Audio Collections of Endangered Arctic Languages in the Russian Federation

Marina Lublinskaya[1] and Tatiana Sherstinova[2]

[1] Institute of Linguistic Research, Russian Academy of Science
Tuchkov per. 9, St. Petersburg, 199053, Russia
lublin@ling.ras.spb.ru,
[2] St. Petersburg State University
Department of Phonetics
Universitetskaya nab. 11, 199034, St. Petersburg, Russia
tanya@ts4306.spb.edu

Abstract. In the Russian Federation 63 minority languages are mentioned in the "Red Book of the Languages of Russia", what means that they are practically dying out. Because of that it is highly important to make and preserve original recordings of these languages and prepare their documentation. Arctic peoples of Russia are demographically small and the number of speakers using them is decreasing dramatically. The paper describes three projects related to two Northern Languages – Nenets and Nganasan: Nenets Audio Dictionary, Nganasan Audio Dictionary and Russian-Nenets Online Multimedia Phrase-book.

1 Introduction. Maintaining endangered languages - the importance of the problem

The dramatic process of social degradation and the dying out of a quantity of minority languages is a serious problem of the world community. This problem "demands an urgent solution as in 10-20 years there may be nothing left to save" [2]. In the Russian Federation there is more than 200 languages spoken by the native population, and "there is no need to prove the thesis that a significant number of languages on the territory of Russia are now under threat of total extinction" [2]. Thus in Russia 63 minority languages are mentioned in the "Red Book of the Languages of Russia" (1994), what means that they are practically dying out [10].

In order to preserve the information about these languages for the future generations and to maintain the knowledge of these languages by the young people, it is necessary "to document as much material on these languages as possible and to use this for language descriptions, grammars, compilation of dictionaries and edited collections of oral and written literature" [3]. It is very important to make and preserve original recordings of these languages, and the less the number of native speakers of some language, the more necessary it is to record and describe it as quickly as possible [7].

It is also important to develop teaching methods, in particular for younger members of certain ethnic groups, who have no sufficient knowledge of their native language and to make them aware of the value of their heritage. In those cases where language revival is impossible, it is very important for linguists and ethnologists to organize a collection of acoustic data and documents on endangered languages [3].

Arctic peoples of Russia are demographically small and the number of speakers using them is decreasing dramatically. The paper describes three projects related with two Northern Languages – Nenets and Nganasan.

2 Concise description of Nenets and Nganasan

Nenets is the small people living along Arctic Ocean between Kola Peninsula and the mouth of river Yenisey. In spite of huge territory, original population counted to 35000 persons in 1989. Nenets is the native language for 77,1% of them. Comparing to other small Northern languages of Russia the Nenets one is rather safe. In spite of its comparative safeness no general description of Nenets phonetics, prosody and grammar have been made up to present.

Nganasans is the northern population in Eurasia, inhabiting the Taymyr Peninsula in Siberia. It seems the size of this population never exceeded 1000-1,500 persons. According to the date of the last All-Union Census (1989), the number of Nganasans was 1278 persons. The majority lives in three villages - Volochanka, Ustj-Avam and Novaya. The administrative center is the town of Dudinka. According to recent field data, only ca. 50% of the entire population (in younger generations – not more than 10-15%) continues to use Nganasan [10]. It is mostly Russian and Dolgan languages that take place of the native language. Nowadays only about one hundred ethnic Nganasans continue to lead a semi-nomadic life of hunters and fishermen, preserving the native language in their everyday life as well.

Both Nenets and Nganasan languages belongs (together with Enets, which is now practically dead) to the Northern-Samoyedic subgroup of the Samoyedic Group of the Uralic language family.

The main danger to both these languages hides in Russian bilingualism of the most of the native speakers. “Pure” Nenets and Nganasan speakers are mainly nomadic (reindeer herdsmen). The parents usually want their children to have a better life conditions in the future, and try to send them studying to towns. In towns national speech ability is being gradually replaced by the Russian practice. Young generation prefers to use Russian as a language of communication in most of cases.

Moreover, as reported in [13], “there are obvious danger signs in both the physical and the national and cultural existence of the Nenets people. Government industrial and military officials determine which settlement areas, positions, rights and privileges are available to them. Nuclear experiments have been carried out unobstructed in Novaya Zemlya. The Norilsknikel concern alone has polluted 5 million hectares of Nenets grazing-lands and almost l million hectares of forests. The pollution of heavy metals has been transferred to the humans through mosses and reindeer meat. The average life expectancy of Nenets is 45-50 years, the suicide rate is unusually high. Only 41% of Nenets are employed and usually in jobs which require low qualification". Practically, the same may be said about the Nganasans.

