

Geographical database of the Uralic languages

Timo Rantanen, Outi Vesakoski, Jussi Ylikoski, Harri Tolvanen

How to cite

When you use the datasets or maps, please cite to the following paper:

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" (submitted manuscript).

The manuscript introduces the whole process from the language data collection, procession and visualization until releasing the data and the maps.

Please include the DOI provided for the Geographical database of the Uralic languages in this release in your citation as well.

Overview

The Geographical database of the Uralic languages consists of past and current distributions of the Uralic languages both as the original digital spatial datasets and as finalized maps. The database has been collected by the interdisciplinary BEDLAN (Biological Evolution and Diversification of LAnguages) research team in collaboration with experts of Uralic languages. The work has been financed by the University of Turku (UTU–BGG), Kone Foundation (UraLex, AikaSyyni), the Academy of Finland (URKO), UiT – The Arctic University of Norway and the University of Oulu, as well as the Finno-Ugrian Society. The data have been compiled for the purposes of doing spatial linguistic and multidisciplinary research, and to visually present the state-of-the-art knowledge of the Uralic languages and their dialects. Geographic distributions are visualized as vector data primarily by using polygon objects (speaker areas or language areas), and in some rare cases, by using points. Based on the language distributions, coordinates for the languages and their dialects (point locations) have also been defined.

Background

The work began in 2016 when Timo Rantanen started to digitalize spatial information of Uralic languages from historical sources such as printed maps and text descriptions as part of the BEDLAN project (Outi Vesakoski as PI, <https://bedlan.net/>). The cartographical digitation work was conducted using geographic information systems (GIS), which enable compilation, procession and harmonization of the data from different sources. At the beginning, the aim of the project was to get locations for the languages to be able to put them on the digital map, and to further be used with other historical data. After going through easily accessible historical material, it was obvious that collection of precise information covering the whole language family was a demanding task. The need of new information concerned both past and current distributions.

At the same time, a writing process of a new Uralic handbook *The Oxford Guide to the Uralic Languages* (OGUL) (Bakró-Nagy et al., in press) was ongoing. The handbook team had an extensive understanding of the Uralic languages. They also had a need for proper language maps, but not the possibility to create uniform state-of-the-art maps of Uralic languages themselves. Jussi Ylikoski, who was contributing in both projects, realized the potential for mutual benefits, and after

negotiations, the collaboration between the projects began offering a great opportunity to acquire the most reliable state-of-the-art knowledge of the Uralic languages.

Previously collected information of past language distributions was utilized in a query, which was created for the responsible authors of the handbook's language chapters. The first step was to visualize the different digitized distributions of a particular language on the base map as overlapping layers. Then, the authors were asked to give their opinion of the accuracy of original sources and language boundaries. Each author's task was to comment on which of the sources most precisely present the traditional distribution, and if any of the sources is inadequate how language area should be determined on the map. Present-day information was also queried from the authors in the process. As a result, plenty of new information on both past and present distributions were received. Then, new information was processed to GIS format. It was stored in the same database with previously collected information of the distributions from original sources producing a comprehensive collection of spatial language data in a consistent form.

Finally, the collected datasets were used as a basis for the language map production. Language maps were constructed in collaboration with the handbook authors, who also had the possibility to affect to the selection of base map features, such as place names, settlements and the physical environment (see a detailed explanation in Rantanen et al., in press). The final settings of overall layout was chosen by the map team. The created language maps are stored and published in *the Geographical database of the Uralic languages*. The same draft maps were also provided to *the Oxford Guide to the Uralic Languages*, where the maps will be published in 2022.

Contents

The Geographical database of the Uralic languages includes the geographical distribution of ca 48 Uralic languages in roughly two time periods: 1) at the beginning of the 20th century – indicating approximately the widest known distribution of the Uralic languages, labeled “traditional”, and 2) the current distribution, covering approximately the beginning of the 21st century up to the present day. For Mator and Kamas, which became extinct in the 19th and 20th centuries, respectively (Dolgix 1960), the “traditional distribution” refers to the beginning of the 19th century. Other exceptions for the traditional time period are Livonian (medieval distribution, ca 1000 AD) and Hungarian (indicating ethnicities in 1495).

