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

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 & Finke 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 of each of them within Romance against the background of the variability observed in genetically diverse inflectional systems.

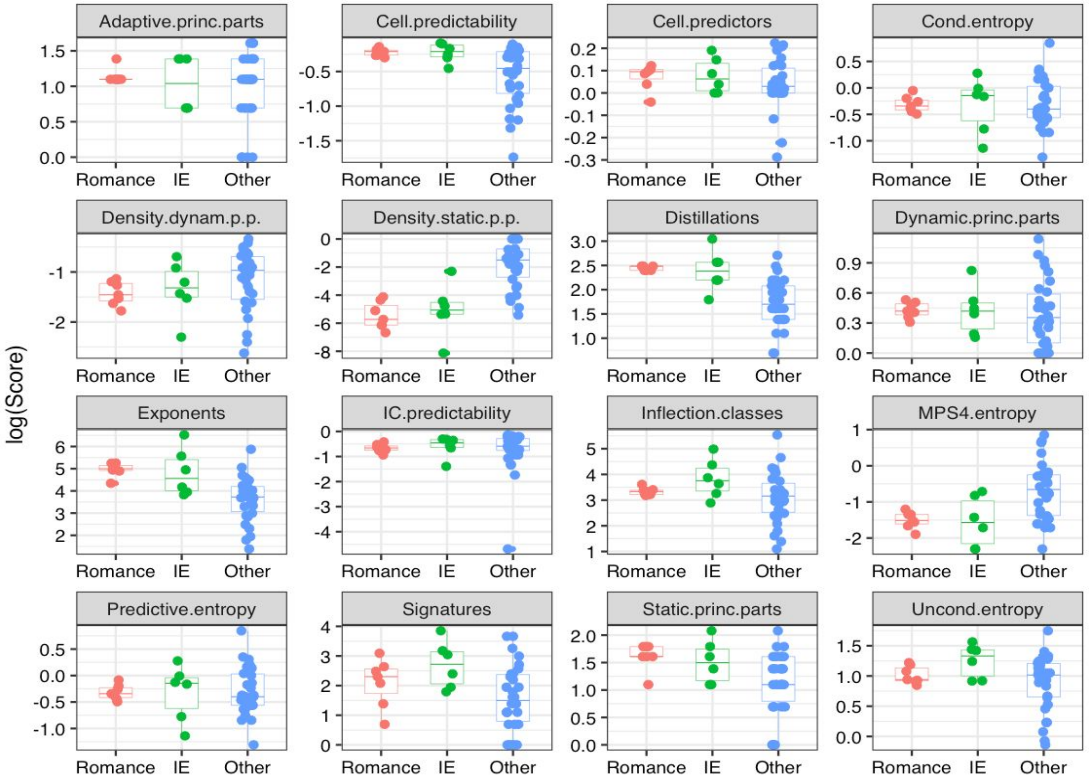

Figure 1: IC traits in Romance, other Indo-European, and unrelated languages

Based on their observed values (see Figure 1), Romance appears to be different from the control sample in most variables in the mean, variance, or both. Difference in variance is particularly interesting because it might suggest differences in relative diachronic stability and as phylogenetic signals of relatedness. Figure 1 shows e.g. that measures like cell predictability and number of distillations show low dispersion in Romance (i.e. high stability), while others, like density of static or dynamic principal parts, show a comparatively large variability (i.e. diachronic instability).

To probe the distinctiveness of Romance further we submitted all variables together to a principle component analysis. The first two components cover over 60% of the variance. They show that Romance verbs have a similar profile and vary within a comparatively small range in the overall design-space (Figure 2), thus further suggesting that these measures are capable of preserving a diachronic signal under the right circumstances.

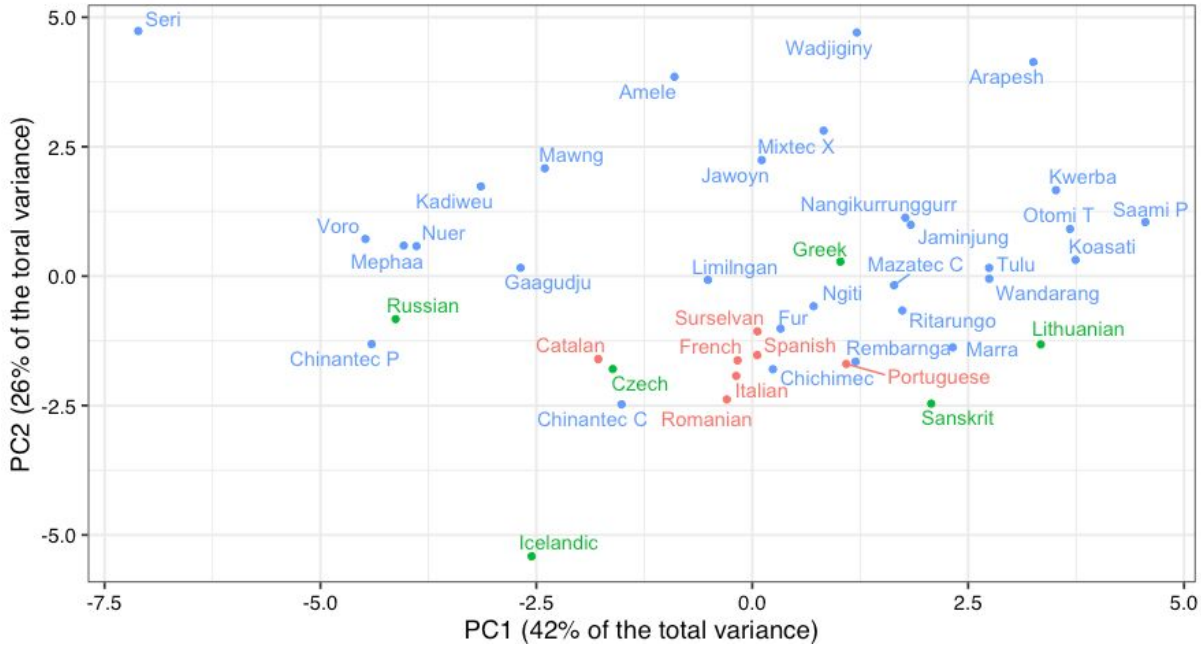

Figure 2: Clustering of inflectional systems in Romance and other languages

If the diachronic stability of (some) paradigmatic predictability relations were confirmed (e.g. if these results were replicated in other families/inflectional systems), entropy and other quantitative traits of inflectional systems could be usefully harnessed by historical morphologists to inform debates on the cognacy and genetic relationship of specific (sub)systems (i.e. tenses, inflection classes, or languages) when the traditional philological or comparative evidence is insufficient or controversial.

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 Blu- menthal, 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 Faruqui 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.