

The Parametric Comparison Method (PCM)

The PCM aims to reconstruct the **historical evolution of human languages** by pursuing a **phylogenetics of grammars** (Longobardi & Guardiano 2009).

To demonstrate the **deep phylogenetic potential** of the PCM (Ceolin et al. 2020, 2021), its genealogical and typological scope must be extended to **broader sets of languages**.

The **PCM Hub** has been developed to support systematic **data acquisition** for cross-linguistic analysis and comparison, while incorporating the **theoretical apparatus** of the parametric system (Crisma et al. 2025).

Building a web-based resource

Ideal requirements for a PCM resource

- Data must be **recorded** (and shared by PCM researchers) through a user-oriented web-based interface
- A **structured network** of connections must be established among data, between data and corresponding manifestations, between manifestations and parameters, and among parameters
- The consequences of **parameter dependencies** must be automatically computed to identify neutralized parameters
- The list of parameters and their associated properties must remain expandable and **modifiable** to ensure that the model can adapt to **ongoing changes** due to revisions to the theory and dataset expansion
- Subsets** of languages and/or parameters must be extractable, enabling **comparison** and **analysis**
- Measures** of dissimilarity and the resulting **classifications** must be automatically derived

Features implemented in the PCM Hub

- Intuitive** web app for informants and language specialists optimized for data recording
- An infrastructure that systematically encodes all **connections** and allows for data checking and data analysis
- A procedure designed to avoid human error in **assigning parameter states** and computing parameter dependencies
- An **expandable architecture** that accommodates theoretical revisions and dataset growth
- Flexible **selection and extraction of data** for cross-linguistic and cross-parametric searches
- Built-in scripts for statistical, taxonomic and phylogenetic processing

A brief history of the PCM

- Longobardi (2003)** → **intuition** → abstract syntactic structures can be used as a **powerful tool for phylogenetic analysis**
- Guardiano & Longobardi (2005)** → introduction of **preliminary tools** for parametric comparison:
 - measuring parametric similarity → the notion of *coefficient*
 - testing the method against **conceptual conditions of adequacy** → *taxonomic mistake, Anti-Babelic Principle*
 - comparison of **historical languages**
- Longobardi & Guardiano (2009)** →
 - Humboldt's problem*
 - Extension of the **empirical basis** (languages and parameters)
 - Formalization of *parameter dependencies*
 - Refinement of the **metric tools** → conversion of coefficients into *distances* and parametric *taxonomies*
- Bortolussi et al. (2011)** → *possible languages* and *possible distances* → **probabilistic evaluations**
- Longobardi et al. (2013) and Ceolin et al. (2020)** → classification of **IE languages**:
 - the PCM is able to identify the **main IE subfamilies** from modern languages only
 - syntactic change** from PIE has taken place at a **slower pace** than lexical replacement
 - Homology conjecture*
- Guardiano et al. (2016)** → analysis of **microvariation** → *ultralocality*
- Ceolin et al. (2021)** → **cross-family comparison**
- Further steps** → towards **global coverage**

Language experts (Users): tools and actions

Preliminary information

- Legal protection** → access and use permitted only upon acceptance of the Privacy Policy, Terms of Use, and Data Contributor License Agreement
- Each **User** is assigned (at least) one language
- Each **language** is associated with a set of **metadata** → genealogical classification, standard identification codes, and geographical coordinates

Entering data

- Each **parameter** is associated with:
 - a set of **YES/NO questions** corresponding to its *manifestations*
 - at least one **example** illustrating the answer YES
 - a list of **instructions**
- For each **YES** answer, at least **two examples** are required, including transliteration (if applicable), gloss, and translation
- For each **NO** answer, a list of motivations is provided: Users select one (or more, when possible), and/or provide an example
- The **Comment** field can be used to add further observations
- Supporting resources** for Users → a glossary of technical terminology and guidelines for data entry and glossing

Submitting data

- Two-colour confidence system**:
 - green** → data are complete
 - red** → data require further review
- After **submission**, data can no longer be modified, but remain **viewable and exportable**