The database consists of 1) geospatial datasets (GIS data as polygons and points) and 2) language maps (map images as png files). The geospatial datasets are further divided to the a) “original distributions” (original published source material) and b) “expert distributions” (state-of-the-art distributions modified by experts during this study). The spatial extent for the language maps originate from the expert distributions. The coordinates for the languages have been determined using expert distributions as a basis as well.

1) Geospatial datasets

The geospatial datasets (original and expert distributions) are available as shapefiles (.shp) in WGS-84 coordinate system. The datasets can be opened and processed with GIS software, such as QGIS and ArcGIS, as well as other software, which are able to handle spatial data. Each dataset consists of spatial information as vector data, either polygons (the usual data type) or points (in few cases). The attribute information tied to each polygon and point consists of the FID (feature ID), name of the language and dialect, information on the branch particular language belongs to, the time period,

original source(s), Glottocode (language ID) and ISO-639-3 (another ID). Additional information is given when necessary.

The datasets cover approximately 48 Uralic languages: South Saami, Ume Saami, Pite Saami, Lule Saami, North Saami, Aanaar Saami, Skolt Saami, Kildin Saami, Akkala Saami, Ter Saami, Finnish, Meänkieli, Kven, Karelian, Ludic, Veps, Ingrian, Votic, North Estonian, South Estonian, Livonian, Erzya, Moksha, Hill Mari, North-Western Mari, Meadow Mari, Eastern Mari, Komi-Zyrian, Komi-Permyak, Yazva Komi, Udmurt, North Mansi, East Mansi, West Mansi, South Mansi, North Khanty, East Khanty, South Khanty, Hungarian, Tundra Nenets, Forest Nenets, Tundra Enets, Forest Enets, Nganasan, Northern Selkup, Tomsk region Selkup, Kamas and Mator. In addition, the datasets include varying amount of dialectal distribution information for these languages. There are 1–8 traditional and 0–2 current distributions available for each language, compiled initially from published sources and secondarily updated and improved by experts in these languages. The total number of datasets is 226.

The database follows a hierarchical structure presenting both the individual branches of the family (e.g. Saami, Finnic, Samoyedic), the individual languages within those branches (e.g. Finnic: Finnish, Karelian, Veps), and some dialectal divisions within individual languages (e.g. Finnish: Southwestern, Tavastian, Southern Ostrobothnian, Mid- and Northern Ostrobothnian, Far-Northern, Savonian, Southeastern). The hierarchical structure of these languages takes no position on how to taxonomically position the (uncontroversial) branches within the family or individual languages within the branches they belong to.

Some research purposes benefit from single exact coordinate locations for languages and dialects. Therefore, in addition to the primary language and dialect distribution data presented as areas (polygons), one coordinate pair (X and Y coordinate) was assigned to indicate the location of each Uralic language and dialect as precisely as possible. The new expert distributions in polygon format formed a basis for the language coordinate determination. The determination of the coordinate locations is complicated, because one point seldom can present the location of a language unequivocally. Here coordinates have been defined using following rules:

- the point must be within the polygon
- in cases where the distributions have been depicted with several polygons, the expected main polygon has been chosen to overlap with a coordinate point
- in sparsely populated regions, the population clusters have a bigger value in the determination of the point location
- the points cannot be located over the big lakes or rivers

2) Language maps

The database contains 45 color maps of Uralic languages. The maps are based on the geospatial datasets of the expert distributions, and they are divided into the following categories:

- I) an overall map of the whole language families (2 maps)
- II) maps for the uncontroversial main branches of Uralic (9 maps)
- III) individual language maps (34 maps)

In some cases, past and present distributions are shown as their own layers on the map, but sometimes there are separate panels for the time periods. The amount of different languages and dialects varies remarkably between the maps, and the layout of each map has been customized independently. To achieve visually clear and easily comprehensible illustrations, the areas of overlapping languages are not shown.

