

## ***THE URALIC LANGUAGES<sup>1</sup>***

**Marianne Bakró-Nagy**

### **1. THE IDENTITY OF THE URALIC LANGUAGES**

#### **1.1. The names 'Uralic' and 'Finno-Ugric'**

With more than thirty languages, the Uralic language family is the largest language family of Northern Eurasia. Its speakers live dispersed over a large geographical area which extends from the northern border of Norway in Scandinavia down south to Hungary (and its neighbouring countries) in East-Central Europe and eastwards to the Ob and Yenisei rivers and their tributaries in Siberia and the Taimyr peninsula in northern Siberia. This language family is also sometimes called Finno-Ugric: in everyday usage the label "Finno-Ugric" is more common and is used as a synonym of "Uralic". However, technically "Finno-Ugric" denotes only one of the two branches of the language family, while "Uralic" refers to the Finno-Ugric branch and the Samoyedic branch together. The term Finno-Ugric was originally created by analogy to the term Indo-Germanic, indicating the two farthest branches of the language family: Finnish + Ugric = Finno-Ugric. The term Uralic was created later from the name of the Ural mountains, referring to the place where the ancient homeland of the common ancestors of Finno-Ugric and Samoyedic peoples are thought to have originally lived and where a lot of them still live to this day, on either side of the mountain range. Speakers of the Uralic languages have come to be separated by great distances over the centuries and can no longer be considered to be genetically (i.e. biologically) related. The genetic relatedness of their languages and the linguistic distances between the languages can only be proven through meticulous and systematic analysis. In terms of the traditional language family model (which indicates the chronology of the splits within the family and which, as we will soon see [cf. section 1.2], can be contested in several points), the family relationships of the Uralic languages are as follows:

see separately!!!

Figure 1. The Uralic language family tree

The names in Figure 1 are the self-designations of the Uralic languages (i.e. the names used by the speakers). For instance, Mansi is the self-designation of the people called Vogul by others – most Uralic peoples have such a second name as well. As far as Uralic languages spoken in Russia are concerned, their second names spread through their usage in Russian and acquired negative connotations over the years. Because of this, the self-designations have gradually become officially used as well, and today these are names used in all contexts. The designations for Saami have gone through a similar process, although the Saami live also outside Russia.

---

<sup>1</sup> Preliminary version.

Figure 1 is not sufficient to indicate all of the Uralic languages and their various names – these are listed in Table 1, together with the extinct languages. In the table, column A lists older names or names used by outsiders, column B lists self-designations, while column C gives the localities where the speakers of the respective language live (i.e. regions where the given people or speech community predominates). The Uralic languages can be found all along the continuum extending from endangered languages to non-endangered ones. The extent of endangerment can be different even between different dialects of the same language. Differentiating between the various dialects of the Uralic languages, especially in the case of the Siberian ones, serves the purpose of indicating such differences. For the various categories of endangerment, see the *UNESCO Red Book of Endangered Languages* (<http://www.tooyoo.l.u-tokyo.ac.jp/Redbook/index.html>).

<i>A</i>	<i>B</i>	<i>C</i>
<b><i>URALIC LANGUAGES</i></b>		
<b><i>FINNO-UGRIC LANGUAGES</i></b>		
<b><i>Finnic languages</i></b>		
Estonian	eesti keel	Estonia and adjacent areas
Finnish	suomen kieli	Finland and adjacent areas
Ingrian (Izhor)	ižoran keeli: endangered	Russia (European part)
Karelian	karjalan kieli	Karelian Republic (European part of Russia), Finland
Livonian	liivõ keel': nearly extinct	Latvia
Lude	l'üüd'in kieli	Karelian Republic (European part of Russia)
Livvi (Olonetsian, Aunus)	livvin kieli	Karelian Republic (European part of Russia)
Veps	l'üd'ikel', bepsan kel'	Russia (European part)
Vote	vad'd'a čeeeli: nearly extinct	Russia (European part)
<b><i>Saami languages</i></b>		
Lappish	Saami	
	<i>Western Saami languages</i>	
	<i>Southern Saami</i>	
	South:seriously endangered	Sweden

	Ume: nearly extinct	Sweden
	<i>Northern Saami</i>	
	Pite: nearly extinct	Sweden, Norway
	Lule	Sweden, Norway
	North	Norway, Sweden, Finland
	<i>Eastern Saami languages</i>	
	Inari: seriously endangered	Finland
	Kemi: extinct	Finland
	Skolt: seriously endangered	Finland, Russia (European part)
	Akkala (Babino): extinct	Russia (European part)
	Kildin: seriously endangered	Russia (European part)
	Ter: nearly extinct	Russia (European part)
<b><i>Volgaic languages</i></b>		
Mordva, Mordvinian	<i>Mordva languages</i>	
	Erzya	Autonomous Republic of Mordovia (Russia, European part)
	Moksha	Autonomous Republic of Mordovia (Russia, European part)
Merya	extinct	
Muroma	extinct	
Meshtser	extinct	
Cheremis	<i>Mari language</i>	Republic of Mari El (Russia, European part)
<b><i>Permic languages</i></b>		
Votyak	Udmurt	Udmurt Republic (Russia, European part)
Zyrian	Komi	Komi Republic (Russia, European part)
Komi-Permyak	Permyak	Perm Region (Russia,

		European part)
<b><i>Ugric languages</i></b>		
Hungarian	Magyar	Hungary and adjacent areas
<b><i>Ob-Ugric languages</i></b>		Khanty-Mansi Autonomous District (Russia, Asian part)
Ostyak	Khanty Southern dialect: possibly extinct Eastern dialect: seriously endangered Northern dialect: endangered	
Vogul	Mansi Southern dialect: extinct Western dialect: possibly extinct Eastern dialect: nearly extinct Northern dialect: seriously endangered	
<b><i>Samoyedic languages</i></b>		
Yurak Samoyed	Nenets Tundra Nenets: endangered Forest Nenets: seriously endangered	Nenets National District (Russia, European part), Yamal Nenets National District (Russia, Asian part)
Yenisey Samoyed	Enets Tundra Enets: nearly extinct Forest Enets: nearly extinct	Taimyr Autonomous Territory (Russia, Asian part)
Tawgi Samoyed	Nganasan: seriously endangered	Taimyr Autonomous Territory (Russia, Asian part)
Ostyak Samoyed	Selkup Northern Selkup: endangered Central Selkup: nearly	

	extinct
Kamass	extinct
Mator	extinct
Yurats	extinct

Table 1. All known varieties of Uralic

## **1.2. The origin and family affiliation of the Uralic language family**

The Uralic languages form an independent language family. Ever since the three big language families – the Indo-European, the Altaic and the Finno-Ugric – of Eurasia were identified in the 19th century, there have been repeated attempts to prove the relatedness of their languages or their ancestors. Thus arose the hypotheses of the relatedness of Indo-European and Finno-Ugric, and, especially, that of Altaic and Finno-Ugric languages. The greatest, actually insurmountable obstacle to the former lies in the typological differences between the two language families in their phonologies and especially in their morphologies: it would require considerable effort to find a common denominator for the basically agglutinative morphology of the Uralic languages and the inflectional morphology of the Indo-European languages. From a typological point of view, relating Altaic with Finno-Ugric would not be as hopeless (this hypothesis has had many followers over the years), but it raises a different problem, namely a methodological one, arising from the Altaic side: the Altaic group, comprising the Turkic, Mongol and Tungusic languages is not a “language family” in the same way that the Indo-European or Finno-Ugric languages form a language family. This, however, does not mean that comparative typological research on Altaic vs. Finno-Ugric languages would be pointless or unimportant. A further possibility of relatedness lies with Yukaghir, an isolate spoken in north-eastern Siberia and grouped together areally with other Paleo-Siberian languages. The greatest problem in connection with hypothesizing genetic relatedness with Yukaghir is that the time-depth involved would be greater than what the comparative method or the diffusionist model would be able to handle – a problem which, in fact, arises in the case of relating Indo-European and Finno-Ugric as well. As we will see below (section 3.2), contact-induced changes exist between and seriously affect Finno-Ugric and all the discussed languages, primarily in their lexicons, but what can be stated with certainty is only that no genetic relatedness between the Uralic language family and other languages and language families can be demonstrated through methodologically consistent analysis.