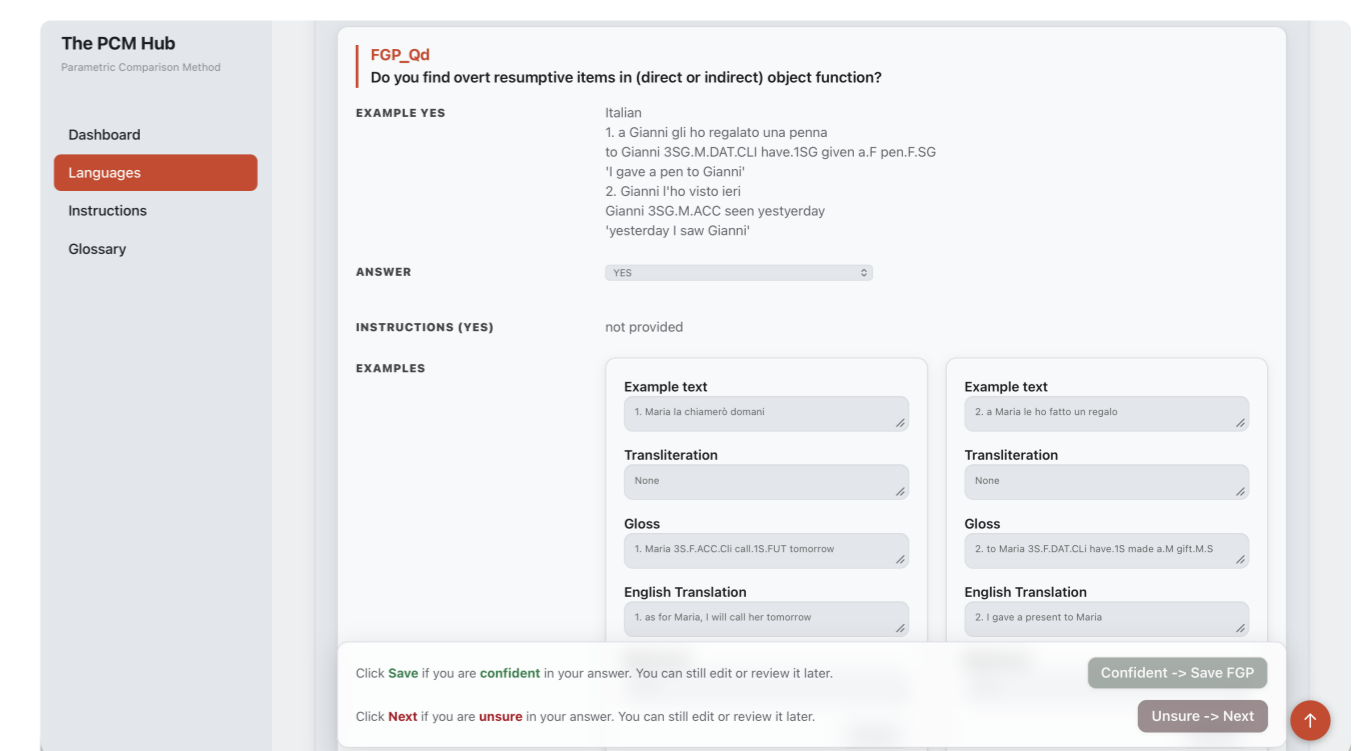


Fig. 1 Data entry interface: answer YES

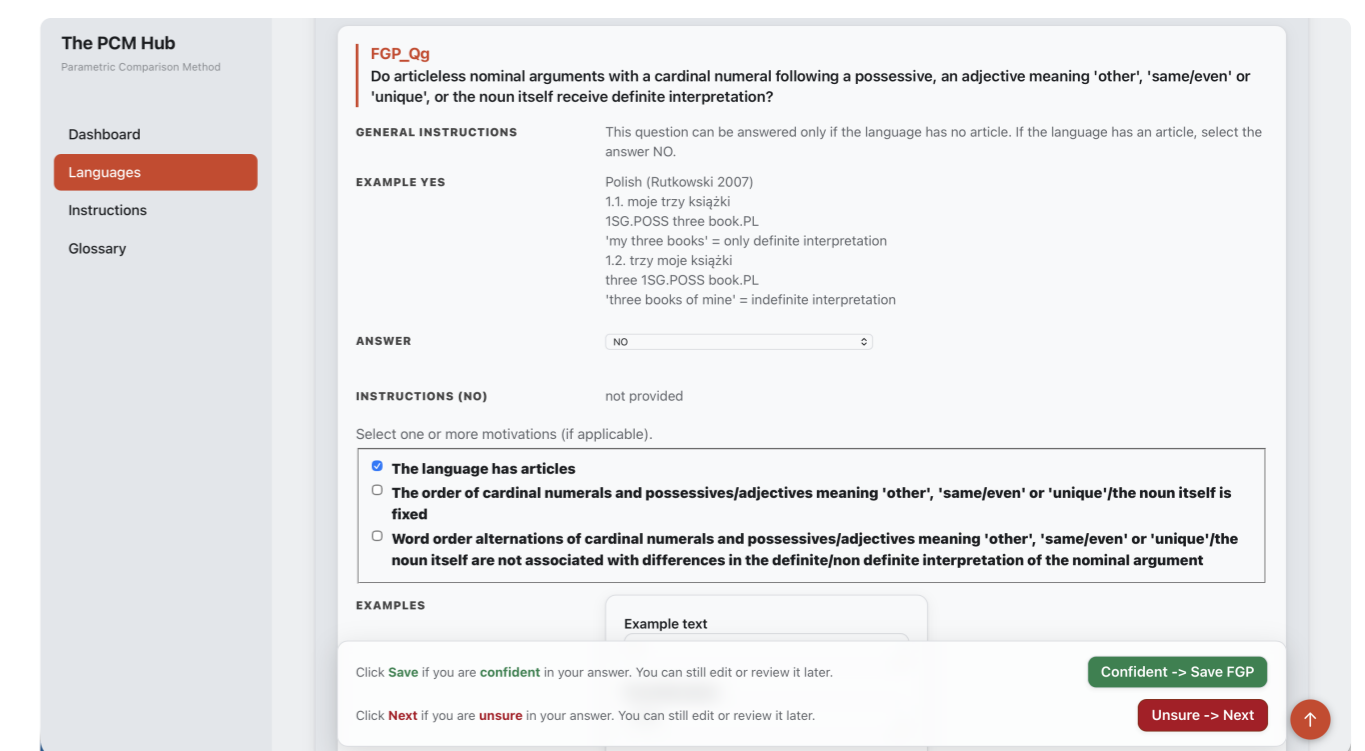


Fig. 2 Data entry interface: answer NO

PCM researchers (Admins): functions and tasks

Assisting Users and reviewing data

- Assign a language to each informant and enter the relevant metadata for each language
- Review User **submissions** to ensure that each example is relevant to the question and consistent with the answer provided
- Ask Users to modify or clarify the examples they entered, if any **inconsistencies** are detected
- Upload and automatically integrate **previously collected PCM datasets** through structured files

Managing the parameter system

- Automatically **compute parameter states** based on User answers and parametric dependencies. At the end of this process, each language is expressed as a sequence of [+], [-], and [0].
- Add and edit** parameters, including implicational condition(s), schema, type and manifestations, or deactivate them **without losing associated data**
- Query the database** and extract specific information about the distribution of parameter states and their manifestations through a system of filters

Performing cross-linguistic comparison

- Compute pairwise **parametric distances** and perform **multivariate analysis**, including Principal Component Analysis (PCA) and hierarchical clustering (HC)
- Export data in **machine-readable formats** for further processing with external tools
- Export materials in non-editable formats for **dissemination**