Base maps (background data) consist of land area, sea, lakes, rivers, topographical information, settlements, administrative borders and place names. All the base maps are in vector format, including the datasets from Natural Earth and DIVA-GIS. Multi-Directional Hillshade used in whole language family maps is from the ESRI archives. The settlements and some environmental features were selected in collaboration with the language experts. Place names' orthography was verified by the experts.

Data release

The datasets are released under Creative Commons Attribution 4.0. The updates for the data can be done in the forthcoming version releases. The authors welcome all contributions for future versions.

Information to ease the usage of data

Folder structure

The 'Geographical database of the Uralic languages' is divided to two main folders: 'Geospatial datasets' and 'Language maps'.

'Geospatial datasets' is the main folder for all GIS data. The actual datasets have been divided to the subfolders of 'Language distributions' (the primary data) and 'Language coordinates' (the supplementary data).

'Language distributions' folder further contains subfolders of 'Original distributions' and 'Expert distributions' with a separate metadata description (Metadata_original_distributions.xlsx and Metadata_expert_distributions.xlsx).

'Original distributions' folder includes the geographical distributions collected and digitized from the original published studies. Each file covers one language from one source, and the files have been divided to the language branch subfolders ('Saami', 'Finnic', etc.). In addition, the original sources which have distributions for the whole branch (or numerous languages within one branch) are further merged to their own files. These files are in subfolders named as branch name_merge, for example 'Finnic_merge'.

'Expert distributions' folder includes the new information of Uralic languages compiled during this work in collaboration with language experts. A subfolder 'Languages' contains the most detailed information collected from individual language-wise approaches. A subfolder 'Language branches_no overlap' have the same information, but the datasets include whole branches in one file. Importantly, the data stored in this folder have been processed: spatial overlap of the separate languages and language branches have been eliminated. These datasets therefore are excellent for example for whole language family visualizations.

'Language coordinates' is a separate folder for the point data created using polygon distributions. The folder contains geographical coordinates for the Uralic languages and their dialects (one coordinate pair per language variant) as .shp, .xlsx and .csv files (file name: Uralic coordinates).

Another main folder 'Language maps' consists of Uralic language maps as .png files. The details of the maps are described in the metadata file Map_list.xlsx.

Systematic naming through database

The language names follow the orthographical rules used in *The Oxford Guide to the Uralic Languages* (Bakró-Nagy et al., in press).

Systematic naming has been used for the geospatial dataset and language map files. In geospatial datasets all the names have been created using the following logic: name of the language_time period_source/author.shp

- > SouthSaami_traditional_Sammallahti.shp (original distributions)
- > Udmurt_current_OGUL.shp (expert distributions)

In the language maps all the names have been based on the following logic: number code space name of the language/branch/family_time period (when needed).png

- > 1 Saami branch.png
- > 1.1 South Saami.png
- > 9.1a Nenets_traditional.png
- > 9.1b Nenets_current.png

Attribute table column names in geospatial datasets

The attribute tables have the same structure including for example same columns in the same order uniformly. The only exception is *Uralic coordinates* file, which is separately described below.

Attribute table column titles in the datasets of language distributions:

FID = feature ID for each object, automatically generated by the software

Shape = used geometry type, most usually polygon, in some cases point

Language = name of the language

Dialect = name of the dialect

Otherinfo = additional information, can contain more detailed definition for the time period or dialectal division, e.g. alternative name for the dialect

Branch = name of the branch a particular language belongs to

Timeperiod = overall time period for the distribution, usually 'traditional' or 'current'

Sources = Used sources for the distribution, can be published source or new information while it is been labeled as p.c. = personal communication

Glottocode = language ID produced by the Glottolog

ISO_639_3 = language ID produced by the International Organization for Standardization

Attribute table column titles in the dataset of language coordinates (= *Uralic coordinates*):

FID = feature ID for each object, automatically generated by the software

Shape = used geometry type, in this case point

LVariant = name of the language variant or dialect

Language = name of the language

Branch = name of the branch a particular language belongs to

Latitude = Y coordinate

Longitude = X coordinate

Otherinfo = additional information, an indication when an entire coordinate originates from Glottolog

Glottocode = language ID produced by the Glottolog

ISO_639_3 = language ID produced by the International Organization for Standardization