## **1.3. The distinctive features**

Typologically, Uralic languages are typically agglutinative languages. Even though as a result of historical changes the agglutinative characteristics of some Uralic languages have become less pronounced, most Uralic languages, to varying degrees, preserve the most prominent features of agglutination, that is, they typically use affixes to express morphological and morphosyntactic relationships.

### **1.3.1. Phonetics, phonology**

As far as their phonological characteristics are concerned, most Uralic languages have a great number of consonants and a relatively small number of vowels. The absolute numbers of each differ in every language and often in different dialects of the same language as well. As far as the quality is concerned, consonants are typically characterised by the opposition of voiced vs. voiceless and that of palatal vs. non-palatal consonants, while the long vs. short opposition is typical for quantity.

In manner of articulation of consonants, Uralic consonant systems all contain, within obstruents, a series of stops (e.g. *p t k*) and fricatives, and among the latter, a series of sibilants (e.g. *s š ś*) and fricatives per se (e.g. *f γ θ*), as well as a series of affricates (e.g. *c č ć*). There are Uralic languages which lack affricates: for instance, Finnish has only one sibilant, *s*. In every Uralic language, sonorants include nasals (e.g. *n m ŋ*), liquids (e.g. *l l' r*) and semi-vowels (e.g. *w j*). Differences between Uralic languages exist in what elements each subclass of consonants contains as well as in the place of articulation of consonants. As far as the latter are concerned, Uralic languages typically have labial, dental, alveolar, palatal and velar consonants, and glottals in the Samoyedic branch. Among glottals, glottal stops play an important role in the morphological processes of Samoyedic languages. In those Uralic languages that have the voiced vs. voiceless opposition as well as the palatal vs. non-palatal opposition in consonants, the number of dental, alveolar and palatal consonants is high: it is highest in Udmurt with a total of 15.

Within obstruents, the voiced vs. voiceless opposition (e.g. *p – b, t – d, k – g*) is lacking in some Uralic languages, for instance in Finnish, Khanty, and Mansi, but is typical in the Permic languages (e.g. *ker* 'taste' ~ *ger* 'plough') and Hungarian, as well as in North Saami and Nenets. Sonorants, which tend to be voiced in the world's languages, are not all voiced in all Uralic languages: voiceless liquids and nasals are found, for instance, in some dialects of Khanty (e.g. the Surgut dialect), Moksha Mordva and some Saami languages.

A palatal series of consonants is present in all Uralic languages except Finnish, with Mordva having palatal counterparts of almost all its consonants (*c – ć, d – d', l – l', n – n', r – r', s – ś, t – t', z – ź*).

The distinction of quality, that is, the long vs. short opposition of consonants (e.g. *p – pp, t – tt, k – kk*) is present in almost all Uralic languages. Some Saami languages and Estonian deserve special mention in this regard as they have three types of consonant length (short, long, and overlong). This triple opposition in length is characteristic of vowels as well as consonants in Estonian. The quality opposition is not as typical in vowels as in consonants in Uralic languages: it is absent in Erzya Mordva, Udmurt, and Komi. In some languages the length distinction in vowels goes together with full vs. reduced phonation. The long vs. short opposition of vowels is present in Hungarian, Finnish, North Mansi, and Forest Nenets, whereas the full vs. reduced opposition is there in Hill Mari, Tromagan Khanty (*e-ě, o-õ, a-ã* etc.), and Tundra Nenets. Several Uralic languages also have diphthongs, some a relatively high number, e.g. Nganasan, which has 19 of them.

Perhaps the most striking feature of Uralic vowel systems is the presence of front rounded and back unrounded vowels (e.g. *ö - ɘ, ü - ɨ*) in a number of the languages, like Udmurt, Komi, Khanty, Mansi etc. Most Uralic languages distinguish between three vowel heights, high, mid, and low. Erzya Mordva has the smallest inventory of vowels among Uralic languages with its five-vowel system (*a e i o u*).

The most important and probably best known phonetic/phonological characteristic of Uralic languages is vowel harmony, which defines the quality of affix vowels in relation to the stem vowel. There are two types of vowel harmony in Uralic languages, backness harmony (e.g. Hungarian *kert-ben* 'garden-in = in the garden' vs. *fal-ban* 'wall-in = in the wall') and rounding harmony (e.g. *föld-höz* 'earth-to = to the earth' vs. *kert-hez* 'garden-to = to the garden'). While backness harmony is present in the majority of Uralic language, rounding harmony is there in only some of them, with many exceptions even there. A phenomenon even less frequent in the world's languages than vowel harmony, gradation is also present in some Uralic languages, namely in Finnish, in several of the Fennic languages, in Saami and in the Samoyedic languages. Gradation is connected with the syllable structure of languages. (Most Uralic languages permit only vowels as syllabic segments. Despite this, some dialects of Khanty, for instance, the Surgut dialect, also have syllabic consonants, which are always sonorants.) The typical simplest structures in Uralic languages are usually CV (open) and CVC (closed) syllables, forming stems of two syllables, which can be expanded to several times their original length through affixation. The simplest manifestation of gradation modifies the last open syllable of a word and the pre-vowel consonants and consonant clusters, triggered by the closing of the last open syllable, which, in turn, is caused by affixation, i.e. the attachment of a consonant affix on a vowel-final form. Gradation is simplest in Finnish, where it operates on three voiceless consonants, *k, t, p*, and their geminated counterparts (*kk, tt, pp*). Estonian and Saami have a much more complex system of gradation, depending, in Saami, also on the position of the vowel inside the word. In Nganasan gradation is manifested even in inflection (e.g. *t ~ δ* gradation: *kamba* 'spring [season]' : *kamba-δu* 'for spring [dative]', *batə* 'fate' : *baδə-* 'fates').

In Uralic languages stress does not play a role in differentiating meaning, and in most of them the primary stress is placed on the word initial syllable. Stress relations are, however, more complex than this: Uralic languages can be roughly divided into two groups depending on whether word stress is fixed or not. It is fixed in all those languages that have word initial stress (e.g. Hungarian), stress on the second syllable (e.g. South Mansi), penultimate stress (e.g. Western Mari and Nganasan: '*koru*' 'house' ~ *ko'ruδə*' 'houses'), and stress on the ultima (Udmurt). In those languages where stress is not fixed, it can be free moving (e.g. Komi *munenj* ~ *munenj* ~ *munenj* 'they go'), it can depend on the quality or quantity of the vowels (e.g. Moksha Mordva, several Khanty dialects, Tundra Nenets), or on the open or closed nature of the syllable (some Khanty dialects). In the languages where stress is fixed, such as on the first syllable in Northern Mansi, the odd number syllables of a word form lengthened by affixation can also receive

stress. The stress on the third syllable in Northern Mansi can be almost as strong as that on the first (*tārati* 'let' ~ *tāratankwe* 'to let [infinitive]').

### 1.3.2. Morphology

Uralic languages are basically synthetic as far as the use of inflection for number and case marking of nouns, person, number, tense and mood marking of verbs etc. is concerned. This does not, however, mean that the Uralic languages have no analytical characteristics: a case in point is the existence of postpositions in all Uralic languages for marking place and time primarily. Postpositions exist in all of them without exceptions (Mordva: *ikēle* 'in front of', Northern Mansi: *xalt* 'between', nyenyec *yeq* 'toward').

Uralic languages are typically regarded as agglutinative, and this is generally true. It is the Eastern Finno-Ugric languages (especially Mansi) and Mari which are closest to the "pure agglutinative" model (with a one-to-one relationship between inflection and function) (e.g. Mari: *pi* 'dog' : *pi-wlä* 'dog-plural=dogs' : *pi-län* 'dog-dative=for dog' : *pi-län-wlä* 'dog-dative-plural=for dogs' etc. Some flecional characteristics are present in the Fennic languages, the Saami languages, and in the Northern Samoyedic languages (especially Nenets and Nganasan). Uralic languages have two types of inflections: prefixes are very rare (there are a total of two: the Finnish privative prefix *epä-edullinen* 'un-favourable', and the Hungarian superlative marker *leg-magasabb* 'most-tall=tallest'), while suffixes are common (e.g. Northern Mansi *ērjγ-nəl* 'from-song=from song'). Uralic languages are, thus, characterised by a dominance of suffixes. Their ordering in the word form follows universal rules, that is, inflectional suffixes are attached peripherally, after derivation suffixes, while derivational suffixes are closer to the stem. There are some exceptions to this, e.g. in Mari and Hungarian (e.g. Mari *puš-āš-so jeη* 'boat-in-ADJECTIVE man = in-the-boat man = man in the boat'). The relative ordering of various inflections within the word form is also not identical, sometimes even within the same language, e.g. in Mari nominal inflection both the possessive+plural and the plural+possessive order is possible (*joltaš-em-wlak* friend-1SGPOSS-PL ~ *joltaš-wlak-em* friend-PL-1SGPOSS 'my friends').