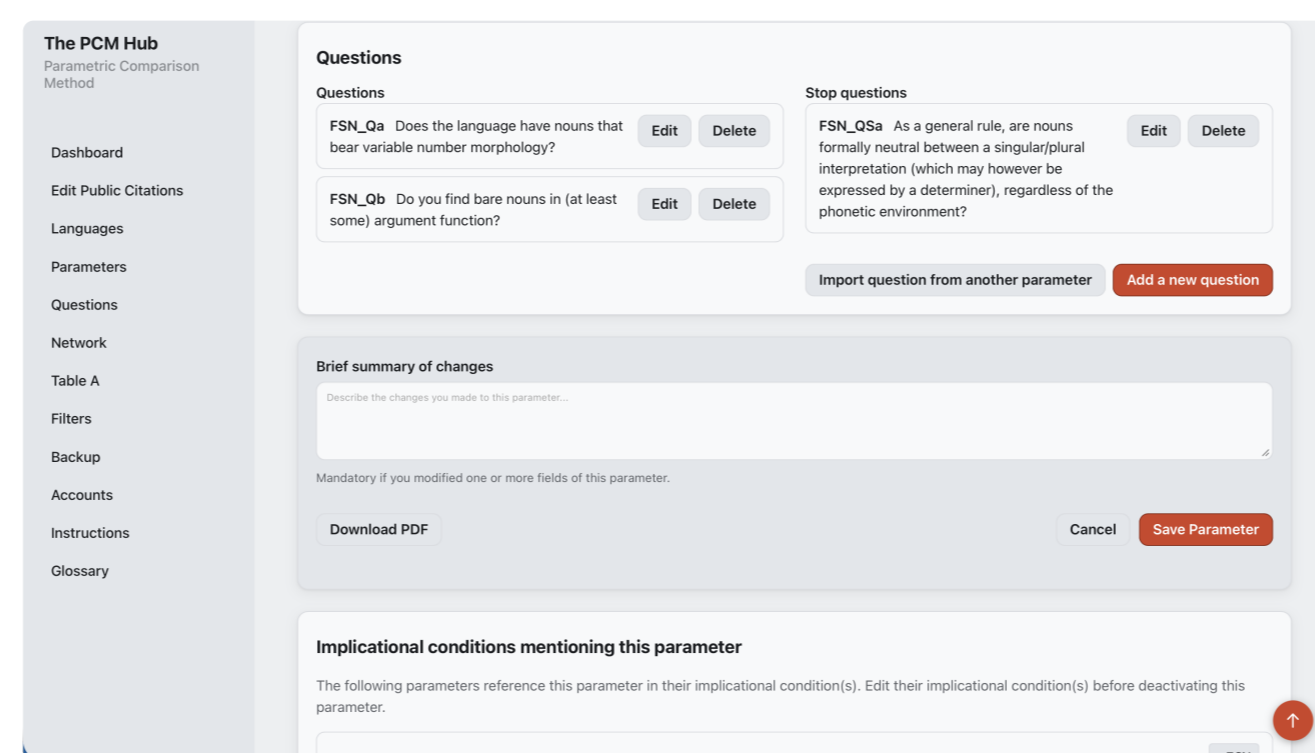


Fig. 3 Parameter edit interface

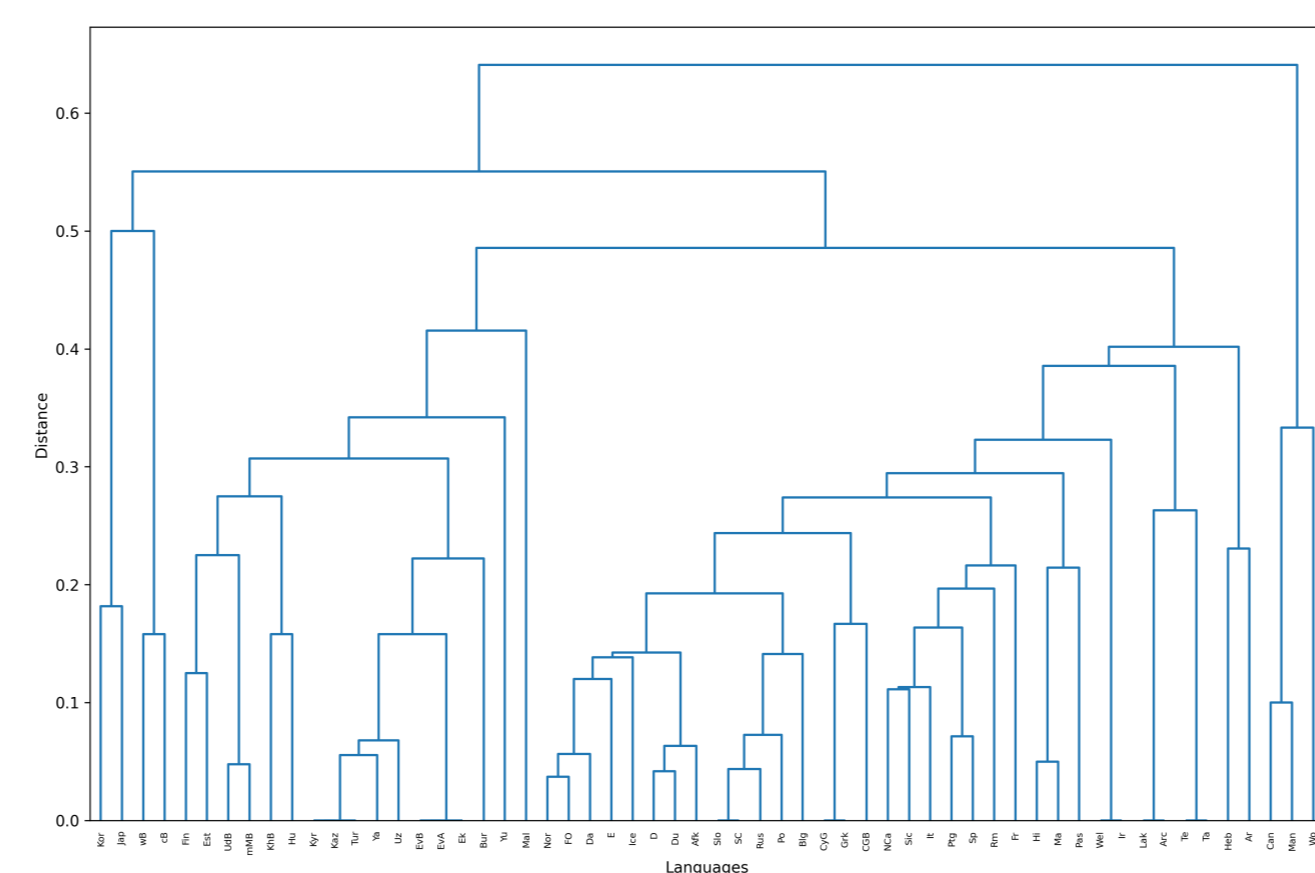


Fig. 4 HC dendrogram – dataset from Ceolin et al. (2021), parameter list from Crisma et al. (2025)

Technical specs

- Scalable platform built on the **Django** framework
- Relational database** with a **PostgreSQL** backend
- Architecture following a clear separation between data management, application logic, and user interface, ensuring stability, maintainability, and future extensibility
- User-friendly interface developed with **responsive web technologies** (HTML5, CSS, JavaScript)
- Structured **data querying**
- Automatic setting** of parameter states according to the User's answers
- Automatic assignment** of [0]s through a logical engine based on **directed acyclic graphs (DAGs)** to parameters whose implicational conditions are not satisfied
- Open source software** and technologies
- Platform design, structure, and all content protected by **copyright**

