



*cordially invites you to a seminar on*

## Increasing the Comparability of Linguistic Data

by

*Johann-Mattis List*

Max Planck Institute for the Science of Human History (Jena, Germany)



### ABSTRACT

During the last 200 years, linguists have been collecting an abundance of data on the world's languages, ranging from small word lists up to detailed grammatical descriptions accompanied by large dictionaries. Unfortunately, these data are often presented in very idiosyncratic ways: transcription systems are largely varying, glossing of concepts is done in an ad-hoc manner, and even names for one and the same language variety often differ widely. This lack of comparability has great consequences for our research: qualitative work requires the knowledge of areal- and source-specific idiosyncrasies before it can be carried out, making certain fields only accessible to highly trained experts, and quantitative research requires that the data are explicitly made comparable, leading to errors and large gaps in cross-linguistic databases. In the talk, I will discuss the problem of data comparability in more detail and present some ideas how we can increase the comparability of existing data and future data in our scientific field.

### ABOUT THE SPEAKER

Johann-Mattis List is a senior scientist at the Max Planck Institute for the Science of Human History. From April on, he will pursue an ERC-funded research project on Computer-Assisted Language Comparison (CALC, 2017-2022). In this project, he will develop methods and interfaces that bridge the gap between computational and classical approaches to historical linguistics in order to shed light on the history of the Sino-Tibetan language family.

In his research, he follows a data-driven, empirical, and quantitative perspective on language change and language history. In contrast to pure computational approaches, he keeps his research closely connected to traditional historical linguistics and theory, following a computer-assisted rather than a computer-based framework of quantitative research in historical linguistics. This research is reflected in various tools and databases, such as the LingPy software library for computational historical linguistics (<http://lingpy.org>), the CLLD Concepticon (<http://concepticon.cld.org>) and the Database of Cross-Linguistic Colexifications (CLICS, <http://clics.lingpy.org>).

DATE:  
Friday,  
24 February 2017

TIME:  
4:30pm

VENUE:  
HSS Conference Room  
(HSS-05-57)

ALL ARE WELCOME