In Uralic languages the morphological categories marked on nouns are person, number, and case. Number is marked on pronouns as well as nouns. Of the Uralic languages, the Saami, the Ob-Ugric and the Samoyedic languages distinguish between singular, dual and plural number (e.g. Northern Mansi *āpa* 'crib-SG' ~ *āpaγ* 'cribs-DU' ~ *āpat* 'cribs-PL').

The case systems of all Uralic languages are rich, but there are great difference among the languages and their dialects as to which one has how many cases. There are, however, six cases that all Uralic languages have: the nominative, the accusative, the genitive, and the local cases typical of the tripartite location marking system of the Uralic languages, namely, the locative (*where?*), the lative (*where to?*) and the ablative (*where from?*). Several of the Uralic languages have a great number of cases in addition to these. In all Uralic languages the nominative always has zero inflection, but in some of them (for instance, in Khanty, Mansi

and Estonian) the accusative, and rarely the genitive is also zero-marked. The tripartite location marking is notable because, based on the three basic local cases, Uralic languages have developed complex systems of local case marking, differentiating between exterior and interior local relationships as well as an additional proximity relationship in Hungarian (e.g. *asztalnál* “near table”, *asztalhoz* ‘to near table’, *asztaltól* ‘from near table’) and a similar system in Komi. The case systems of Uralic languages have become even more complex through the application of local cases to abstract relations. Within the category of nominals in Uralic languages an important role is played by the system of personal possessive markers, which can refer to the person and number of both the possessor and the thing possessed. In those languages where personal possessive markers have disappeared, such as some of the Fennic languages, possessive pronouns have developed instead. The non-marking of the possessor as a result of the absence of the genitive case occurs in Mansi, for instance: the possessor precedes the thing possessed and is unmarked, while the thing possessed is marked for the appropriate person and number (Northern Mansi *āyirís-ət kola-nəl* ‘girl-plural house-3PLPOSS = girls’ house’). In some of the Uralic languages the possessor is marked for genitive case while the thing possessed is also marked with the personal possessive suffix (this is characteristic primarily of Mari, but also of Mordva and the Permic as well as of Samoyedic languages). The “inverse situation” is found in the Fennic and Saami languages, where the possessor is marked for the genitive case, while the thing possessed is unmarked (e.g. Finnish *miehe-n talo* ‘man-genitive house = man’s house’).

Within the class of nominals, in addition to nouns and adjectives functioning as nouns, various pronouns, namely, personal, (nominal) interrogative, indefinite, negative, universal and demonstrative pronouns, can also be inflected in their own individual paradigms. Another characteristic feature of Uralic languages is the absence of grammatical gender. Despite this, some of them differentiate between animacy and inanimacy in various grammatical contexts and through various means. The morphological marking of definiteness, for instance, in Mordva, has to be noted specially. Among Uralic languages those that have definite articles constitute exceptions: even though some languages have “article suspect” forms, only Hungarian has true articles. The definiteness of a nominal is marked by inflection in Mordva (*kudo-ś* ‘house-the = the house’).

In both Mordva and Nenets nominals can be marked by verbal person/number markers, and in Mordva even by tense markers: Mordva *lomań* ‘man’, *lomańan* ‘I am a man’, *lomańal’iniń* ‘I was a man’.

One of the most important universal characteristic of adjectives is their gradability. While in Hungarian and a number of Uralic languages their gradability is a matter of fact phenomenon and is marked through inflection, in other Uralic languages – for instance, in Mordva and Mansi – adjectives are not gradable at all, and, instead, various syntactic constructions express comparative and superlative meaning. In the former, comparative meaning is expressed by the base form of an adjective preceded by the demonstrative pronoun inflected for the ablative case, and superlative meaning by the base form of the adjective preceded by an inflected

demonstrative pronoun: *śed'e mazij* 'more beautiful', *śex mazij* 'most beautiful'. In Mansi, forms expressing comparison carry the function of comparatives and superlatives. The basis of comparison is in the ablative, and what is compared to it is in a base form following it: Mansi *wōt an<sup>u</sup>mnal jōr* 'wind-from.me-strong = the wind is stronger than me'.

While the central inflectional categories of nominals are number marking, cases, and the personal possessive inflections, those of verbs are the marking of person, number, tense, mood, aspect, and genus verbi.

Almost all Uralic languages have person marking in first, second, and third person singular, (dual), and plural numbers (although Hungarian constitutes an exception with zero marking of 3rd person singular in the indefinite conjugation: *hoz* 's/he brings', *lát* 's/he sees', *jár* 's/he goes'). The presence of indefinite and definite conjugations is not at all general among Uralic languages; besides Hungarian, it is there in Khanty, Mansi, and some Samoyedic languages, and, most significantly, in Mordva, where all verb forms express the person and number of the object as well.

Uralic languages morphologically distinguish between only two tenses – present and past – whereas future is expressed syntactically or through auxiliaries (in those languages where it is marked at all), e.g. with the auxiliary *fog* in Hungarian. Aspect is not a morphological category in Uralic languages, while mood is. Indicative mood is unmarked in all of them, and likewise all mark, with inflection, the imperative as well as another mood, desiderative, which is similar to conditional. A characteristic feature of Uralic languages such as Estonian, Mari, Udmurt, Komi, Mansi and the Samoyedic languages is the evidential mood: it is used if the subject was not an eye witness to the event in question.

### 1.3.3. Syntax

There exists an image of the Uralic languages meeting the typological observations that agglutinative languages tend to have SOV word order and that sentence constituents can be ordered freely in the sentence. In the following Estonian sentence and its Hungarian equivalent, for instance, this seems to be true since the constituents can be ordered in any of the possible combinations: *Tihti taevas tähti nähti. Gyakran az égen csillagokat láttak.* 'often in the sky stars they saw = They often saw stars in the sky.' Despite this, if we consider the basic word order of Uralic languages in greater detail, we can distinguish between three groups of languages. The first group comprises languages with what is considered the most general order, SOV. In these languages it is not only the object that is placed between the subject and the verb but all other sentence elements as well. This is characteristic primarily of Eastern Uralic languages (that is, the Ob-Ugric and Samoyedic languages) and, to some extent, of Mari and Udmurt as well. E.g. Northern Mansi *ńawramət wōrnəl χājteγət* '(the) children from (the) forest run (out) = The children run out of the forest.'. In the second group of languages, the subject precedes the verb, but the object (and other sentence elements) follow it, that is, the SVO order is predominant. Fennic languages and Mordva belong in this group. E.g. Erzya Mordva *te'am kundaś ka isak* 'my father caught a fish

yesterday’. In the third group of languages both of the above mentioned orders are equally common and generally used, such as in Komi and Nganasan.

Which Uralic languages are characterised by freer or more fixed word order does not depend on either geographical (areal) characteristics or the genetic grouping of Uralic languages. Just like Khanty and Mansi, the Samoyedic languages have a fixed word order, with the exception of Nganasan, in which the basic order is SOV, but depending on the topic-focus relations of the sentence, constituents can be moved to a focus position, thereby producing SVO or OVS order. SOV: *mə̀nə́ n̄iəmə́ d’embijə̀ndim* ‘I my child dress = I dress my child’, SVO: *s̄ij̄i h̄odətəmur̄h̄<sup>u</sup>antu hot̄iur̄əj* ‘s/he is writing letters’, OVS: *obədə hiritim̄iŋh̄<sup>i</sup>anti n̄em̄jbt̄i<sup>u</sup>amə́* ‘lunch cooks my wife = my wife is cooking lunch’.

