Mapping the distribution of the Uralic languages

Timo Rantanen, Outi Vesakoski & Jussi Ylikoski

The maps in *The Oxford Guide to the Uralic Languages* (with the exception of those in chapter 2) have been produced within the project "Geographical Database of the Uralic Languages" by the interdisciplinary BEDLAN research team (Biological Evolution and Diversification of LANguages) and Jussi Ylikoski. The work has been financed by the Kone Foundation (Outi Vesakoski), UiT – The Arctic University of Norway, and the University of Oulu (Jussi Ylikoski), as well as the Finno-Ugrian Society and the University of Turku (Timo Rantanen). The maps are based on earlier studies on Uralic languages, but all underlying data have been updated in line with the information and comments provided by the authors of the present volume. The cartographical digitalization work has been conducted using geographic information systems (GIS). With GIS it has become easy to compile, combine, modify, and visualize detailed spatial information. The full potential of the GIS approach has not been previously used for language distribution areas. The detailed description of the Geographical Database of the Uralic Languages will be published elsewhere (Rantanen et al., submitted ms). The compiled data sets are freely available in GIS format (Rantanen et al. 2021) as well as via an interactive spatial data platform (Uralic Historical Atlas 2021).

There is a long tradition of presenting the geographical distribution of Uralic languages as printed maps and descriptions. As for the most important geographical depictions of the entire language family, the most reliable and comprehensive maps have been created and published by the Finno-Ugrian Society (Helsinki) ever since Otto Donner's map *Verbreitung der Finnisch-ugrischen völker* ('Distribution of the Finno-Ugric peoples') and its accompanying written account (Donner 1885, 1886). While Donner's description did not include the Samoyed branch (cf. chapter 34), later maps, such as "Geographical Distribution of the Uralic languages" edited by Janhunen (1980) and revised by Grünthal and Salminen (1993) have depicted the Uralic family in its entirety. As for the individual branches of the family, the most exemplary modern map is that of Finnic by Grünthal and Sarhimaa (2004/2012).

Modern maps often stem from the earliest depictions of the speaking areas. As they copy earlier sources without critical evaluation, earlier cartographical decisions are repeated from map to map. However, there are also maps that have adopted the most recent knowledge of the past and present distribution areas of individual Uralic languages. The general pattern is naturally that the maps for individual languages include more precise and updated

information than the general overviews of the family. For technical reasons (restrictions in visualization), the detailed information and latest amendments in the individual language maps have not been transferred on to maps of the entire family. The Geographical Database of the Uralic Languages is a synthesis of the existing information for individual languages. With this database and modern cartographical methods (GIS), it is then possible to produce overview maps with a selected level of detail.

The Geographical Database of the Uralic Languages includes information both on the traditional speaker areas of each language and distribution of the more current speaker areas. The aim of the traditional distributions is to show the widest known distribution of a given language. In practice, the outcome depends entirely on the available sources, and is usually equivalent to a representation of the situation around the beginning of the twentieth century. The information for current speaker areas also comes to some extent from the earlier maps and descriptions, but mainly from the other authors contributing to *The Oxford Guide to the Uralic Languages*.

The overall digitalization process involved multiple phases including digitalization, modification, and visualization of the geographical data using ArcGIS software. In order to make it possible to use spatial data sets in GIS, the source data (scanned and electronic language maps) were tied to an appropriate coordinate system by georeferencing. In this process, the exact location on Earth was determined with selected control points from already referenced maps (e.g. online base maps such as National Geographic, OpenStreetMap). Then, the speaker area distributions were digitized. Here, digitization refers to the process whereby speaker areas are picked from a georeferenced map as polygon objects. The new data were defined in vector format, which enables the representation of the exact shape and location of objects and thereby creates a good basis for further comparisons of spatial data and map-making. Frequently, there were multiple different spatial extents of language distributions, all of which were collected in the database.

The collected language distributions were visualized as overlapping layers with carefully selected supplementary features. In many cases these included the layers of physical geography (land and sea areas, inland waters, and the Ural Mountains) and current and past settlements, administrative borders, as well as nomenclature. Supplementary data were mainly freely available and downloaded from web portals (Natural Earth, DIVA-GIS) and from online services of ESRI (Multi-Directional Hillshade). At times, the spatial data sets were partly incomplete, in which case some modifications and improvements were made. The primary purpose of the visualizations was to illustrate the distributions of the languages as clearly as possible and thus systematic attempts to mitigate the deficiencies of background features were avoided.

After the language distributions in the map were visualized, the outputs were sent to the authors of each language description for evaluation. There were often considerable differences between the proposed geographical extents of language distributions in the original studies. Therefore, the authors were asked to decide which of the numerous options

were the most appropriate and, if necessary, how the boundaries of language areas should be modified. Their other task was to comment on the other factors to be represented on the map, including environmental factors, settlement information, and orthography of the place names. The questions were customized separately for particular languages. As a result, in some cases the geographical distributions strictly follow earlier sources, but in other cases the authors provided new, unpublished distributions for the languages. Most commonly, the final outcome is based on original sources, but with varying modifications proposed by the authors. The sources of the maps and the impact of each author are comprehensively indicated in the figure legends.

