

Spring school 2022 afternoon workshop
“Interdisciplinary and digital Uralistics”

Organisers:

University of Turku: Outi Vesakoski, Dept of Finnish language and Finno-Ugric linguistics, and the BEDLAN team (<https://bedlan.net/>): Timo Rantanen and Meeli Roose (Dept of Geography), Jenni Santaharju (UH, Dept of Biology)

Uppsala University: Michael Dunn, Department of Linguistics and Philology (LingFil), General Linguistics section; Yingqi Jing (LingFil, General Linguistics / Language Technology sections); Miina Norvik (Department of Modern Languages, Finno-Ugric language section)

The course includes:

1. Lectures: Interdisciplinary perspectives on human history in the North-Western Eurasia
2. Introduction of two new data sets to support Uralic linguistic studies:
 - Uralic typological data “UraTyp”
 - Cartographical data of Uralic language speaker areas
3. Tutorials: Basics of visualisation and statistics of linguistic typological data using R
4. Tutorials: Basics of GIS methods with QGIS software

On-line and *in situ* teaching:

Monday lectures can be followed via Zoom. Tuesday-Friday will also be broadcast via Zoom, but interactive participation for on-line students is not guaranteed as tutorials include not only short lectures but also practical exercises. The instructions for the practical exercises will be provided via Moodle, and all students can follow them with their own computers. However, we can only offer technical help to students participating *in situ*. Chat and zoom can be used for on-line students to help each other.

Schedule in short:

Monday:

- Introduction to the course. Short review of environmental, cultural, linguistic and genetic history of the North-Western Eurasia.
- Lectures by Outi Vesakoski and Kristiina Tambets (Univ of Tartu)

Tuesday-Wednesday:

- Uralic typological data and data visualisation with R.
- Tutorials by Michael Dunn, Yingqi Jing and Miina Norvik and Jenni Santaharju

Thursday-Friday:

- Spatial data of Uralic language speaker areas and map making with QGIS
- Tutorials by Timo Rantanen and Meeli Roose

Before the course:

You don't need to have skills in R or GIS, but you should

1. Read the Moodle materials (2 "need-to-know" articles)
2. Study these web apps
 - Uralic Language speaker areas as well as some other Uralic data is found from the **Uralic Historical Atlas**: sites.utu.fi/urhia
 - Uralic typological data is found from **Uralic Areal Typological Online** <https://uralic.cld.org>
3. Please complete the course preparation questionnaire by the beginning of the course in Moodle or by sending the answers to Jenni.Leppanen at Helsinki.fi
4. Get your computer ready:
 - Install R (<https://cran.r-project.org/mirrors.html>), Rstudio (<https://www.rstudio.com/products/rstudio/download/#download>)
 - Download the related data sets and run the test R-code (Rmarkdown template: https://github.com/bedlan/tartu-spring-school-R-workshop-2022/blob/main/Rmd_template/Rmarkdown-template.Rmd). Open this using RStudio
 - Install QGIS 3.24.1 'Tisler' on your computer <https://www.qgis.org/en/site/forusers/download.html>

If you are not successful with installing the programs, visit the Zoom trouble-shooting sessions:

- R-problems (Yingqi Jing): Thursday 1st April, 2-3pm Swedish time or 3-4pm Estonian time
<https://uu-se.zoom.us/j/6891098977>
- QGIS-problems: Meeli.Roose at utu.fi. /Timo.Rantanen at utu.fi

Learning aims

The student should achieve the following key skills

Part 1. Interdisciplinary Uralistics: and general aims:

1. After the course, you will be able to describe the main points of the **environmental history** of the North-Western Eurasia.
You will be able to chat over coffee breaks on how the climatic and ecological variation could have affected human biological and cultural evolution.
Note the human evolution in interaction with the environment fluctuation is not much studied yet, thus we cannot offer facts but food for thoughts and for future studies!
2. You will know the major turns in the **cultural prehistory** of the area.
After the course you will be able to follow archaeological presentations, for you know the basics of archeological periodisation: What is Mesolithic and Neolithic era, what

are Metal ages and how all these are characterized? You know which societies were hunter-gatherers, which were farmers.