The languages with SOV word order usually have postpositions, while those with SVO word order have prepositions. Even though some prepositions can be found in several Uralic languages, the Fennic and Saami languages have most of them, even though postpositions still outnumber prepositions in these languages. As far as modifying phrases are concerned, in Uralic languages (including the Fennic languages), adjectival, numeral, and possessive phrases always precede the sentence element they modify, just as it would be expected in SOV languages. The situation is the same as regards participial constructions as well: in Uralic languages these always precede the sentence element they modify.

The concord which marks close grammatical relations between sentence elements through formal means – through inflection in agglutinative languages, for instance – can vary in extent and type among languages. Uralic languages do not use agreement to the same extent and in the same way. The most general type of concord is that between subject and verb in person and number. E.g. Northern Mansi *luw χājti* ‘(the) horse runs, *luwət χājteγət* ‘(the) horses run’. If a numeral quantifier premodifies the subject or if the subject is a collective noun, both the subject and the verb are singular. There are, however, types of concord which are not shared by all Uralic languages. Thus, for instance, there are two kinds of concord that exist in Uralic languages between the head of the noun phrase and its attributive premodifier. In most Uralic languages, the premodifier does not agree in number and case with the head noun: Northern Mansi *jomas χum* ‘good man’, *jomas χumitnəl* ‘from good men’. Number and case concord is most complete and most widespread in the Fennic languages. Typologically, an important feature of concord in Uralic languages is the agreement of the demonstrative pronoun in number and case systematically in Komi and Hungarian and less systematically in Finnish, Saami, and Mari. E.g. Hungarian *azokban a házakban* ‘those-in the houses-in = in those houses’; *azok mögött a házak mögött* ‘those behind the houses behind = behind those houses’.

As far as syntactic structures are concerned, Uralic languages are characterised, on the one hand, by the lack of overt linguistic material (e.g. conjunctions) that would mark the relationship of clauses in subordinate constructions – a feature that is found in all of them (although this, to a great extent, is the result of contact effects in them). On the other hand, they are

characterised by non-finite (infinitival and participial) structures integrating in a simple sentence the relationships that are usually expressed in coordinated clauses. An example to the latter is the following Northern Mansi sentence (involving an infinitival construction replacing an object): *akmatērtñ mātāpriś kwonəl kwālurŋkwe noməlmatas* 'suddenly mouse out.to.go thought = suddenly the mouse decided to go out'. Such structures are more common in the Eastern Uralic languages, that is, in Samoyedic and Ob-Ugric languages.

**1.3.4. Alphabet and spelling system**

The Uralic languages can be divided into two groups as far as their writing systems are concerned, and this division follows directly from their geographical location and the history of their literacy (cf. 4). The Uralic languages in Russia (with the exception of Fennic languages) use the Cyrillic alphabet (which is often not very useful, since this system does not indicate a lot of phonetic and phonological distinctions despite the use of various symbols and diacritics in addition to letters). The Saami of Russia have also used the Cyrillic alphabet since 1982, even though they used the Latin alphabet before then. The other Uralic peoples use the Latin alphabet. From a historical point of view the Old Komi writing system, the *Abur* (named after the first two letters of the alphabet: *an + bur*) deserves special mention (see Figure 3 for an example inscription): it was invented in the 14th century by St. Stephen of Perm by combining Ancient Greek, Aramaic, Old Slavonic letters and owners' marks and was in use for almost 300 years, until the 17th century.

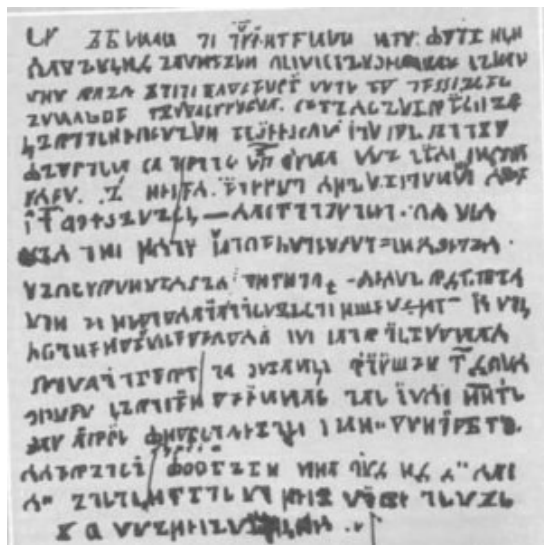


Figure 3. An Abur inscription.

**2. THE HISTORY OF THE URALIC LANGUAGES**

**2.1. The emergence and periodization of the Uralic languages**

Uncovering the formation and history of the Uralic languages is made difficult by issues which do not arise in the case of Indo-European languages at all or only to a much lesser extent and which are due to the fact that most of the Uralic languages are unwritten. This means that reconstruction provides the only means for positing their common ancestors as well as for providing clues for the history of the individual languages, relying on similarities between the synchronic systems of the languages, on conclusions drawn from comparing dialects, and, whenever possible, on language contact effects with visible results. Reconstruction of the Uralic protolanguage can only start after that. Uncovering the history of Uralic languages is further complicated by the fact that, again, unlike in the case of proto-Indo-European, non-linguistic (archeological, historical and anthropological etc.) evidence providing information on the structure and cultural anthropological characteristics of the society in which proto-Uralic was spoken is scarce due to the small number of sources and their lack of detail. One last, practical detail plays a great role in making investigation even more difficult: the smaller Uralic languages such as the Permic, Volgaic, Ob-Ugric and Samoyedic languages are spoken in areas of considerable geographical isolation in Russia (i.e. the former Soviet Union) which, due to historical and political reasons were almost completely inaccessible for decades. With these factors in mind it might be easier to see why it is difficult to describe the histories of these languages and why making tentative statements about them is sometimes the only possible option.

In outlining the histories of Uralic languages it is best to start with where, to the best of our present-day knowledge, the ancient proto-homeland of the ancestors of the speakers of the Uralic languages was and approximately when they lived there. Its exact geographical boundaries are not known, but it most likely was between the Baltic area and western Siberia, where the Uralic speakers moved to from the Ural mountains further south. This presupposition is supported by solid evidence and synthesises competing hypotheses. Historical linguists also agree that the Uralic proto-homeland was smaller than the area where Uralic peoples live today. The proto-homeland was a scarcely populated forested area where the ancestors of the Uralic speakers lived in a Neolithic community of natural economy in the 6000–4000 BC. The Uralic population most likely arrived here from different directions: from the area of present-day Ukraine, the lower Volga region, and from the direction of the Caspian Sea and the Aral Sea in the south. The first split of the unified Uralic population is hypothesised around 4000 BC. Even though there is no complete agreement in the periodization, it is almost certain that the first split involved the Samoyedic group. The descendants of proto-Samoyedic speakers live east of the Ural mountains today. After the split within proto-Samoyedic a number of Samoyedic languages (like Yurats) must have disappeared without a trace. The death of various Samoyedic language continues to this day: the southern Samoyedic language Mator became extinct (displaced by Turkic languages) in the early 19th century, while another, Kamass died when its last speaker died in 1989.

The Finno-Ugric branch is a branch of Uralic larger than Samoyedic, with more speakers and much more complex in its varieties. As far as the grouping of

the languages within the Finno-Ugric branch is concerned, there are two competing views. According to one, the proto-Finno-Ugric population broke up with the separation of the proto-Ugric population (the ancestors of the Ob-Ugric speakers and of Hungarians) at the beginning of the Bronze Age around 2000 BC. At this time the Ugric peoples still lived east of the Urals, while the Fennic peoples lived in the forested area of Eastern Europe extending from the Urals to the Baltics. This split was followed by the separation of the speakers of proto-Fennic-Volgaic (i.e. the ancestors of Fennic, Volgaic and Permic languages) and then their split into three. According to the other view, it was the Western Finno-Ugric languages, that is, the Fennic and the Saami languages, that separated first, not very long after the Samoyedic split, with proto-Fennic-Saami splitting in the second half of the 3rd millennium BC. This was then followed by the splitting off of Mordva, then of Mari, and finally of proto-Permic. This second view refers, first, to the shared similarities and elements that can be traced to a common ancestor, second, to all the foreign and areal influence that can be detected in these languages, and third, to their shared innovations. According to this view, then, the “traditional” family tree of the Uralic languages is as given in Figure 3:

see separately!!!

Figure 3. Alternative Uralic family tree.

