# Challenges for Computer-Assisted Language Comparison

Frederic Blum¹², Johann-Mattis List¹², Alejandra Regúnaga³

¹ Max Planck Institute for Evolutionary Anthropology

² University of Passau

³ Universidad Nacional de la Pampa

## Abstract

In contrast to purely computational approaches in historical linguistics that try to do without human annotation, computer-assisted language comparison aims to reconcile computational and classical approaches in historical language comparison by providing interactive workflows in which data are passed back and forth between humans and machines (List 2017b). Computational approaches are mostly employed to preprocess linguistic data, while interfaces then allow experts to refine and correct computational annotation.

While some workflows and tools have been published that illustrate the benefits of computer-assisted as opposed to purely computer-based or purely manual approaches in historical linguistics (Wu et al. 2020; Hill and List 2017; List 2017a; Segerer and Flavier 2015; Starostin 2000), there remain many tasks in historical linguistics where computational and computer-assisted approaches are lacking so far. These include, for example, *semantic reconstruction* (Urban 2015), *phonological reconstruction* (Anttila 1972), and *sound law induction* and the establishment of *relative chronologies* (Fortson 2014).

In the workshop we want to invite both classical and computational linguistics to present open problems and potential solutions in historical language comparison that have so far not been addressed sufficiently in computer-assisted approaches. Our hope is that we find a good mix of contributions in which classical linguists present problems that cannot be addressed by computational approaches and illustrate how they solve them manually, while we also hope for contributions by computational linguists who share recently developed workflows that can be applied to data in historical linguistics. In both cases, we hope that scholars present problems and solutions in such a form that they can be applied across different languages and families, rather than solutions focusing on individual problems in particular languages and particular language families.

## References

Anttila, Raimo. 1972. *An Introduction to Historical and Comparative Linguistics*. New York: Macmillan.

Fortson, Benjamin W. 2014. “Relative Chronology.” In *Encyclopedia of Ancient Greek Language and Linguistics. Volume 3*, edited by Georgios K. Giannakis, 219–24. Leiden; Boston: Brill.

Hill, Nathan W., and Johann-Mattis List. 2017. “Challenges of Annotation and Analysis in Computer-Assisted Language Comparison: A Case Study on Burmish Languages.” *Yearbook of the Poznań Linguistic Meeting* 3 (1): 47–76.

List, Johann-Mattis. 2017a. “A Web-Based Interactive Tool for Creating, Inspecting, Editing, and Publishing Etymological Datasets.” In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics. System Demonstrations*, 9–12. Valencia: Association for Computational Linguistics.

———. 2017b. *Computer-Assisted Language Comparison. Reconciling Computational and Classical Approaches in Historical Linguistics [Research Project, 2017–2022]*. Leipzig: Max Planck Institute for Evolutionary Anthropology.

Segerer, Guillaume, and S. Flavier. 2015. “RefLex: Reference Lexicon of Africa.” Paris; Lyon. 2015. [http://reflex.cnrs.fr](http://reflex.cnrs.fr/).

Starostin, Sergej Anatolévič. 2000. *The STARLING Database Program*. Moscow: Online ressource; RGGU. [http://starling.rinet.ru](http://starling.rinet.ru/).

Urban, Matthias. 2015. “Lexical Semantic Change and Semantic Reconstruction.” In *The Routledge Handbook of Historical Linguistics*, edited by Claire Bowern and Bethwyn Evans, 374–92. London; New York: Routledge.

Wu, Mei-Shin, Nathanael E. Schweikhard, Timotheus A. Bodt, Nathan W. Hill, and Johann-Mattis List. 2020. “Computer-Assisted Language Comparison. State of the Art.” *Journal of Open Humanities Data* 6 (2): 1–14.