Both Nenets and Nganasan languages are registered in the UNESCO Red Book on Endangered Languages and in the Red Book of the Peoples of the Russian Federation [11, 10].

3 How new information technologies may help to save and popularize the endangered languages

Modern information techniques, especially speech databases, represent convenient tools to preserve acoustic data of different kind. Besides, they allow to combine in a single application both written information about the language and its actual recordings. In addition to reading a written text, a scientist can now listen to its recording at the same moment. That makes linguistic databases and similar computer systems the proper instrument for preserving and studying small and disappearing languages.

The impetuous development of the Internet extends applications of speech databases. Until recently they could be accessible only in large archives or libraries. Nowadays there are many examples of representative speech databases in the Web. We consider both speech and Internet technologies to be the basis for actual maintenance of the endangered languages, as the publication of any linguistic database in the Web will allow access to its textual and audio data to all users of the Internet.

There was a special program proposed and started by the Department of Phonetics of St. Petersburg State University, which aims in creation of Phonetic Funds and Multimedia Dictionaries for the minority languages of Russia. The Department of Phonetics has been always characterized by its interest in minority languages of Russia: here not only the phonetics of the languages of the peoples of the North and the Far East are studied, but earlier in last century also the necessary recommendations for creating written languages were given [3].

In addition, there are several joint projects of the Department of Phonetics of St. Petersburg State University, Institute of Linguistic Research of the Russian Academy of Science, and the Department of Linguistics, Groningen University, The Netherland – "Writing and Teaching of Samoyedic" (phonetic project, related to the peoples and languages of Russian North) and "Nenets Speech Varieties on the Internet" (Web-publication of the Russian-Nenets Phrasebook).

First of all it was considered important to inventory and study the existing recordings, which are currently preserved in collections of state archives. The work started with making a digital catalogue for collections of the Sound Archive of the Institute for Russian Literature of the Russian Academy of Sciences – the Pushkinsky Dom, which is the richest audio collection of speech and folklore of the peoples of the Russian Federation. The finished part of this catalogue is available in the Web: `http://www.speech.nw.ru/phonetics/homepage.html.` The fragments of the spoken language, folksongs and fairy tales preserved in the Phonogrammarchive may be also found on this site. As the quality of the old archival recordings usually is far from perfect, many of the archival collections deserve serious reconstruction, decoding and digitization.

In order to make new recordings of the minority languages it is necessary to organize scientific expeditions, which imply field recordings with handheld good sound recording equipment. The other way is to invite native speakers of the language to the main research centers in St. Petersburg or Moscow and to made the professional recordings using studio facilities. Fieldwork expeditions and studio recordings allowed us to collect substantial sound data in Nenets and Nganasan.

4 Nenets Audio Dictionary

Nenets Audio Dictionary was created in St. Petersburg Institute of Linguistic Research of the Russian Academy of Science in cooperation with the Department of Phonetics of St. Petersburg State University. Its main aim is to preserve specialties of Nenets phonetics.

The Audio Dictionary structure is based on the Nenets-Russian Dictionary composed by N.M.Tereshchenko, where from it borrows the completeness of translation, text examples, linguistic remarks and comments. It was elaborated in the form of the database created by means of DBMS Access 97/2000 utilities. Audio data are stored in the database in the raw-format, which allows their processing and analysis in Windows-compatible applications. The number of words in the core database is limited to 4000 (as presented in the School Nenets Dictionary for Tundra Nenets dialect composed by N.M.Tereshchenko). Variants of pronunciation are given for word likewise in Yurak-Samojedisches Worterbuch, composed by T.Lehtisalo [5]. Along with the recording of standard word pronunciation, another variants of pronunciation in subdialects of Nenets Tundra dialect may be presented. Supplementary phonetic information concerns word accent and number of syllables.

Native Nenets speakers from different region – Kanin tundra, Bolshaya Zemlya, Yamal Peninsula and Taymyr – recorded all the words and phrase examples. At present dictionary articles consist of the following parts (database forms): "Concise article form", "Complete article form", "Phonetic form", "Russian meanings of the word", and "Phrase and phraseological examples in Nenets". The user may choose one of the main four forms to work with the Dictionary. They are mutually connected and allow the data search. The detailed description of Nenets Audio Dictionary may be found in [7].