### **3. THE GEOGRAPHY OF THE URALIC LANGUAGES**

#### **3.1. The expansion of the Uralic languages**

While significant changes in climatic conditions (such as the shrinking of the ice sheet and the subsequent changes in flora and fauna, resulting in changes in natural economy) most likely played a role in populating the Uralic proto-homeland as well as in the splitting of the proto-Uralic population, in the later movements of the Uralic peoples the influence of other populations was also significant. The splitting and wandering off of subsequent groups of Uralic resulted in the geographical location of today’s Uralic languages discussed in sections 1.1 above and 3.2.1 below. The effect of changing climatic and geographical conditions on the changes in the settlements is best illustrated by the prehistory of the speakers of Ugric languages.

The fact that the three Ugric languages – Khanty, Mansi and Hungarian – had to have a shared past is more or less commonly agreed on. It is supported by shared vocabulary and shared phonological and grammatical characteristics which are not shared by any other Uralic language. If a proto-Ugric language ever existed, however, it most likely existed for only a short period of time. In the 2nd millennium BC the life-style characteristic of proto-Ugric peoples involved animal husbandry and some basic agriculture (horse breeding became predominant under Iranian influence – Hungarian *ló* ‘horse’, *nyereg* ‘saddle’, *fék* ‘bridle’ and their Ob-Ugric equivalents date back to the Ugric era), and they most likely lived in the Ural mountains and in Western Siberia. Due to the drier weather in the 1st millennium BC, the Ob-Ugric peoples (the ancestors of the Khanty and the Mansi) moved northeast to the region of the Ob river (which is also where their name

comes from), while the ancestors of Hungarians started to move south around 800-700 BC, partly in reaction to another climatic change. Whether this happened on the western or eastern side of the Urals is debated. (The split of the Mansi from the Khanty happened on the western, European side of the Urals even though they moved to the eastern, Asian side later.)

As a result of the great nomadic wanderings, the ethnic composition of Eurasia went through considerable changes in the first centuries AD. The great movement of populations started by the Huns in the 3rd-4th centuries brought the first larger groups of Turkic language speakers of Central Asian origin to the steppes of Western Siberia and Eastern Europe. The ancestors of the Hungarians were also forced to move at this time, relocating to the area between the Volga river and Ural mountains. The greater part of ancient Hungarians left Bashkiria most likely around the middle of the 8th century, probably due to their displacement by Volga Bulgarians coming from the south, and moved first to the land of the Khazars (the area between the Don river and the Sea of Azov), moving further west to between the Dnieper and the Dniester rivers around 850, and then on to their final destination in the Carpathian Basin in 895.

The ancestors of the Ob-Ugric peoples lived over an extensive area back in the middle ages. The Ob-Ugric groups that moved to the western Urals (i.e. the ancestors of the Mansi) stayed there until about the 17th-18th century, which is when they were forced to move to the eastern side of the Urals. Since then the Ob-Ugric peoples have been living in the taiga and tundra regions of western Siberia, where their life-style had to undergo considerable change. Their old life-style and culture is preserved only in their rich folklore and belief systems, which, among other things, testify to their horse breeding past.

### **3.2. Uralic in contact**

The contact of Uralic languages with other languages is best examined from two different aspects: first, what kind of contact phenomena can be found in proto-Uralic, proto-Finno-Ugric and other proto-languages of the language family, and second, what contact phenomena exist in the individual Uralic languages.

As far as the language contacts of proto-Uralic and proto-Finno-Ugric are concerned, the most important source of contact is proto-Indo-European. As has been referred to above (see section 1.2), this poses some serious methodological problems. Despite this, there are quite a few words that are part of the basic vocabulary of proto-Uralic and proto-Finno-Ugric which show surprising similarity to Indo-European forms. Some examples are as follows: Finnish *nimi*, Hungarian *név*, Mansi *nam* etc. < proto-Uralic *\*nime* 'name' and Latin *nōmen*, Avesta *nāma*, *naṃa* etc. < pre-Iranian *\*en(o)m̥n̥*, *\*nōm̥n̥* 'name' or Finnish *vesi*, Hungarian *víz*, Mansi *wit* etc. < proto-Uralic *\*wete* 'water' and Avestan *vaidi*, Gothic *watō*, Old Church Slavic *voda* etc. < pre-Iranian *\*wed*, *\*wod*, *\*ud* 'water' and others, that is, approximately between 20 and 25 Uralic and Finno-Ugric words share similarity of form and meaning with Indo-European words. Archeological findings show that language and cultural contacts played a considerable role in the early history of the Finno-Ugric peoples. They lived in

proximity to Indo-European peoples but also had contacts with the ancestors of Paleo-Siberian peoples (Yukaghir, Ket). Between 2000 and 1000 B.C., the contacts with peoples living to the south played an especially important role in the life of the Finno-Ugric peoples. Indo-European peoples speaking most probably proto-Iranian lived on the Eastern European and Western Siberian steppe at this time, engaging in agriculture and animal husbandry, and possessing the knowledge to make metal. The above and similar examples as well as words referring to metal working, trade, and animal husbandry indicate contact with and, most likely, borrowing from these peoples. The Permic, Volgaic and Ob-Ugric languages borrowed early Iranian words later as well, their number totalling almost one hundred.

Such contacts existed not only between Uralic and Indo-European peoples or between Uralic, Finno-Ugric and Turkic peoples, speakers of Mongolian, Tungusic and even Yukaghir. The following examples stand out: Finnish *ala*, Hungarian *al-*, Mordva *al* stb. < proto-Uralic \**ala* 'under, lower part' and Turkic \**al* 'side, lower part'; Finnish *nuoli*, Hungarian *nyíl*, Udmurt *ñil*, *ñel* 'arrow' etc. < proto-Uralic \**ñele* 'arrow' or Tungusic *ñur*, Manchu *niru* 'arrow'; Finnish *keri*, Hungarian *kéreg*, Mansi *kēr* < proto-Finno-Ugric \**kere* '[tree] bark' és Yukaghir *χar* 'hidem skin' etc. Further research is needed to clarify what kind of contacts resulted in these similar and identical forms.

Bilingualism and multilingualism is typical among the speakers of Uralic languages today, with the exception of various proportions of speakers of Hungarian, Finnish, and Estonian. This follows directly from facts such as the following: speakers of Saami live in four countries, and the speakers of the other smaller Uralic languages live in Russia, where all of them have learned (or are learning) Russian as a second language. The whole picture has, however, always been more diverse and complex. When we systematise the contacts of Uralic languages by the contact languages in question, two types can be distinguished: contacts with other Uralic languages vs. contacts with non-Uralic languages.

Language contact can be best traced through its lexical effects. With the help of the loanwords borrowed by Uralic languages we can recreate the map showing the location of Uralic languages and their non-Uralic neighbours: starting in the west, the Fennic and Saami languages borrowed words from Germanic and Baltic languages, and Komi borrowed from Baltic languages to a lesser extent as well; Mordva, Mari and Udmurt borrowed from Chuvash; Mordva, Mari, Udmurt and Ob-Ugric languages from Tatar; Khanty and proto-Samoyedic from Tungusic etc. After the break-up of proto-Samoyedic, the individual Samoyedic languages were in contact with almost all the neighbouring languages, that is, with Turkic (Khakas, Tuva, Dolgan), Mongol (Buryat), Tungusic (Evenki) and Yeniseic (Ket, Kott, Arin) languages.

The fact that Fennic–Germanic language contact had structural effects has been known for longer than, for instance, what kind of effects the contact between Mari and Chuvash, or between Nganasan and Russian had. At the same time when, in the case of the Mari–Chuvash contact, non-Finno-Ugric patterns emerge in Mari as a result of code-copying, on the Taiymir Peninsula data from older Nganasan

speakers indicate that, as a result of language mixing, a pidgin, namely, Taiymir Pidgin Russian existed, which preserved Uralic morphosyntactic patterns but acquired Russian lexical and grammatical elements. In some cases typological similarities emerged, for instance, between Selkup and Ket, or between Sayan Samoyedic (Kamass, Mator) and Khakas or Tuva.