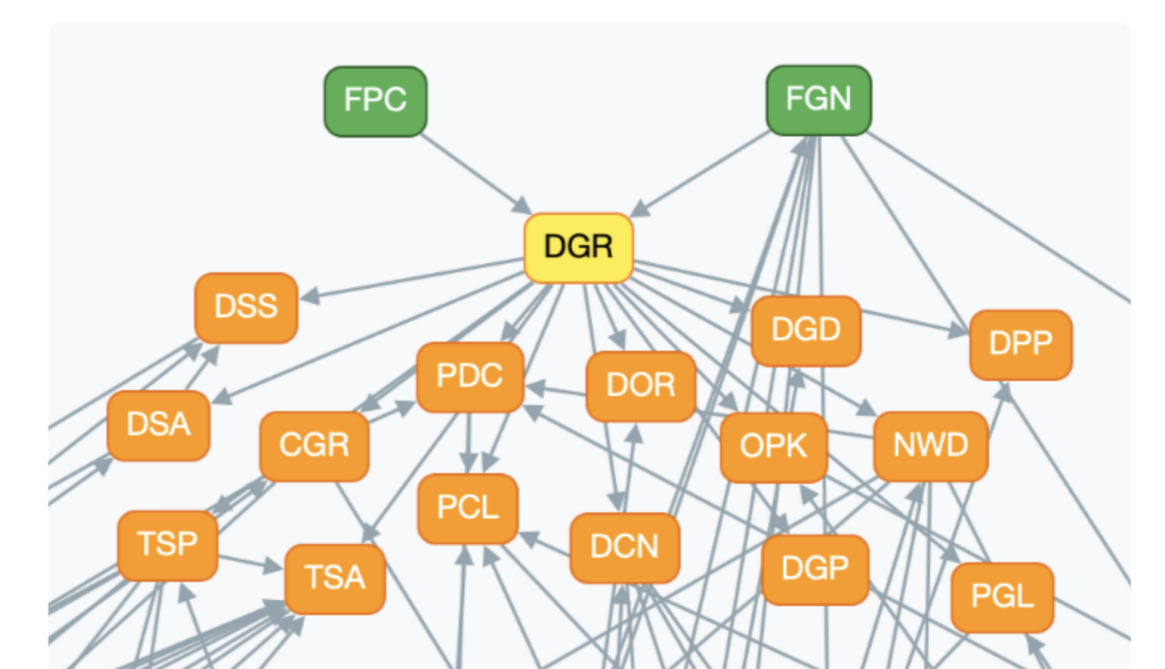


Fig. 5 Network of parametric dependencies

Links

- To access the PCM Hub, contact us at: pcm_lab@unimore.it
- The link to the PCM Hub will be made available at: www.parametriccomparison.unimore.it, where published PCM materials and resources can also be accessed.



References

- Bortolussi, L., A. Sgarro, G. Longobardi & C. Guardiano. 2011. How many possible languages are there? In ed. by G. Bel-Enguix et al., *Biology, Computation and Linguistics*, 168–179. Amsterdam: IOS.
- Ceolin, A., C. Guardiano, M. A. Irimia & G. Longobardi. 2020. Formal syntax and deep history. *Frontiers in Psychology* 11, 488871.
- Ceolin, A., C. Guardiano, G. Longobardi, M. A. Irimia, L. Bortolussi & A. Sgarro. 2021. At the boundaries of syntactic prehistory. *Philosophical Transactions of the Royal Society B* 376, 20200197.
- Crisma, P., G. Fabbri, G. Longobardi & C. Guardiano. 2025. What are your values? Default and asymmetry in parameter states. *Journal of Historical Syntax* 9, 1–26.
- Guardiano, C. & G. Longobardi. 2005. Parametric comparison and language taxonomy. In ed. by M. Battlori et al., *Grammaticalization and parametric variation*, 149–174. Oxford: OUP.
- Guardiano, C., D. Micheloudakis, A. Ceolin, M. Irimia, Longobardi, Giuseppe, N. Radkevich, G. Silvestri & I. Sitaridou. 2016. South by Southeast. A syntactic approach to Greek and Romance microvariation. *L'Ybica diabetologica* LXXVII, 95–166.
- Longobardi, G. 2003. Methods in parametric linguistics and cognitive history. *Linguistic Variation Yearbook* 3(1), 101–138.
- Longobardi, G. & C. Guardiano. 2009. Evidence for syntax as a signal of historical relatedness. *Lingua* 119, 1679–1706.
- Longobardi, G., C. Guardiano, G. Silvestri, A. Boattini & A. Ceolin. 2013. Toward a syntactic phylogeny of modern Indo-European languages. *Journal of Historical Linguistics* 3(1), 122–152.