Personal communications

The significant impact of expert(s) have been indicated with personal communication (p.c.) label in the sources. Personal communications were used when creating state-of-the-art distributions for the languages. In these cases, the received information is often unpublished, and it is implemented to the new language distributions in different ways. Following experts have given a personal communication:

Ante Aikio: North Saami

Svetlana Burkova: Forest Nenets and Tundra Nenets

Ulla-Maija Forsberg: East Mansi, West and South Mansi

Riho Grünthal: Karelian, Ludic

Markus Juutinen: Skolt Saami

Olga Kazakevich: Northern Selkup and Tomsk region Selkup

Maja Lisa Kappfjell: South Saami

Elena Markus: Ingrian, Votic

Sergej Maksimov: Udmurt

Tamara Merova: North Mansi, East Mansi, West and South Mansi

Tat'jana Merova: North Mansi, East Mansi, West and South Mansi

Peeter Päll: South Estonian

Eszter Ruttkay-Miklián: North Khanty

Anneli Sarhimaa: Karelian, Ludic

Zsófia Schön: East Khanty, South Khanty

Florian Siegl: Forest Enets and Tundra Enets

Mária Sipos: North Khanty, South Khanty

Elena Skribnik: North Mansi, East Mansi, West and South Mansi

Taarna Valtonen: Aanaar Saami, Skolt Saami

Jussi Ylikoski: South Saami, Lule Saami, North Saami

Acknowledgments

We are grateful for the collaboration with the editors and corresponding authors of the *Oxford Guide to the Uralic Languages* including Marianne Bakró-Nagy, Elena Skribnik, Johanna Laakso, Ante Aikio, Svetlana Burkova, Svetlana Edygarova, Ulla-Maija Forsberg, Riho Grünthal, Arja Hamari, Markus Juutinen, Olga Kazakevich, István Kenesei, Gerson Klumpp, Eino Koponen, Nikolay Kuznetsov, Elena Markus, Matti Miestamo, Karl Pajusalu, Michael Rießler, Sirkka Saarinen, Anneli Sarhimaa, Zsófia Schön, Florian Siegl, Mária Sipos, Taarna Valtonen and Beáta Wagner-Nagy. The authors of language descriptions were assisted by Enn Ernits, Maja Lisa Kappfjell, Vesa Koivisto, Sergej Maksimov, Tamara Merova, Tat'jana Merova, Mehmet Muslimov, Helena Omma, Marjatta Palander, Peeter Päll, Helka Riionheimo, Eszter Ruttkay-Miklián and Håkan Rydving to whom we show our gratitude – simultaneously we apologize for not thanking those assisting persons who were not named in the correspondence. We also thank other researchers, who have contributed to the development of the database at its various stages including Rogier Blokland, Michael Dunn, Hannah Haynie, Mervi de Heer, Terhi Honkola, Lotta Jalava, Sofia Koskela, Dmitry Kuznetsov, Unni-Päivä Leino, Luke Maurits, Miina Norvik, Tua Nylén, Niko Partanen, Meeli Roose, Tapani Salminen, Jenni Santaharju, Kaj Syrjänen and Juho Vehviläinen.