Contact among various Uralic languages was prominent in the northeast and in the Baltic region among Finnish, Estonian and Saami, where, for example, Estonian borrowed from both Finnish and Saami. In addition to historical sources, lexical borrowings also attest to intense contact among Komi, the Ob-Ugric languages, and the Samoyedic languages (especially Nenets, and, to a lesser extent, Selkup). This contact situation is clearly supported by their relative geographical location and, in the case of the Ob-Ugric and Nenets speakers, the reindeer herding and shepherding life style they share to the present. Also, Komi–Mansi contacts have been shown by sources to exist beginning with the 10th century, which was due to the vigorous trade relations the Komi maintained.

Finally, contact with Russian has been considerable: there is no Uralic language in Russia which has not had contact with it since the 16th-17th centuries. The type and intensity of contact between Russian and the individual Uralic languages varies greatly, however. While, for instance, Mordva and Komi have had long term intense social and cultural contact with Russian and extensive Mordva–Russian and Komi–Russian bilingualism, respectively, in the case of Mari, Udmurt, the Ob-Ugric and Samoyedic languages Russian has become the main contact language and source of bilingualism only beginning with the late 19th or early 20th century. It is not surprising that such long term contact has resulted not only in lexical borrowing but also in contact induced structural effects. There are two significant details to point out about contact with Russian: first, that bilingualism with Russian is asymmetrical, in that while speakers of Uralic languages in Russia are, to various degrees, all bilingual in Russian, incomparably fewer native speakers of Russian are bilingual in one of the Uralic languages, and second, Uralic languages have affected Russian to an incomparably lesser extent than the other way around.

### **3.3. The speakers of the Uralic languages**

see separately!!!

Of the peoples speaking Uralic languages, the following do not have their own nation state: the Saamis of Scandinavia, the Livonians of Latvia, and the Uralic peoples of Russia. The latter live in one of two types of autonomous administrative units, which are geographically identical to the ancient territories of these peoples. The Mordva, the Mari, the Udmurt, the Komi, and the Karelians have their own republics, while the Komi-Permyacs (only until 2005), the Khanty, the Mansi, and the Nenets live in autonomous districts. The former peoples belong to the numerically bigger, the latter to the smaller Uralic peoples. The remaining small Uralic peoples – such as the Selkup or the Nganasan – do not have any kind of

territorial autonomy in Russia. Almost without an exception, the Uralic peoples of Russia live as minorities in the autonomous administrative units where they are the people after whom the administrative unit is named. All the relevant data would be too much to mention here, but the following two examples from 2002 census data are typical. In Karelia, the total population was 716,300 in that year, of whom 76.6% were of Russian nationality, 9.2% Karelians, 2.0% Finns, 0.7% Vepse, and the rest of other nationalities. In the Udmurt Republic the total population was 1,600,000, of whom 60.1% were Russian, 29.3% Udmurt, 6.9% Tatar, and the rest of other nationalities. At the same time, it is important to know that the number of people by a certain nationality is not the same as the number by mother tongue. In all, about 23 million people speak the languages of the Finno-Ugrian language family. Getting a reliable number of mother tongue speakers of the small Uralic languages is difficult for the same reason it is with small or minority populations in general: it depends on people choosing to self-report themselves as mother tongue speakers, and on the proportion of people who really know and use the language. Choosing to self-report as a native speaker and knowing and using that language is not necessarily the same thing. The data below refer to self-reports and not to actual proficiency in and use of the mother tongue. It is also important to note in advance that mother tongue use among Russia's Uralic language mother tongue speakers has declined in the past 20-30 years – steadily among Karelian, Udmurt and Mari speakers, and drastically among Khanty speakers. According to reliable observations, this tendency is typical among Uralic language speakers in Russia. Also in accord with language shift situations elsewhere, it is the older speakers and rural residents who tend to maintain and use their mother tongue. Table 2 below summarises census data from 1989 and 2002, providing a possibility for comparison and for identifying tendencies of change. The Saami languages are not included in this table: there is no data on the Saami populations in the Scandinavian states, Russia or elsewhere, or on the number of mother tongue users. Their total number is estimated at around 20,000. According to the 2002 Russian census, there were 1,991 Kildin-Saami in Russia, 39.5% of them mother tongue users.

Language	1989		2002	
	number of speakers	% mother tongue speakers	number of speakers	% mother tongue speakers
Mordva	1,153,987	67.1	977,381	58.1
Mari	670,868	80.8	678,932	68.4
Udmurt	746,793	69.6	640,028	67.0
Komi	496,579	70.3	434,248	cca. 67.0
Karelian	130,929	47.8	93,344	cca. 50.0
Nenets	34,665	77.1	41,302	cca. 75.0
Khanty	22,521	60.5	28,678	cca. 47.0
Veps	12,501	50.8	8,240	cca. 69.0
Mansi	8,474	37.1	11,432	cca. 23.0

Selkup	3,612	47.6	4,249	cca. 38.0
Nganasan	1,278	83.2	834	cca. 60.0
Ingrian	820	36.8	327	cca. 92.0
Enets	209	45.4	237	cca. 42.0
Votic	very few	very few	73	no data
Total	3,283,236		2,919,305	

Table 2. The number of speakers of Uralic languages in Russia in 1989 and 2002.

The data from 2002 on mother tongue use are of questionable reliability. This is partly due to the fact that questionnaires were modified right before the census had taken place, and partly to a methodological issue of the census: it provided valid and reliable data only for populations over 400,000. For smaller populations, the figures should only be taken as estimates and referring only to the total of speakers of a language in Russia regardless of whether it is their mother tongue or not.

In some cases, for instance in the case of Khanty and Mansi, the number of the population is increasing, while the number of mother tongue users is decreasing. This is partly due to the fact that in these territories, which are rich in oil and natural gas reserves, there is steady industrialisation going on, accompanied by in-migration.

Language maintenance or shift is influenced by the joint work of a number of factors, ranging from economic, political, administrative relations through changes in the ecological, demographic situation to strategies of language policy and changes in ethnic awareness etc. The chances of survival of the Uralic languages in Russia do not look very promising based on the data above, which indicates language shift. It does not look like an unlikely prediction that the number of mother tongue speakers will decrease by half by the end of this century.

### **3.4. The Uralic languages and pluricentricism**

In the strictest sense of the term, pluricentricity among Uralic languages exists only in the case of Hungarian. In addition, speakers of Finnish and Estonian live outside their respective countries. The speakers of Saami languages live in three countries in Scandinavia and on the Kola Peninsula in Russia. Of the Fennic languages, altogether 15 mother tongue speakers of Livonian live in some coastal villages of Northwest Courland and in Latvia (where a few hundred people profess to be of Livonian nationality but do not speak the language).

Of the numerically stronger Uralic languages of Russia, Mordva, Mari and Permian are spoken outside their own republics. Mordva live outside the Mordva Republic, for instance, around Nizhnij-Novgorod, Orenburg and Saratov, in the Chuvash, Tatar and Bashkir Autonomous Republics, in the area between the rivers Volga and Belaja and in smaller groups all the way from Ukraine to Vladivostok and Sakhalin. Only approximately half of the Mari live in their own republic, the other half live in the Bashkir Republic, and in a relatively large pocket near the town of Ufa, speaking an eastern variety of the Meadow Mari dialect. About one third of those professing to be of Udmurt nationality live in neighbouring republics, that is, in Bashkiria, Tatarstan, as well as in the Perm' region and

scattered all over Russia, mainly in cities. About 40-50% of the latter have shifted to Russian. Finally, the Komi live outside the Komi Republic in relatively homogeneous Komi settlements on the Kola Peninsula and in Western Siberia.