3. -You can talk about different hypotheses on **Uralic linguistic history**: place-of-origin of Proto-Uralic, chronology of the family, structure of family.
You are able to read phylogenetic trees and outputs of admixture models.
You have at least 2 ideas on how to study different hypothesis about Uralic linguistic history.
4. You can explain to your colleagues what are mitochondrial, Y-chromosomal and autosomal data and why all of them are used for studies of human diversity.
You know what is the difference between modern DNA and aDNA.
You are able to read the main genetic result figures.
You can pinpoint why Y-chromosomal haplotype N and “Siberian component” are discussed when talking about genetics of Uralic speaker populations.
5. You are able to identify the major components of the **linguistic and genetic diversity** of the area.
6. After the course you feel that you know enough of each of these disciplines to follow scientific talks on the matter. You will be able to bring in the linguistic perspective of Uralic history if brain storming with students of archaeology or genetics.

Part 2. R and Uralic Typological data workshop

1. You are able to describe to your colleague what is UraTyp data. What does it include, how is it collected.
2. You can apply the basic principles of ‘tidy data’ to organising your data for analysis and visualisation (Wickham 2014). You will have some practical experience in how to make ‘untidy’ data into tidy data, and will know the sorts of things you can do with it.
3. You will be familiar with the ideas behind compiling and using quantitative typological data. This includes both working with pre-existing databases (in particular, UraTyp data from both Zenodo/GitHub repositories and from the Uralic Areal Typology web app), and having practical ideas about extending pre-existing databases and creating new ones.
 - You are able to download data from zenodo (<https://zenodo.org/record/5910343>)
 - You have created at least 2 ideas how you could use UraTyp data and Uralic Areal Typology web app in your studies.
4. You will know how to get started with producing quantitative analysis and data visualisation in a manner which is maximally simple, reproducible, and collaborative (Wilson et al. 2017).

Part 3. Cartographical workshop

1. You will be able import geospatial data into QGIS
2. You are able to create and edit both point and polygon data; and you understand the difference.

3. You are able to create maps with QGIS
4. You know the steps in importing, creating, curating and visualizing geospatial data
5. You can explain why spatial data platforms are needed.
6. You have given 2 ideas on how URHIA (Uralic Historical Atlas) could be tuned to better serve students of Uralic linguistics (work in progress, your ideas are valuable!)

Core pre-reading material:

Rantanen, T., Vesakoski, O. & Ylikoski, J. 2022: Mapping the distribution of the Uralic languages. Marianne Bakró-Nagy, Johanna Laakso & Elena Skribnik (eds.), *The Oxford Guide to the Uralic Languages*. Oxford University Press. Preprint: https://www.academia.edu/70120154/Mapping_the_distribution_of_the_Uralic_languages

Norvik, M., Jing, Y., Dunn, M. Forkel, R., Honkola, T., Klumpp, G., Kowalik, R., Metslang, H., Pajusalu, K., Piha, Saar, E., Saarinen, S. & Vesakoski, O. Uralic typology in the light of new comprehensive data set, forthcoming, PDF attached

Nice-to-know reading material:

Rantanen, T., Tolvanen, H., Roose, M., Ylikoski, J. & Vesakoski, O. Best practices for spatial language data harmonization, sharing and map creation – a case study of the Uralic languages (conditionally accepted; PDF attached)

Moilanen, U., Pesonen P., Norvik, M., Saipio, J., Vesakoski, O., Immonen, V. & Onkamo, P. 2021: New tools for studying Finnish archaeology and Uralic languages. *Antiquity* 95(323), E30. <https://doi.org/10.15184/aqy.2021.113>.

Roose, M., Nylén, T., Tolvanen, H. & Vesakoski, O. 2021: User-centred design of multidisciplinary spatial data platforms for human-history research. *SPRS International Journal of Geo-Information* 10(7): 467. <https://doi.org/10.3390/ijgi10070467>.

Wickham, Hadley. 2014. “Tidy Data.” *Journal of Statistical Software* 59 (September): 1–23. <https://doi.org/10.18637/jss.v059.i10>.

Wilson, Greg, Jennifer Bryan, Karen Cranston, Justin Kitzes, Lex Nederbragt, and Tracy K. Teal. 2017. “Good Enough Practices in Scientific Computing.” *PLOS Computational Biology* 13 (6): e1005510. <https://doi.org/10.1371/journal.pcbi.1005510>.