References

- Abondolo, D. (ed.). (1998). *The Uralic languages*. London: Routledge.
- Aikio, S. (2017). *Davvisámi báikenamat*. Kárášjohka: ČálliidLágádus.
- Barancev = Баранцев, А. П. (1978). *Образцы лудиковской речи. Образцы корпуса лудиковского идиолекта I. Деревня Пелдожа и ее окрестности*. Петрозаводск: Карелия.
- Bartens, R. (1999). *Mordvalaiskielten rakenne ja kehitys* (MSFOu 232). Helsinki: Suomalais-Ugrilainen Seura.
- Boiko, K. (1998). 'Liivlased ja liivi keel'. In J. Õispuu (ed.), *Kaheksa keelt, kaheksa rahvast*. Tallinn: Tallinna Pedagoogikaülikool. 5–13.
- Brykina, M. & V. Gusev. (2015). 'Temperature terms in Nganasan'. In M. Koptjevskaja-Tamm (ed.), *The linguistics of temperature*. Amsterdam: John Benjamins Publishing Company. 537–569.
- Burkova = Буркова, С. И. (2016a). 'Ненецкий лесной язык'. In В. Ю. Михальченко (ed.), *Язык и общество. Энциклопедия*. Москва: Азбуковник. 313–319.
- Burkova = Буркова, С. И. (2016b). 'Ненецкий тундровый язык'. In В. Ю. Михальченко (ed.), *Язык и общество. Энциклопедия*. Москва: Азбуковник. 319–328.
- Castrén, M. A. (1856). *Nordische Reisen und Forschungen 2. Reiseberichte und Briefe aus den Jahren 1845–1849*. St. Petersburg: Buchdruckerei der Kaiserlichen Akademie der Wissenschaften.
- Csepregi = Чепреги, М. (2017). *Сургутский диалект хантыйского языка*. Ханты-Мансийск: ОУИПИИР.
- Dolgix = Долгих, Б. О. (1960). *Родовой и племенной состав народов Сибири в XVII веке*. Москва: Издательство АН СССР.
- Dolgix & Fajnbere = Долгих, Б. О. & Л. А. Файнберг. (1960). 'Таймырские нганасаны'. In Б. О. Долгих, (ed.), *Современное хозяйство, культура и быт малых народов Севера*. Москва: Издательство АН СССР. 9–62.
- Dolgix = Долгих, Б. О. (1963). 'Происхождение долган'. In Б. О. Долгих, (ed.), *Сибирский этнографический сборник*. Москва: Издательство АН СССР. 9–62.
- Dunfjeld, M. (2006). *Tjaalehtjimmie. Form og innhold i sørsamisk ornamentikk*. Snåsa: Saemien Sijte.
- Eesti murrete sõnaraamat = Haak, A, E. Juhkam, M. Must, M. Mäger, H. Neetar, S. Nigol, E. Niit, V. Oja, V. Pall, E. Ross, A. Univere & H. Viies. (1994). *Eesti murrete sõnaraamat. I. A–J*. Tallinn: Eesti Keele Instituut.
- Eesti Rahva Museum (ERM). (2009). Soome-ugri ja samojeedi rahvaste asualad. Map collected by Indrek Jääts. Available in <<https://blog.erm.ee/?p=4245>>
- Ezykin, L. (2015). Wikipedia maps on Saami languages made by Ljuboslav Ezykin. Available in: <https://en.wikipedia.org/wiki/Akkala_Sami#/media/File:Sami_dialects_and_settlements_in_Russia_map.svg>
- Ermuškin = Ермушкин, Г. И. (1984). *Ареальные исследования по восточным финно-угорским языкам (эрзя-мордовский язык)*. Москва: Наука.
- Feoktistov, A. P. (1990). 'Die Dialekte der mordwinischen Sprachen / Диалекты мордовских языков'. In Kaino Heikkilä (ed.), *H. Paasonens mordwinisches Wörterbuch* (LSFU XXIII, 1). Helsinki: Finno-Ugrian Society. XXXI–LXXXVI.
- Feoktistov, A. & S. Saarinen. (2005). *Mokšamordvan murteet* (MSFOu 249). Helsinki: Suomalais-Ugrilainen Seura.
- Fil'čenko & Potanina = Фильченко, А. & О. Потанина. (2012). 'Хантыйские тексты'. In А. Фильченко, О. Потанина, Е. Крюкова, В. Гусев, А. Байдак, Н. Максимова, В. Лемская & А. Ким, (eds.), *Сборник аннотированных фольклорных и бытовых текстов обско-енисейского языкового ареала*. Томск: Вайар. 67–190.