The most interesting part of the Dictionary is the "Phonetic form", which consists of: sound button for standard pronunciation of the word; speaker's name; transcription of the word (in present version this field is left to be filled in by the user himself, using Nenets transcription symbols); number of syllables; accent detachment of syllable (meaning 0/1 for every syllable); word accent information; sound buttons for another variants of pronunciation and corresponding speaker names. The variants of pronunciation vary in different regions. Because of that the Dictionary may be completed by another speakers, as well as by new words, new phrase examples, and texts of different types. The Audio Dictionary may be used to represent and study all aspects of Nenets phonetics and grammar.

5 Nganasan Audio Dictionary

The necessity to create Nganasan Audio Dictionary was determined by the fact that at present no representative collections of Nganasan audio data exist (either in Russian institutes or libraries or abroad). The oral tradition plays a central role in the Nganasan culture, because the first alphabet for the language was proposed just in the last ten years. Now only few aged persons speak Nganasan really perfectly. Young teachers with native knowledge of Nganasan live far from larger educational centers in St. Petersburg or Moscow. Even at the Taymyr Regional Center for Advanced Education there are no Nganasan speaking persons younger than forty.

It is only in the last years that certain steps were made to support the Nganasan language and culture: several textbooks for children and adults as well as folklore materials have been published, mostly in Dudinka. The less becomes the number of qualified native speakers of Nganasan, the more important and urgent becomes the recording and documentation of this language – before the native competence diminishes and ultimately disappears.

Nganasan Audio Dictionary should fill a gap in accessible audible data on this language. The text basis for the dictionary is the Nganasan-Russian dictionary, which is now in print. It was compiled by two speakers of Nganasan – Nadezhda Kosterkina and Alexander Momde, and edited by T. Zhdanova and V.Gusev. The records were made by several Nganasan speakers, including Nadezhda Kosterkina herself and the wife of Alexandr Momde – Anna. Alexandr himself couldn't articulate when the recordings were made because of the hard disease; he died in winter 2000.

Speech corpus for Nganasan Audio Dictionary consists of recordings of Nganasan Dictionary made by two Nganasan speakers. The whole number of words is about 4000. Many of them are accompanied by phrase examples or by idiomatic expressions.

The supporting database proposes the following parameters of description: Nganasan word, grammatical characteristics of the word, word inflections, Russian translation, Nganasan phrase examples and Nganasan idiomatic expressions with their translations into Russian, sound file names for each Nganasan word or phrase. Working with the Dictionary the user has the possibility to listen to each Nganasan word/phrase by means of simple clicking on the corresponding button.

6 Nenets online Multimedia Phrase-book

Nenets Multimedia Phrase-book was prepared for the Web-publication within the framework of the joint project "Writing and teaching Samoyedic", made by the Research Group on Phonetics and Ethnolinguistics, University of Groningen, The Netherland, and the Department of Phonetics of St. Petersburg State University, Russia.

The Russian-Nenets phrase-book embraces 21 general topics. These topics and the majority of Russian expressions were borrowed from Nganasan-Russian phrase-book [1]. This approach is justified by the fact that Nenets and Nganasan people being geographical neighbors have much in common. The phrase-book was recorded by three native Nenets speakers, who represent three main dialects of the Nenets language: central (Bolshezemelsky), eastern (Yamalsky), and western (Kaninsky). All of them were born in tundra, and from their childhood they took part in nomadic reindeer graze. Now they are the students of St. Petersburg State Pedagogical University named after A.A.Herzen, and perfectly speak Russian. It should be mentioned here that now it is impossible to find a Nenets speaker, younger than 70 years old, who does not know Russian rather good. The speakers were asked to make translations to each phrase of the phrase-book and then to record all the phrase-book. Creating the database we tried to preserve the original spelling made by the speakers, though frequently one and the same word might be spelt in different ways, because there is not any normative spelling for Nenets still. Moreover, Nenets is mostly used as a spoken language. In the following two phrases one and the same underlined word "hour" is spelt in two quite different ways:

Сян <u>часна</u> маня ядабтагунь'? — *When (at what hour) we shall meet?*
Сидя <u>сяс</u> ваерамад обт' тэвгунь. — *Let's meet in two hours.*

The speakers could not translate into Nenets some phrases, describing atypical phenomena for the language. Thus, no speaker could translate "*At 5:15 p.m., Moscow time*". It turned out that the speakers from Yamal better know the deer-raising, and identify local birds and animals. However, it was interesting to find that in the speech of these young girls the beasts are called by the ancient allegorical names. For example, the following expressions have the meaning "*wolf*":

ӈэсында хэвхана пивна [ядерта] — *wolf* (Yamalsky dialect)
пиня'мэна эсыд хэвхана — *wolf* (Kaninsky dialect)
яхӑнанда тёняӈы — *wolf* (Bolshezemelsky dialect)

Two first phrases may be word-by-word translated as "*who* [*goes*] *around the settlement*". The speaker of Bolshezemelsky subdialect called the wolf "*the neighbor of our land*". The bears are called "*spring earth inhabitant*", the eagles are named "*big birds*":

нара ям' мядота — *bear* (Yamalsky dialect)
аркаюм' тиртя — *eagle* (Kaninsky dialect).

