

# Fundamentals of computer-assisted language comparison

Tiago Tresoldi, Mei-Shin Wu, and Nathanael E. Schweikhard

## Abstract

The comparative method, the core of historical linguistics, originated in the 19th century, linguists compared lexical items in different languages to examine the genealogical relationships among languages. Through repetitively analysis various sets of lexical items among target languages, linguists have categorized various large and small language families in the world. Even though, we have gained in-depth knowledge about certain language families after more than 300 years of studies, but many more questions about the rest of the world await linguists to answer. The issue of comparative method, which relies on experts' manual judgements, has reaches its practical limit as the amount of language data piles up through time.

On the other hand, digitalized language data are also accumulated and able to access easily, it opens up the era which experts can focus on applying their knowledge to examine the results, and let computer power to handle the repetitive and boring tasks. In our term, it is called **computer-assisted language comparison**. Combining qualitative and quantitative approaches is not a completely new concept, it has been widely used in biology and natural language processing. With the help of computing power, the cross-linguistic data can be easily compiled and manipulated from various individual studies, the cognate detections as well as sound correspondences are detected in merely few hours, plus the algorithms can be flexibly connected into workflows and apply onto many linguistic aspects. The experts could evaluate and make changes to the data at any stage so to increase the accuracy.

In today's three talks, we organized as the following:

- (A) Softwares and methods (Tiago Tresoldi)
- (B) Interface and workflows (Mei-Shin Wu)
- (C) Data and Modeling (Nathanael E. Schweikhard)

With these three talks, we introduce the core value, the methods, the tutorial and the applications of our lab and the tools we are developing.

18 , , , ,