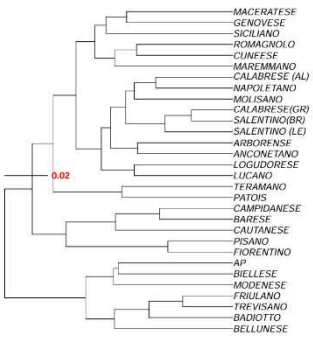
When the Parametric Comparison meets the CP:
a preliminary taxonomy of Italo-Romance varieties

Background: The Parametric Comparison Method (PCM) is an innovative tool for language comparison that strives to reconstruct linguistic phylogeny and to provide new linguistic taxonomies relying on the notion of syntactic parameter (Longobardi 2005; Longobardi & Guardiano 2009, 2017; Guardiano et al. 2020, *i.a.*). To achieve this goal, the PCM is grounded in the Modularized Global Comparison (MGC), testing an indefinite parametric database across an unrestricted language inventory within a single domain (Longobardi 2003). Over the last twenty years, studies primarily involved the nominal domain, releasing a list of 97 parameters tested on 69 languages across 13 language families (Crisma et al. 2020). This presentation, rather, delves into an unexplored syntactic domain, the complementizer phrase (CP), selecting a language sample of Italo-Romance varieties spoken in various regions. The goal is twofold: first, proving that the PCM does not exclusively function with the nominal domain, but can also be successfully expanded to other structural modules, provided that an effective parametric database is built, second, investigating and categorizing the greatly recognized micro-variation in Italo-Romance varieties through a trustworthy parametric tool capable of depicting minimal variational patterns and offering a revised taxonomy.

Building the Parametric System: The formulation of a parametric database regularizing the CP was framed with the cartographic approach (Rizzi 1997, *i.a.*). Each head of the split-CP is independently treated to detect the most salient patterns of variation and the respective parameters regulating them. Broader phenomena (e.g. the realization of a clause-type, the occurrence of a discourse-type exponent etc.) are mapped with core (macro-)parameters, which frame more detailed (micro-)parameters expressing fine-grained patterns of variation (e.g. the operation involved in the realization of clause-type, the syntactic nature of the items involved etc.). The core parameters were constructed according to a parameter *schema* (Longobardi 2005; Gianollo et al. 2008) asking whether a feature F is *grammaticalized*, *checked*, *spread* or *strong*. Therefore, it was necessary to retrieve a bundle of universally definable features and verify whether these operations occur. The selection of the feature analysed for each head shows a clear deviance from Rizzi (1997) who postulated a one-to-one association between a functional feature and a functional head. Once the core featural-driven parameters, regulating major structural-building operations, were established, other parameters, which regulate salient patterns of variation, were framed within them. The valuation of parameters in the *schema* sheds light on other structural phenomena, whose (micro-)parameterization is contingent on the value assigned to the parameters in the *schema*, displaying a significant cross-linguistic variation among the languages tested. In total, 97 parameters were produced and tested.

The Language Database: Contrary to the previous applications of the PCM, which constitute an essential testing ground for standard languages, this presentation prioritized non-standard varieties in an attempt to determine the degree of linguistic relatedness between languages that are unequivocally close given their common ancestor. Thirty dialects were investigated, including the dialects of northern Italy, the dialects of central Italy, the dialects of southern Italy and Sardinian. Data were collected by conducting a recorded oral translation task, corroborated by individual questions aimed to better grasp the variation.

Methods & Results: Each parameter in the database received a value (+, -, 0) for each language tested. Afterwards, the syntactic distance δ for each language pair was calculated by dividing the number of different values by the sum of different and identical values. Ultimately, the syntactic distances were computationally analysed generating the tree in Figure 1 using the UPGMA method.



It is undeniable that broadly speaking the results largely correspond to the expectations based on traditionally recognised classification thereby clearly distinguishing NIDs, CIDs and SIDs. However, some exceptions are noteworthy: (1) the status of Genovese, Cuneese and Romagnolo clustered within SIDs, (2) the distribution of Sardinian varieties, (3) the distribution of Tuscan varieties, (4) the idiosyncrasies of Valdôtain Patois.

Discussion: The phylogenetic tree permits us to draw some preliminary conclusions, as a long as additional research hypotheses to further develop. In the first place, the efficacy of the PCM in detecting cross-linguistic variation and in producing a reliable language taxonomy is attested in relation to the CP as well. Second, it is inevitable to draw a comparison between the tree generated by Guardiano et al. (2020) testing the nominal domain on a set of Italo-Romance varieties with the tree represented above: the DP/NP looks more compact, whereas the results obtained for the CP reveal several inconsistencies, predicting that this structural domain is more sensitive to cross-linguistic variation. Third, given the tree above, it is reasonable to advance a new classification of Italo-Romance based on the CP. In fact, the traditional taxonomies of Italo-Romance are primarily based on phonology or morphology, greatly overlooking numerous syntactic phenomena, specifically if idiosyncratic of a small handful of verities. This parametric investigation, rather, intends to highlight the CP variation and to define a language classification based on phenomena available in this functional area. Lastly, the quantitative analysis presented here seeks to uncover patterns of variation that can serve as a foundation for typological analysis. A key, yet underdeveloped step, is to understand the source of this variation—specifically, to explain why the CP classification appears as it does by identifying the most prominent parametric differences.

Conclusions: Although the investigation is still in its early stages, the results are compelling and offer insights into the linguistic evolution of Italo-Romance. Notably, they reveal several unexpected patterns of variation, which have been effectively analysed using a detailed parametric approach. Further research could delve deeper into this classification to uncover the fundamental sources of variation.

References

Crisma, Paola, Giuseppe Longobardi & Cristina Guardiano. 2020. Syntactic diversity and language learnability. *Studi e Saggi Linguistici* 58(2).

Gianollo, Chiara, Cristina Guardiano & Giuseppe Longobardi. 2008. Three fundamental issues in parametric linguistics. *The Limits of Syntactic Variation*, 109–142. Amsterdam: John Benjamins Publishing Company.

Guardiano, Cristina, Giuseppe Longobardi, Guido Cordoni & Paola Crisma. 2020. Formal syntax as a phylogenetic method. *The Handbook of Historical Linguistics* 2. Wiley Online Library, 145–182.

Longobardi, Giuseppe. 2003. Methods in parametric linguistics and cognitive history. *Linguistic Variation Yearbook* 3, 101–138.

Longobardi, Giuseppe. 2005. A minimalist program for parametric linguistics. *Organizing grammar: linguistic studies in honor of Henk van Riemsdijk* 86. Mouton de Gruyter Berlin, Germany, 407.

Longobardi, Giuseppe & Cristina Guardiano. 2009. Evidence for syntax as a signal of historical relatedness. *Lingua* 119(11), 1679–1706.

Longobardi, Giuseppe & Cristina Guardiano. 2017. Phylogenetic reconstruction in syntax: The parametric comparison method. *The Cambridge handbook of historical syntax*, 241–271. Cambridge University Press. (Rizzi, Luigi. 1997. The Fine Structure of the Left Periphery. In Liliane Haegeman (ed.), *Elements of Grammar* (Kluwer International Handbooks of Linguistics), 281–337. Dordrecht: Springer Netherlands.