



**CHAIRMANSHIP OF CROATIA**  
Council of Europe  
May - November 2018

**PRÉSIDENTE DE LA CROATIE**  
Conseil de l'Europe  
Mai - Novembre 2018

COUNCIL OF EUROPE



CONSEIL DE L'EUROPE

# Minorities and Minority Languages in a Changing Europe

Conference on the occasion of the 20<sup>th</sup> anniversary of the Framework  
Convention for the Protection of National Minorities and the European  
Charter for Regional or Minority Languages

Council of Europe, Strasbourg, France  
Palais, room 1

18 – 19 June 2018

## **Panel Session: Minority Language 2.0 – Digitalisation and other new Developments**

**Olga Kazakevich, Research Computing Centre, Lomonosov Moscow State University**

### **Endangered minority languages of Russia in the internet**

In Russia two groups of autochthonous minorities are distinguished: “large” minorities counting over 50 thousand people (up to several million), many of them have autonomies (autonomous republics) of their own, where their languages have the official status of national languages (alongside with Russian), and “small” minorities, ethnic groups counting less than 50 thousand people. Minorities belonging to the second group have the official status of autochthonous lesser numbered peoples of Russia, and their languages are respectively called languages of autochthonous lesser numbered peoples, or lesser used minority languages. The fact is that practically in all lesser numbered minority ethnic groups ethnic language speakers are less than the number of people in the group. As far as lesser used minority languages concerns, the main problem is language loss. Even in ethno-local groups of lesser numbered minorities in the Caucuses where ethnic languages were quite strong up to recently signs of the beginning of language shift can be traced. In Siberia, the Far East and the North of European Russia language shift is at hand in all lesser numbered minority groups and in the last three decades it is developing at a great speed.

In the sociolinguistic encyclopedia “Language and Society” published recently by the Institute of Linguistic, Russian Academy of Sciences (2016), a description of the functioning of 53 lesser used minority languages is given. The authors of these descriptions are all field linguists working with the languages and knowing the situation in local communities. Among 53 “super-minority” languages 14 have no more than 20 speakers, all over 60 – 80 years old (e.g. Aleut in Russia is spoken by 4 people, Votic is spoken by less than 10 people, all well over 80, Oroque (Uilta) is spoken by 5 people etc.). And we are loosing languages. In this millennium three languages have been lost: Babino (Akkala) Saami (Murmansk region), Kerek (Chukotka) and Oroch (Amur region).

So, in our epoch of digitalization what can be done by those who still speak these dramatically endangered minority languages? The last speakers of these languages are often conscious of the heritage they keep, and some of them try to write down what they know, to or/and to share their knowledge with linguists. Quite often when we, linguists, come to communities where only the elder generation speaks the ethnic language we appear to be those who came to accept the community heritage and to keep it for the future generations hoping that they would wish to accept it from us. And these elderly people, preservers of what is left of their cultural and linguistic heritage, pass it to us saying: “All right, record all I tell you. My children are not interested in my stories, they are too busy, my grandchildren think all that is old fashioned. Who knows, maybe, my great grandchildren will turn to all that, one day they would like to know how their ancestors used to live”. Thanks the technical progress, now we have proper equipment for language documentation: audio recorders of high quality, portative video cameras, and special software to analyze the data and to archive them. Today the role of a linguist working in a communities where ethnic language is kept by a handful of speakers or even by a single speaker becomes vital, as s/he is responsible not only for the quality of data collection, procession and preservation and for its presentation to academic community, but also for the return of the data and the results of its

processing to the community members and their descendants. It should be stated that linguistic work in a community elevates the prestige of the community (or heritage) language.

Data in lesser used minority languages posted on the web make a miracle awakening interest in heritage language among young people and teenagers. A grandson would never listen to his grandmother's confusing chat at home, but her stories placed on the web become interesting for him.

Endangered minority languages are present in the web either on web-sites, or in social nets.

The majority of web-sites containing data on these languages are created by linguists and address the academic community, primarily linguists. They contain audio (possibly also video) and graphic representation of texts, mostly morphologically annotated, and/or sounding dictionaries. As a rule, it is problematic for far-away linguistic community members to use the data from these sites, as the software used for data representation demands special programs on the user's computers and stable internet connection, and that's what often lacks in the Siberian villages.

