

Quantitative detection of cognacy in the predictive structure of inflection classes: Romance verbal conjugations against the broader typological variation

Borja Herce & Balthasar Bickel
University of Zurich & ISLE

Quantitative approaches to inflectional systems have become widespread, particularly in relation to the Paradigm Cell Filling Problem (Ackerman et al. 2009), which captures how, even in morphologically complex languages, speakers can produce any form in a lexeme's paradigm on the basis of an incomplete input. Information Theory (with its core notion of entropy) has provided the theoretical background for a lot of empirical research in this domain in recent years (e.g. Milin et al. 2009, Ackerman & Malouf 2013, Stump & Finkel 2013, Sims & Parker 2016, Cotterell et al. 2019). Stump and Finkel (2013), for example, introduced a series of metrics that capture different aspects of an inflectional class system's complexity (i.e. of how easy/difficult it is to infer some forms on the basis of others). These included static and dynamic principal parts, different kinds of predictability and predictiveness, etc.

So far these metrics have chiefly been used to assess synchronic states. Here we explore their potential for capturing patterns in language change and phylogenetic relatedness. Specifically, we probe different aspects of an inflectional system for their stability within one language family, Romance, and for the degree to which they distinguish this family from unrelated and less closely related languages. Drawing on Beniamine et al. (2020), and other sources (Bonami et al. 2014, Calderone et al. 2019, Barbu 2009, Kirov et al. 2018, Perea & Ueda 2010), we compiled a database with 100 cognate verbs in 7 Romance languages. On the basis of this, we calculated Stump & Finkel's (2013) predictive complexity measures (and some additional ones), and evaluated the variability

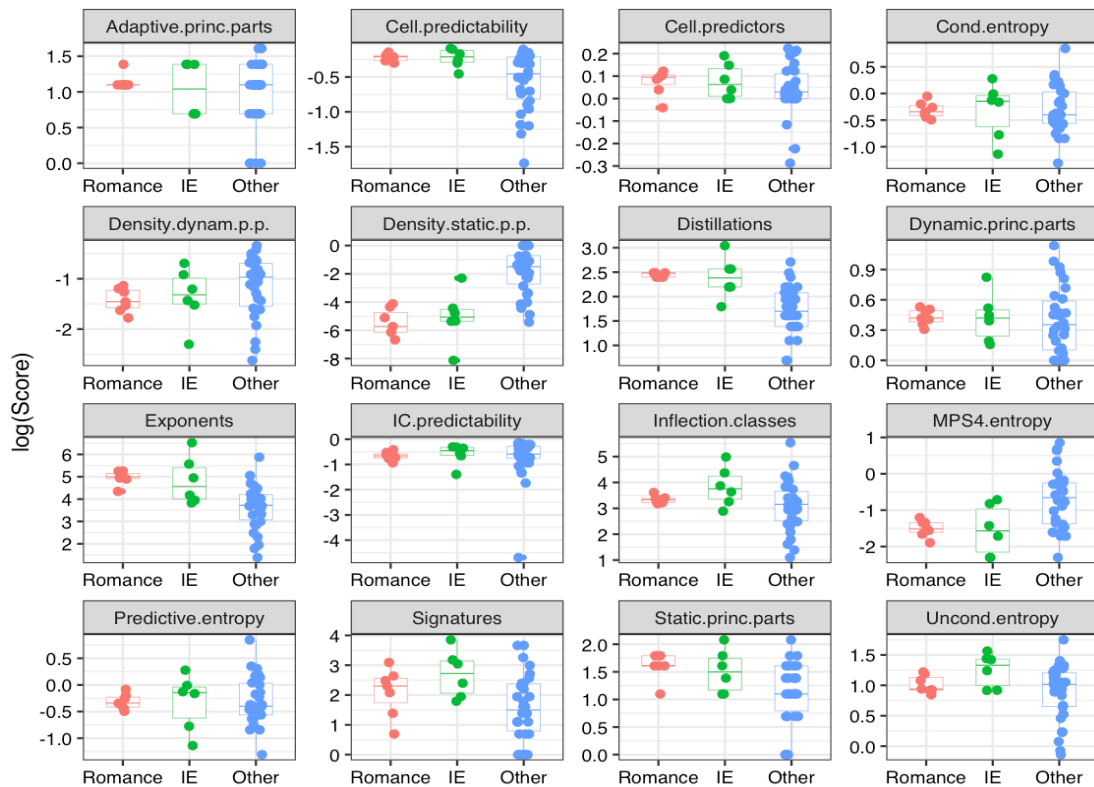


Figure 1: IC traits in Romance, other Indo-European, and unrelated languages

References

- Ackerman, Farrell, James P. Blevins & Robert Malouf. 2009. Parts and wholes: Patterns of relatedness in complex morphological systems and why they matter. In James P. Blevins & Juliette Blevins (Eds.), *Analogy in grammar: Form and acquisition*: 54-82.
- Ackerman, Farrell, & Robert Malouf. 2013. Morphological organization: The low conditional entropy conjecture. *Language*, 89, 3: 429-464.
- Barbu, Ana Maria. 2008. Romanian lexical databases: Inflected and syllabic forms dictionaries. In *Proceedings of the Sixth International Conference on Language Resources and Evaluation (LREC 2008), Marrakech, Morocco*. European Language Resources Association (ELRA).
- Beniamine, Sacha, Martin Maiden & Erich Round. 2020. Opening the Romance Verbal Inflection Dataset 2.0: a CLDF Lexicon. In *Proceedings of The 12th Language Resources and Evaluation Conference*, pp. 3027-3035.
- Bonami, Olivier; Gauthier Caron & Clément Plancq. 2014. Construction d'un lexique flexionnel phonétisé libre du français. In Franck Neveu, Peter Blumenthal, Linda Hriba, Annette Gerstenberg, Judith Meinschaefer and Sophie Prévost, editors, *Actes du quatrième congrès mondial de linguistique française*: 2583-2596.
- Calderone, B., Pascoli, M., Sajous, F., & Hathout, N. (2017). Hybrid Method for Stress Prediction Applied to GLAFF-IT, a Large-scale Italian Lexicon. In J. Gracia, F. Bond, J. P. McCrae, P. Buitelaar, C. Chiarcos, & S. Hellmann (Eds.), *Language, Data, and Knowledge* (pp. 26-41). Cham: Springer International Publishing.
- Cotterell, Ryan; Christo Kirov, Mans Hulden, and Jason Eisner. 2019. On the complexity and typology of inflectional morphological systems. *Transactions of the Association for Computational Linguistics*, 7.
- Kirov, Christo, Ryan Cotterell, John Sylak-Glassman, Géraldine Walther, Ekaterina Vylomova, Patrick Xia, Manaal Faruqi et al. 2018. UniMorph 2.0: universal morphology. *arXiv preprint arXiv:1810.11101*.
- Milin, Petar, Victor Kuperman, Aleksandar Kostic, and R. Harald Baayen. 2009. Paradigms bit by bit: An information theoretic approach to the processing of paradigmatic structure in inflection and derivation. *Analogy in grammar: Form and acquisition*: 214-252.
- Perea, Maria-Pilar, & Hiroto Ueda. 2010. Applying Quantitative Analysis Techniques to La flexió verbal en els dialectes catalans. *Dialectologia et Geolinguistica* 18, 1: 99-114.
- Sims, Andrea D. & Jeff Parker. 2016. How inflection class systems work: On the informativity of implicative structure. *Word Structure* 9, 2: 215-239.
- Stump, Gregory & Raphael A. Finkel. 2013. Morphological typology: From word to paradigm. Vol. 138. Cambridge: Cambridge University Press.