- Grünthal, R. & T. Salminen, (eds.) (1993). *Geographical distribution of the Uralic Languages*. Helsinki: Finno-Ugrian Society.
- Grünthal, R. & A. Sarhima (eds.). (2004/2012). *Itämerensuomalaiset kielet ja niiden päämurteet*. Helsinki: Suomalais-Ugrilainen Seura.
- Grünthal, R. (2011). 'Population decline and the erosion of the Veps language community'. In R. Grünthal & M. Kovács (eds.), *The ethnic and linguistic context of identity: Finno-Ugric minorities*. Uralica Helsingiensia 5. Helsinki: Suomalais-Ugrilainen Seura. 267–293.
- Haarmann, H. (1974). *Die finnisch-ugrischen Sprachen. Soziologische und politische Aspekte ihrer Entwicklung*. Hamburg: Buske.
- Hammarström, H., Forkel, R., Haspelmath, M. & Bank, S. (2021). *Glottolog 4.4*. Leipzig: Max Planck Institute for Evolutionary Anthropology. <https://doi.org/10.5281/zenodo.4761960>
- Hasselbrink, G. (1981–1985). *Südlappisches Wörterbuch*. Uppsala: A.-B. Lundequistska bokhandeln.
- Helimski, E. (1997). *Die matorische Sprache. Wörterverzeichnis – Grundzüge der Grammatik – Sprachgeschichte*. Unter Mitarbeit von Beáta Nagy. SUA 41. Szeged: JATE Finnougor tanszék.
- Helimski, E. (1999). 'Einige Angaben zur Toponymie und Ethnonymie des sajanamojedischen Siedlungsgebietes aus den Jahren 1739–1740 bei G. F. Müller'. In C. Hasselblatt, & P. Jääsalmi-Krüger (eds.), *Europa et Sibiria – Beiträge zu Sprache und Kultur der kleineren finnougriken, samojedischen und paläosibirischen Völker*. Gedenkband für Wolfgang Veenker (VSUA 51). Wiesbaden: Harrassowitz. 181–189.
- Itkonen, T. I. (1948). *Suomen lappalaiset vuoteen 1945. II*. Porvoo: WSOY.
- Ivanov = Иванов, И. Г. (1981). *Марий диалектологий*. Йошкар-Ола: Марий государственный университет.
- Joki, A. J. (1944). *Kai Donners Kamassisches Wörterbuch nebst Sprachproben und Hauptzügen der Grammatik*. Lexica Societatis Fenno-Ugricae VIII. Helsinki: Suomalais-Ugrilainen Seura.
- Kannisto, A. (1951). *Wogulische Volksdichtung I* (MSFOu 101). Helsinki: Suomalais-Ugrilainen Seura.
- Katz, H. (1980). 'Zum Namen der Kamassen', *Acta Linguistica Hungarica* 30: 239–246.
- Kazakevich, O. (2005). 'Local dialects of the Northern Selkups: the present-day situation'. In *Congressus Decimus Internationalis Fenno-Ugristarum. Yoshkar-Ola 15.08.-21.08.2005. Pars II. Summaria acroasium in sectionibus. Linguistica*. Yoshkar-Ola. 199–201.
- Kiss, J. (ed.) (2001). *Magyar dialektológia*. Budapest: Osiris.
- Klement'ev & Šlygina = Клементьев, Е. И. & Н. В. Шлыгина (eds.). (2003). *Прибалтийско-финские народы России*. Москва: Наука.
- Kocsis, K., P. Tátraí, N. Agárdi, D. Balizs, A. Bognár, Z. Bottlik, A. Kovács, & Á. E. Varga (2015). *Changing ethnic patterns of the Carpatho-Pannonian area*. Third edition. Budapest: MTA research Centre for Astronomy and Earth Sciences, Geographical Institute.
- Komi jazyk = Федюнева, Г. В. (ed.) (1998). *Кomi язык: энциклопедия*. Москва: Издательство ДИК.
- Koryakov, Y. B. (2016). Unpublished map drawn by Yuri Koryakov in 2016.
- Kuznetsova, N., E. Markus & M. Muslimov. (2015). 'Finnic minorities of Ingria: The current sociolinguistic situation and its background'. In H. F. Marten, M. Rießler, J. Saarikivi & R. Toivanen (eds.), *Cultural and linguistic minorities in the Russian Federation and in the European Union: Comparative studies on equality and diversity*. Heidelberg: Springer. 127–167.
- Köppen, P. (1867). *Erklärender Text zu der ethnographischen Karte des St. Petersburger Gouvernements*. St. Petersburg: K. Akademie der Wissenschaften.
- Laanest = Лаанест, А. (1966). *Ижорские диалекты. Лингвогеографическое исследование*. Таллин: Академия Наук Эстонской ССР.