I found only five web resources with the materials in lesser used minority languages of Russia addressing speakers of these languages and all interested users (including linguists). All of them are multimedia. Four of them contain data in Evenki dialects. Two sites are created in Yakutsk: "Evengus" (<http://www.evengus.ru/>) with texts and training materials allowing to learn Evenki, exists since 2007, and "Evenkiteka" (<http://www.evenkiteka.ru/>), an on-line library with 300 books in Evenki, describing Evenki, teaching Evenki and on the Evenki history and culture (since 2011). Two others resources are being developed on the basis of Amur State University in Blagoveshshensk: Save Language (<https://savelanguage.amursu.ru/>) containing texts and sounding dictionaries in the dialects of Amur region, China and Mongolia (since 2016) and a corpus of Evenki dialects of Amur region and adjacent territories (<https://linguacorporus.amursu.ru/base>) containing texts in Evenki dialects in audio, video and graphic representation.

The fifth site was created on the basis of Laboratory for Computational Lexicography, Research Computing Centre, Lomonosov Moscow State University in 2012 and is being developed since then. The site developers keep two target groups in focus: on the one hand, they address the academic community – linguists, anthropologists, historians, on the other hand, they address the speakers of the languages presented on the site, those who teach or learn these languages at schools and colleges, who want to learn them or just to get acquainted with languages and cultures of Siberian peoples.

The site developers use the latest relevant hardware and software, which enables cross-browser compatibility and cross-platform operation of the site. The site is to be displayed properly on different devices including net books, tablets and other mobile devices.

The primarily informational source for the site is the Multimedia Computer Archive of materials in local dialects of Selkup, Ket and Evenki, recorded in the course of linguistic expeditions regularly organized and led within the laboratory, but we also draw relevant information from published works and other archives as well. For the moment the site contains data in Selkup, Ket and Evenki, but data in other Siberian languages are supposed to be added.

On the site the following types of data can be found:

1) sociolinguistic data characterizing the functioning and degree of preservation of Selkup, Ket, or Evenki in the surveyed villages;

- 2) sounding thematic dictionaries of Selkup, Ket and Evenki local dialects recorded in the field within the last decade and a half and loaded into a database, which is to be developed for that purpose;
- 3) annotated Selkup, Ket and Evenki text corpora containing texts of various genres in local dialects of these languages both recorded in the course of our expeditions of the last decade and a half and extracted from archives (first of all, from the archive of Peter the Great Museum of Anthropology and Ethnography, Russian Academy of Sciences, Saint Petersburg);
- 4) Russian texts recorded in the field from the Russian population of the area living side by side with the Selkups, Kets and Evenkis and telling about life, traditional ways and rituals of their neighbours;
- 5) information on the grammar structure of the three languages in the form of grammar reference book or reference database;
- 6) Selkup, Ket and Evenki language learning software (training games);
- 7) documentaries showing the functioning of the languages and the life of their speakers;
- 8) photos representing the life of the villages and their residents, informants and the nature surroundings.

According to the Google statistics, the site is most frequently visited by users from Siberia (Krasnoyarsk Territory, Yamalo-Nenets autonomous area, Tomsk region) – places of residents of the Selkups, Kets and Evenkis.

The presence of minority languages on one of the most popular in Russia social web networks – Vkontakte (<http://vk.com>) – was investigated in the National Research University Higher School of Economics in Moscow (Orekhov et Al. 2016; Zaydelman 2016). In the whole the presence of 43 minority languages was traced on the network automatically using lexical markers. For 30 minority languages at least one Vkontakte community was found. In the whole 1633 communities containing at least one text in a minority language were considered. The majority of the traced languages were “large” minority languages, such as Tatar, Yakut, Udmurt etc., but there were also several lesser used, endangered minority languages among them – Chukchi, Evenki, Tundra Nenets, Nanai, Nivkh, South Yukagir. The texts in the communities are not only in the ethnic languages, but also in Russian, and for most of the languages Russian texts considerably outnumber texts in minority languages. Among lesser used minority languages only for Tundra Nenets communities texts in Tundra Nenets outnumber Russian texts. Finally, texts in minority languages are mostly pretty small – their average length is about five words. Geographically the members of the web communities using endangered minority languages are often located in big cities, and for Nivkh and South Yukagir the majority of the web community members are located in St. Petersburg.

Thus, we can see the appearance of lesser used endangered minority languages of Russia in the internet. So, a new communication sphere arose for them, and it seems, they do have, or at least can have future in this sphere.

#### References

- Orekhov B., Krylova I., Popov I., Stepanova E., Zaydelman L. 2016. Russian minority languages on the web: descriptive statistics. In *Computational Linguistics and Intellectual Technologies: Proceedings of the International Conference “Dialogue 2016”*. Moscow.
- Zaydelman L. 2016. Modeling speech behavior of native speakers of minority languages of the Russian Federation in a social network. Magister thesis.