#### **4. THE AUXILIARIES OF THE URALIC LANGUAGES**

##### **4.1. Institutions and language planning**

As we could see, only some of the smaller Uralic peoples live in autonomous republics or districts in Russia. For those that do, the use and cultivation of the native languages and cultures is theoretically guaranteed, since language laws were passed in all of these territories in the mid-1990s. However, it is also important to bear in mind that by today all of the smaller Uralic peoples constitute minority populations even in their own republics and autonomous districts, that is, in the places where they are autochthonous, and this fact does not help them assert their own interests or shape institutionalized language policy. This situation is the result of specific historical development. In the late 18th and especially in the first half of the 19th century a whole range of translations of church texts and translations of the Bible (especially of the Gospel according to Matthew) were published in Uralic languages spoken in Russia (e.g. Karelian 1804, Mordva 1821, Livonian 1863 etc.), and this brought about a desire for the introduction of literacy. It was propelled by forceful (and not always peaceful) evangelizing and proselytizing efforts as the result of which the number of converts to Russian Orthodox Christianity visibly increased – at least as far as records attested. In the early 20th century the minority and language policy of the young Soviet state supported the development of literacy by attempting to establish literary norms (in the case of Mordva, Mari, and Komi two norms each were created at the same time), unify orthographies, standardise spelling and grammatical rules, and create norms of language use and terminology. The most immediate goal of the 1920s was publishing school textbooks and newspapers and starting up radio broadcasting. This period brought about a boom for these languages and for the development of minority language literature as well. The 1940s, however, brought a great change: forceful Russianising tendencies started to operate directly against these processes to the extent that the publishing of books and newspapers in minority languages was minimised, minority language education became limited, not to mention the persecution of minority intellectuals which not only decimated but almost completely annihilated them, as in the case of the Udmurt. The prestige of the minority languages decreased drastically as a result. This half a century long series of attacks against minority languages continued until 1989 and had grave consequences for their speakers.

##### **4.2. Linguistic resources**

The earliest sources available on Uralic languages form an integral part with travellers' accounts of these peoples, written by early cartographers mapping up the geography of the enormous Tsarist empire. The linguistic information in these usually consists of bi- or multilingual word lists. Another kind of sources is the enormous amount of amassed material collected in the field by scholars

investigating Uralic peoples and languages in the 19th and 20th century. Their job was not easy in the Siberian environment, and many of them gave their health or even lives for carrying it out. The texts (some of them still waiting to be processed and published) collected by them formed the basis of large scale dictionary and grammar writing projects.<sup>2</sup> Since the 1990s carrying out fieldwork has become much easier, since the places where Uralic peoples live are no longer restricted areas where travel is prohibited.

#### **4.2.1. Conventional resources**

The Finnish Matthias Alexander Castrén (1813-1853) was one of the first pioneers of the study of Uralic languages. He collected language data on more than ten Uralic and a number of non-Uralic languages spoken in Russia and described the circumstances of their speakers. A large part of his life work was published in twelve volumes shortly before his death under the title *Nordische Reisen und Forschungen von Dr. M. Alexander Castrén* ([Nordic travels and researches of Dr. M. Alexander Castrén] 1853–1862), although some of his notes are still unpublished. His Samoyedic materials have been unparalleled to this day (*Samojedische Volksdichtung* [Samoyedic folk poetry] 1940, *Samojedische Sprachmaterialien* [Samoyedic language materials] 1960). Among the early investigators are Russian ethnographer Serafim Patkanov who collected data among the Khanty (*Die Irtysh-Ostjaken und ihre Volkspoesie* [The Irtysh-Ostyaks and their folk poetry] 1897-1900), and the Hungarian Antal Reguly (1819-1858), who carried out work among the Udmurt, the Mansi, the Khanty and the Mari (*Osztyák hősénekek* [Ostyak heroes' songs] 1944, 1951, 1963, 1965). The second half of the 19th century and the early 20th century are the period in Uralic studies when the most extensive data collection was carried out. A large collection of Mansi folk poetry is the result of the work of the Hungarian Bernát Munkácsi (1860-1937; *Vogul népköltési gyűjtemény* [Vogul folk poetry] 1892-1921, 1952, 1963) and of the Finnish Artturi Kannisto (1874-1943; (*Wogulische Volksdichtung* [Vogul folk poetry] 1951, 1955, 1956, 1958, 1959, 1963); their materials were largely published posthumously. The folk poetry of the Komi became known as the result of the work of the Hungarian Dávid Fokos-Fuchs (1884-1977; *Syrjanische Volksdichtung* [Zyrian folk poetry] 1916), that of the Mordva due to the Finnish Heikki Paasonen (1865–1919; *Mordwinische Volksdichtung* [Mordva folk poetry] 1938-1947), that of the Nenets due to the Finnish Toivo Lehtisalo (1887–1962; *Juraksamojedische Volksdichtung* [Yurak-Samoyedic folk poetry] 1947), that of the Saami due to Eliel Lagercrantz (1894-1973; *Lappische Volksdichtung* [Lappish folk poetry] 1957-1966), to mention only a few names. Their multi-volume publications present materials with translations in German or Hungarian, appended with linguistic and other explanations, usually together with highly detailed ethnographic descriptions which are often the only existing sources on the traditional culture, belief systems, one time customs and social conditions

---

<sup>2</sup>Due to space limitations only the most significant works can be mentioned here.

of these peoples. Among the very detailed bi- or trilingual dictionaries comprising information on all the dialects of the given language the following have to be mentioned: Toivo Lehtisalo's *Juraksamojedisches Wörterbuch* [Yurak-Samoyedic dictionary] 1956) for Nenets, Kustaa Fredrik Karjalainen's (1871–1919) *Ostjakisches Wörterbuch* [Ostyak dictionary] 1948) and Wolfgang Steinitz's (1905-1967) *Dialektologisches und etymologisches Wörterbuch der ostjakischen Sprache* ([Dialectological and etymological dictionary of the Ostyak language] 1966-1991) for Khanty, Yrjö Wichmann's (1868–1932) *Wotjakischer Wortschatz* [Votyak vocabulary] 1987) for Udmurt, Dávid Fokos-Fuchs's *Syrjanisches Wörterbuch* ([Zyrian dictionary] 1959) and Ferdinand Johann Wiedemann's (1805-1887) *Syrjänisch-deutsches Wörterbuch* ([Zyrian-German dictionary] 1880/1964) for Komi. All of these dictionaries served as the main starting points for further investigations.

#### **4.2.2. Electronic resources**

Electronic resources for the smaller Uralic languages have only recently started to appear. No larger corpuses are available electronically yet, but smaller amounts of language data (often together with audio recordings), morphological analysis software and sources on various aspects of the Uralic peoples (including their literatures and arts) have become more and more widely available electronically recently, providing – among other things – excellent means of raising these people's ethnic and language awareness.

### **5. PRESENT AND FUTURE ROLE OF THE URALIC LANGUAGES**

Probably the most difficult aspect of the Uralic languages (with the exception of Finnish, Estonian and Hungarian) to predict is the present and future role of these languages. The data in the tables in sections 1.1 and, especially, 3.2 indicate clearly that most Uralic languages, especially those spoken in Siberia, but the smaller Fennic languages as well, are on the brink of extinction. In our age, when languages die at rates never seen before, this might not even seem so shocking. It is, however, important to keep in mind how many factors have been identified by sociolinguistic and language policy investigations as affecting the chances of language maintenance or language shift (inner ones, deriving from the inner state and attitudes of the speech community as well as outer, e.g. economic, political etc. factors). A detailed and comprehensive investigation of these factors is largely lacking for various reasons and especially as far as the small Uralic peoples of Russia are concerned. Some tendencies can, however, be identified.

The languages whose speakers number a few dozens or a couple of hundreds have basically no chance of surviving. This category includes all Saami languages with the exception of Northern Saami, Votic and Ingrian from the Fennic group, as well as the Ob-Ugric and Samoyedic languages. In the case of the latter the most threatening factor is not only their lack of motivation for language maintenance (Samoyedic speakers have no economic or social interest in maintaining their languages) but the fact that it is coupled with a lack of

institutional support of these languages. The assimilative measures of the Soviet Union in the 1930s and 1940s resulted in irreversible processes. The state of public education today exemplifies this: additive bilingualism is clearly not the goal in the system where students study the native language of their ethnic group as a foreign language, in minimal numbers of classes per week. No language policy factors that could change this situation exist today. Some attempts at raising language awareness and a consciousness of ethnic identity as well as attempts at language revitalization have been made, for instance, among the Khanty, with the creation of the ethnographic and language archive as a result of the work of the Hungarian scholar Éva Schmidt in Western Siberia (Beloianski, Khanty-Mansi District). The enormous amount of material gathered there is a collection of sources and traditions of a culture and a language, the result of work that should be completed soon for all the other endangered Uralic languages as well.

The languages with more speakers – such as Northern Saami, and the bigger Finno-Ugric languages such as Mordva, Mari and Permian languages – have a much better chance of survival since they have more chances of asserting their interests. The Northern Saami are in the most promising situation in this respect, since the Norwegian Constitution of 1988 states that the Norwegian state is responsible for the maintenance and development of the language, culture and life-style of the Saami, and the Saami Language Act of 1990 declares that Saami is an official language in six municipalities. In Finland the Saami Language Act of 1990 states that the Saami have the right to use their native language for all government services, and in 2003 Saami became an official language in five municipalities. In Sweden Saami became an officially recognised minority language in 2002. In these three countries the Saami have their own parliaments as well.