- Lehtiranta, J. (1992). *Arjeploginsaamen äänne- ja taivutusopin pääpiirteet* (MSFOu 212). Helsinki: Société Finno-Ougrienne.
- Levina = Левина, М. З. (2014). Мокшень диалектологиясь. Диалектология мокшанского языка: учебное пособие. Саранск: Издательство Мордовского университета.
- Linkola, A. & M. Linkola. (2000). *Kolttsaamelaiset – The Skolt Sámi*. Kuopio: Inari Sámi Museum.
- Lytkin = Лыткин, В. И. (1962). *Коми-пермяцкий язык*. Введение, фонетика, лексика и морфология. Кудымкар: Коми-пермяцкое книжное издательство.
- Lytkin et al. = Лыткин, В. И., К. Е. Майтинская & К. Редеев (eds.). (1976). *Основы финно-угорского языкознания. Марийский, пермские и угорские языки*. Москва: Наука.
- Maksimov, S. (2001). 'Pohjoisudmurtin murteiden ja komin kielen areaalisia leksikaalisia yhtäläisyyksiä', *Journal de la Société Finno-Ougrienne* 89: 167–184.
- Maksimov = Максимов, С. (2009). *Диалектологический атлас Удмуртского языка: карты и комментарии*. Вып. 1. Ижевск: Регулярная и хаотическая динамика.
- Mets, M., A. Haak, I. Triin, G. Juhkason, M. Kalmus, M. Norvik, K. Pajusalu, P. Teras, T. Tuisk & L. Vaba. (2014). *Lõunaesti keelesaarte tekstid*. Tallinn: Eesti Keele Sihtasutus.
- Mustonen, T. & K. Mustonen (eds.). (2011). *Eastern Sámi atlas*. Lehtoi: Snowchange.
- Nickul, K. (1952). *Saamelaisalueen saamelainen väestö 1.1.1949*. Saamelaisasiain komitean mietintö. Komiteanmietintöjä 1952, 12. Helsinki: Valtioneuvosto.
- Nissilä, V. (1967). *Die Dorfnamen des alten lüdischen Gebietes* (MSFOu 144). Helsinki: Suomalais-Ugrilainen Seura.
- Omma, H. (2019). 'Guovtti sámegiela gaskkas? Gielladutkiid ja hubmiid iežaset oainnut Jiellevári ja Unna tjerutja suopmaniidda', *Sámi dieddalaš áigečála* 1/2019: 7–29.
- Pahomov, M. (2017). *Lyyläiskysymys: Kansa vai heimo, kieli vai murre?* Helsinki: Helsingin yliopisto & Lyyläinen Seura.
- Pajusalu, K., T. Hennoste, E. Niit, P. Päll & J. Viikberg. (2020). *Eesti murded ja kohanimed*. Kolmas, kohendatud ja täiendatud trükk. Tartu: EKSA.
- Palander, M., H. Riionheimo & V. Koivisto (eds.). (2018). *On the border of language and dialect*. *Studia Fennica Linguistica* 21. Helsinki: Finnish Literature Society.
- Pennanen, J. & K. Näkkäläjärvi. (2000). *Siiddastallan – Siidoista kyliin: luontosidonnainen saamelaiskulttuuri ja sen muuttuminen*. Inarin saamelaismuseon julkaisuja n:o 3. Saamelaismuseosäätö. Oulu: Kustannus Pohjoinen.
- Popov = Попов, А. А. (1948). *Нганасаны: материальная культура*. Вып. 1. Москва: Издательство АН СССР.
- Popov = Попов, Н. А. (2011). По пяти рекам Авама (путевые заметки.) Дудинка.
- Rydving, H. (2008). 'Sørsamisk'. In A. Dalen, J.R. Hagland, S. Hårstad, H. Rydving & O. Stemshaug: *Trøndersk språkhistorie: språkforhold i ein region*. Det Kongelige Norske Videnskabers Selskab, Skrifter 3. Trondheim; Tapir akademisk forlag. 357–397.
- Rydving, H. (2016). 'Sydsamisk eller umesamisk? "Södra Tärna" i det samiska språklandskapet', *Fenno-Ugrica Suecana Nova Series* 15: 159–173.
- Salminen, T. (2017a). Personal communication on the traditional Nenets language area 1.11.2017.
- Salminen, T. (2017b). Personal communication on the current Nenets language area 1.11.2017.
- Sammallahti, P. (1998). *The Saami languages. An introduction*. Kárášjohka: Davvi Girji.
- Sarhimaa, A. (2017). *Vaietut ja vaiennetut: karjalankieliset karjalaiset Suomessa*. Helsinki: Suomalaisen Kirjallisuuden Seura.
- Sepeev = Сепеев, Г. А. (1975). *Восточные марийцы*. Йошкар-Ола: Марийское книжное издательство.