In the finalization process we constructed and visualized each map individually. This process involved decisions on suitable scale, map projection, and overall layout. For maximal uniformity and clarity of the maps, we made the following decisions: 1) The language distributions are illustrated as speaker areas instead of being presented as points around the settlements. On this basis, each language or dialect has been visualized with a certain colour or pattern. 2) The maps are divided into three categories: individual language maps, maps for the main branches of Uralic, and an overall map of the whole language family. All maps are based on the same source data, but the most detailed information exists in the individual language maps, which predominantly also include the distributions of the dialect areas. 3) In this volume, overlapping language areas are not shown on the map, even though bi- and multilingualism commonly occur in the region.

In all, the maps are created to support text chapters. The choice of background features, details of outputs, and the number of maps were made in consultation with the authors of each language chapter. For example, in some cases previous and current language distributions were represented on the same map, but in other cases different time periods were displayed on separate maps.

It is a universal fact that maps are generalized and simplified views of a phenomenon and they are created according to a specified purpose. Here, the purpose was to represent geographical distributions of the Uralic languages in past and present time by relying on the best possible available knowledge. The general guideline of the process was to show accurate information but to avoid overly precise details. The final outcome is highly dependent on the quality of available data, but there are many other things which also had significant impacts on the visualizations and which one should be aware of when viewing the maps and drawing conclusions.

Firstly, the maps are estimations of reality, no matter whether they display languages or ethnicities or, for example, distribution ranges of animal or plant species. The borders of distribution ranges of any species do not mean that the objects in question occur over the whole area: Humans and their languages and cultures are spread much like other species, aggregated in the landscape with gaps in between. Neither can the borders of a map be considered the ultimate outer limit of the occurrence area, as surely some animals or speakers exist or will migrate outside this area. This point is important especially in the case of

mapping migrating populations, such as nomadic reindeer herders. In the absence of sufficient and reliable data, some maps knowingly conform with traditional yet somewhat distorted portrayals of language areas, instead of presenting partly impressionistic revisions. Further, in the maps of the present volume, large cities, such as Moscow or Helsinki, have not been included as speaker areas of minor Uralic languages, even though they may have significant populations of these speakers. When necessary, the aggregations outside the main distribution ranges are explained in the text.

Secondly, it must be noted that the size of the speaker area certainly does not correspond with the number of speakers of a particular language. The more detailed speaker areas are available of the more thoroughly mapped western languages, which are also often languages with a large number of speakers, such as Mordvin and many Finnic languages. Instead, the less-studied languages, such as Nganasan and Nenets, are very generally defined with extensive speaker areas. Further, the areas with high linguistic diversity have traditionally been defined with more precise language distributions. An illustrative example is a comparison of the languages spoken in the middle Volga river area and northernmost Eurasia.

Thirdly, minority languages often occur without clearly distinguishable borders, and dialect continua and bilingualism are common phenomena. Even though the language distributions are visualized here with strict borders, the natural smooth transitions of language varieties must be kept in mind.

The maps in this volume represent the state-of-the-art knowledge of the geographical distribution of the Uralic languages. It is clear that there is room for further language geography studies on Uralic languages and with the data in digital format it is feasible to complement them with details, corrections, and additions. The possibility of getting more precise information about earlier speaker areas is naturally limited, but future studies will certainly add to our knowledge of the ancient distributions of Uralic languages.

Acknowledgements

References

Donner, Otto (1885). *Verbreitung der Finnisch-ugrischen völker. Karte entworfen vom secretär der fin. ugr. gesellschaft 1885*. Helsinki: Finnisch-Ugrische Gesellschaft.

Donner, Otto (1886). 'Die Finnisch-ugrischen Völker', *JSFOu* I: 120–129.

Grünthal, Riho, and Tapani Salminen (eds.) (1993). *Geographical distribution of the Uralic Languages*. Helsinki: Finno-Ugrian Society.

Grünthal, Riho, and Anneli Sarhimaa (eds.) (2004/2012). *Itämerensuomalaiset kielet ja niiden päämurteet*. Helsinki: Suomalais-Ugrilainen Seura.

Janhunen, Juha (ed.) (1980). *Geographical Distribution of the Uralic languages*. Helsinki: Finno-Ugrian Society.

Rantanen, Timo, Harri Tolvanen, Meeli Roose, Jussi Ylikoski, and Outi Vesakoski (submitted ms). 'Best practices for spatial language data harmonization, sharing and map creation - a case study of Uralic'.

Rantanen, Timo, Outi Vesakoski, Jussi Ylikoski, and Harri Tolvanen (2021). 'Geographical database of the Uralic languages (Version v1.0)'. Zenodo. http://doi.org/10.5281/zenodo.4784188

Uralic Historical Atlas (2021). https://sites.utu.fi/urhia/