Nenets borrowed many words from Russian. For example, they did not have welcome greetings and borrowed it from Russian, as well as a number of "polite" expressions (e.g., see topics "Agreement/Disagreement", "Farewell", "Surprise", "Gratitude", "Approval").

The presented topics describes both the traditional Nenets life and housekeeping (e.g., topics "Family. Family relations. Age", "Hunting. Fishing. Deer-raising", "Flora and Fauna") and the new phenomena for the tundra life ("Shop", "Post. Telegraph. Telephone").

The audio phrase-book was created in the form of speech database, and then the database was adapted to the Internet and publish in the site: `http://www.speech.nw.ru/Nenets/`.

This phrase-book represents a good manual to get acquaintance with the Nenets language. Our further steps will be the translation of the phrase-book into English so that it may become accessible to more scientists and other users of the Web.

7 Conclusion

In the process of development of audio collections for endangered languages we have faced a number of challenges, which complicate the process of data collection and processing. First, there is no accepted norm both in pronunciation and spelling in most of minority languages. Because of that it is difficult to select the unique spelling. Without uniform spelling it is impossible to carry out the automatic search for words in the databases. The other problem is a high differences between generations of the minority people. The younger the speaker – the less is his (her) native language practice. Besides, there is some technical problems (e.g., national fonts representations for the Internet), which deserves further development and standardization.

The developments presented in the paper may be considered as the model for creation sound collection and speech databases for other minority languages of the Russian Federation.

8 Acknowledgments

The project "Nenets Audio Dictionary" and expeditions to Taymyr in 1996 and 2000 were supported by Russian Humanitarian Scientific Fund. The development of Nganasan Audio Dictionary is supported by Endangered Language Fund (project "Collection of Audio Material in The Nganasan Language"). The creation of Nenets online Multimedia Phrase-book is supported by Groningen University, the Netherlands within the framework of the project "Writing and teaching Samoyedic" – "Nenets Speech Varieties on the Internet".

References

1. Aaron, T., Momde, A.: Nganasan-Russian phrase-book. Dudinka (1985)
2. Berkov, V.P.: How to save the disappearing languages of Russia? In: Archives of the Languages of Russia. Groningen-St. Petersburg (1996) 39–44
3. Bondarko, L.V., de Graaf, T.: Archives of the Languages of Russia. Introduction. Groningen-St. Petersburg (1996) 5–12
4. Janhunen, J., Salminen, T.: UNESCO Red Book On Endangered Languages: Northeast Asia
   http://www.helsinki.fi/~tasalmin/nasia_report.html#TNenets
5. Lehtisalo, T.: Jurak-Samojedisches Worterbuch. Helsinki (1956)
6. Lublinskaya, M.D.: Audio dictionary as a mean for saving and studying endangered languages. In: The Proc. of the IX-th International Finno-Ugric Congress in Tartu. Materials of the Congressus Nonus Internationalis Finno-Ugrisrarum. 7 - 13.8.2000. Tartu (2000) 344 – 345
7. Lublinskaya, M.D., Sherstinova, T.Y., Kuznetsova, E.Y.: Multimedia Dictionary of Nenets. In: Proceedings of International Workshop on Computational Lexicography and Audio Dictionaries "COMLEX'2000", September 22-23, 2000, Patras, Greece (2000) 105–107
8. Nenets-Russian Dictionary: Moscow. "Sovetskaja Encyclopedia" (1965)
9. Nenets-Russian, Russian-Nenets Dictionary: Leningrad, "Prosveshchenije" (1982)
10. "Red Book of the Languages of Russia": Moscow, "Academia" (1994)
11. "Red Book of the peoples of the Russian Empire":
    http://www.eki.ee/books/redbook/nenets.shtml
12. Skrelin, P., Sherstinova, T.: Models of Russian Text / Speech Interactive Databases for Supporting of Scientific, Practical and Cultural Researches. In: Proceedings of II International Scientific Conference LREC-2000 (Language Resources and Evaluation), Athens (2000) 185–189
13. Vaba, L., Viikberg, J.: The Endangered Uralic Peoples. Short Reference Guide. Editor Andres Heinapuu. http://www.suri.ee/eup/nenets.html