- Siegl, F. (2013). *Materials on Forest Enets, an indigenous language of Northern Siberia* (MSFOu 267). Helsinki: Suomalais-Ugrilainen Seura.
- Soviet Census 1926 = *Всесоюзная перепись населения 17 декабря 1926 г. Краткие сводки*. Москва: ЦСУ Союза ССР, 1927–1929.
- Тепляшина & Лыткин = Тепляшина, Т. И. & В. И. Лыткин (1976). 'Пермские языки'. In В. И. Лыткин, К. Е. Майтинская & К. Реден (eds.), *Основы финно-угорского языкознания. Марийский, пермские и угорские языки*. Москва: Наука.
- Tönnisson, Evald (1970). 'Muistsed liivlased ja Kuramaa liivlased', *Keel ja Kirjandus* 13: 461–466.
- Trefurt, F. L. (1785). 'Fortgesetzte Nachricht von den Tschuden'. In F. K. Gadbusch (ed.), *Versuche in der livländischen Geschichtskunde und Rechtsgelehrsamkeit*. II Band, 2. Stück. Riga: Johann Friedrich Hartknoch. 89–122.
- Tučkova et al. = Тучкова, Н. А., А. И. Кузнецова, О. А. Казакевич, А. А. Ким-Мелони, С. В. Глушков & А. В. Байдак. (2004). *Мифология селькупов. Энциклопедия уральских мифологий. Том 4*. Томск: Издательство Томского университета.
- Tugarinov = Тугаринов, А. Я. (1926). 'Последние калмажи', *Северная Азия* 1: 73–88.
- Tunkelo, E. A. & E. Leskinen. (1946). *Vepsäläisten asuma-alueiden kartta*. Helsinki: Suomalaisen Kirjallisuuden Seura.
- Vadja keele sõnaraamat* = Grünberg, S. (ed.). (2013). *Vadja keele sõnaraamat*. Tallinn: Eesti Keele Sihtasutus.
- Verbov = Вербов, Г. Д. (1973). 'Диалект лесных ненцев'. In *Самодийский сборник*. Новосибирск. 3–190.
- Virtaranta, P. (1967). *Lähisukukielten lukemisto*. Helsinki: Suomalaisen Kirjallisuuden Seura.
- Volžanina = Волжанина, Е. А. (2007). 'Лесные ненцы: расселение и динамика численности в XX веке, современная демографическая ситуация', *Археология, этнография и антропология Евразии* 2: 143–154.
- Wagner-Nagy, B. (2018). *A grammar of Nganasan*. Leiden: Brill.
- Wilbur, J. (2012). *A grammar of Pite Saami*. Studies in Diversity Linguistics 5. Berlin: Language Science Press.
- Zen'ko-Nemčinova = Зенько-Немчинова, М. А. (2006). *Сибирские лесные ненцы: историко-этнографические очерки*. Екатеринбург: Баско.
- Ziker, J. P. (2002). *Peoples of the Tundra – Northern Siberians in the Post-Communist transition*. Illinois: Waveland Press.
- Zsirai, M. (1937). *Finnugor rokonságunk*. Budapest: Magyar Tudományos Akadémia.