Perfective auxiliation in italo-romance. the complementarity of historical and modern cross-dialectal evidence. In P. Benincà, A. Ledgeway, & N. Vincent (Eds.), *Diachrony and dialects. grammatical change in the dialects of italy* (pp. 48–70). Oxford University Press.
- Štichauer, P., & Ripamonti, F. (2025). Mixpar database v1.0. <http://hdl.handle.net/11234/1-5982>
- Wiecha, N., Hoppin, J. A., & Reich, B. J. (2025). Two-stage estimators for spatial confounding with point-referenced data. *Biometrics*, 81(3), 1–10.

Acknowledgements

The support of the project **The MIXPAR project: A quantitative approach to mixed perfective auxiliation patterns in Italo-Romance**, funded by the **Czech Science Foundation** (grant no. 24-10205S), is gratefully acknowledged. This poster is based on a paper forthcoming in the special issue *Perspectives on Romance inflection*, guest-edited by Borja Herce and Chiara Zanini, in *Word Structure*. We also thank both guest editors for their valuable feedback.