The main goal and aim of the speakers of the smaller Uralic languages would definitely be the situation where, as a result of appropriate language policy measures, they would have the institutional right to maintain their native languages and preserve their cultures in circumstances of additive bi- or multilingualism.

## **6. BIBLIOGRAPHY**

- ABONDOLO (Daniel), ed. *The Uralic Languages* (London-New York: Routledge, 1998).
- ALHONIEMI (Alho), *Grammatik der tscheremissischen (Mari) Sprache*. (Hamburg: Buske Verlag, 1993).
- ARISTE (Paul), *A Grammar of the Votic Language* (Bloomington-The Hague-Indiana University: Mouton, 1968).
- AUSTERLITZ (Robert), *Ob-ugric metrics* (Helsinki: Finnish Academy of Sciences, 1958).
- AUSTERLITZ (Robert), „L’ouralien”, in MARTINET (André), ed. *Le Language: Encyclopédie de la Pléiade* (Paris: Gallimard, 1968), pp. 1331-1389.
- BAKRÓ-NAGY (Marianne), „The responsibility of literati”, *Lingua*, 115 (2002), pp. 1053-1062.

- BAKRÓ-NAGY (Sz. Marianne), *Die Sprache des Bärenkultes im Oburischen* (Budapest: Akadémiai Kiadó, 1979).
- BERECZKI (Gábor), *Grundzüge der tscheremissischen Sprachgeschichte* (Szeged: Universitas Szegediensis de Attila József nominata, „Studia uralo-altaica 35”, 1992-1994).
- BERGSLAND (Knut), *The Lapp dialects south of Lappland* (Helsinki: Société Finno-Ougrienne, „Mémoires de la Société Finno-Ougrienne 125”, 1962).
- BUBRIX (Dmitrij V.), *Erzya-mordovskaja grammatika* (Saransk: Mordovian State Publishers, 1947).
- COLLINDER (Björn), *Survey of the Uralic Languages* (Stockholm: Almqvist-Wiksell, 1957).
- COLLINDER (Björn), *Comparative Grammar of the Uralic Languages* (Stockholm: Almqvist-Wiksell, 1960).
- COMRIE (Bernard), *The Languages of the Soviet Union* (Cambridge: Cambridge University Press, 1981).
- CSÚCS (Sándor), *Chrestomathia votiacica* (Budapest: Tankönyvkiadó, 1990).
- DÉCSY (Gyula), *Einführung in die finnisch-ugrische Sprachwissenschaft* (Wiesbaden: Harrassowitz, 1965).
- ERDÉLYI (István), *Selkupisches Wörterverzeichnis* (Budapest: Akadémiai Kiadó, 1969).
- FEOKTISTOV (Aleksandr P.), „Mordovskie jazyki”, in LYTKIN (Vasilij I.) et al., red. *Osnovy finno-ugorskogo jazykoznanja 2.* (Moscow: Nauka, 1975).
- GRUNDSTRÖM (Harald), *Lulelappisches Wörterbuch* (Uppsala: Instituts für Mundarten und Volkskunde, 1946-1954).
- GULYA (János), *Eastern Ostyak Chrestomathy* (Bloomington: Indiana University, 1966).
- HAARMANN (Harald), *Die Finnisch-ugrischen Sprachen: Soziologische und politische Aspekte ihrer Entwicklung* (Hamburg: Buske Verlag, 1974).
- HAJDÚ (Péter), DOMOKOS (Péter), *Die uralischen Sprachen und Literaturen* (Budapest-Hamburg: Akadémiai Kiadó-Buske Verlag, 1987).
- HAJDÚ (Péter), *The Samoyedic Peoples and Languages* (Bloomington-The Hague-Indiana University: Mouton, 1963).
- HAJDÚ (Péter), *Chrestomathia Samoiedica* (Budapest: Tankönyvkiadó, 1968).
- HAJDÚ (Péter), *Introduzione alle lingue uraliche* (Turin: Rosenberg-Seiler, 1990).
- HARMS (Robert T.), „The Uralo-Yukaghir focus system: a problem in remote genetic relationship”, in HOPPER (Paul), ed. *Studies in Descriptive and Historical Linguistics* (Amsterdam: Benjamins, 1977), pp. 301-316.
- HASSELBRINK (Gustav), *Südlappisches Wörterbuch I-III* (Uppsala: Instituts für Mundarten und Volkskunde, 1981-1985).
- HELIMSKI (Eugene), „Two Mator-Taigi-Karagass vocabularies from the 18th century”, *Journal de la Société Finno-Ougrienne*, 81 (1987), pp. 257-267.

- HELIMSKI (Eugen), *Die matorische Sprache*. (Szeged: Universitas Szegediensis de Attila József nominata, „Studia uralo-altaica 41”, 1997).
- HONTI (László), *Chrestomathia ostiatica* (Budapest: Tankönyvkiadó, 1984).
- HONTI (László), „Ugrisches”, *Linguistica Uralica*, 16 (1990), pp. 298-299.
- HULTKRANTZ (Ake), „Swedish Research on the Religion and Folklore of the Lapps”, *The Journal of the Royal Anthropological Institute of Great Britain and Ireland*, 85 (1955), pp. 81-99.
- ITKONEN (Erkki), *Lappische Chrestomathie mit grammatikalischem Abriss und Wörterverzeichnis* (Helsinki, „Apuneuvoja suomalais-ugrilaisten kielen opintoja varten 7”, 1960).
- ITKONEN (Erkki), *Suomalais-ugrilaisen Kielen- ja historian tutkimuksen alalta* (Helsinki: Suomalais Kirjallisuuden Seura, 1961).
- ITKONEN (Erkki), *Inarilappisches Wörterbuch I-IV* (Helsinki: Société Finno-Ougrienne, 1986-1991).
- ITKONEN (Terho I.), *Koltan- ja kuolanlapin sanakirja I-II* (Helsinki: Société Finno-Ougrienne, 1958).
- JANHUNEN (Juha), *Samojedischer Wortschatz. Gemeinsamojedische Etymologien* (Helsinki: Castrenianum, 1977).
- JANHUNEN (Juha), „Uralic languages”, in BRIGHT (W.), ed. *International Encyclopedia of Linguistics IV* (New York-Oxford: Oxford University Press), pp. 205-210.
- JANHUNEN (Juha), „On the structure of Proto-Uralic”, *Finnisch-Ugrische Forschungen*, 4, (1981), pp. 23-42.
- JANHUNEN (Juha) „Uralilaisen kantakielen sanastosta”, *Journal de la Société Finno-Ougrienne*, 77 (1981), 9. pp. 219-274.
- JOKI (Aulis), *Uralier und Indogermanen* (Helsinki: Société Finno-Ougrienne, „Mémoires de la Société Finno-ougrienne 151”, 1973).
- KÁLMÁN (Béla), *Vogul Chrestomathy* (Bloomington-The Hague: Indiana University-Mouton, 1965).
- KÁLMÁN (Béla), *Wogulische Texte mit Glossar* (Budapest: Akadémiai Kiadó, 1976).
- KATZ (Hartmut), *Generative Phonologie und phonologische Sprachbünde des Ostjakischen und Samojedischen* (München: Wilhelm Fink, 1975).
- KATZ (Hartmut), *Selcupica I-IV* (München: Finnisch-Ugrisches Seminar der Universität München, 1975-1988).
- KERESZTES (László), *Chrestomathia Morduinica* (Budapest: Tankönyvkiadó, 1987).
- KORENCHY (Éva), *Iranische Lehnwörter in den obugrischen Sprachen* (Budapest: Akadémiai Kiadó, 1972).
- KORHONEN (Mikko), *Johdatus lapin kielen historiaan* (Helsinki: Suomalaisen Kirjallisuuden Seura, 1981).
- KÜNNAP (Ago), *Enets* (München: Lincom Europa, 1999).

- KÜNNAP (Ago), *Kamass* (München: Lincom Europa, 1999).
- LAAKSO (Johanna), *Our Otherness. Finno-Ugrian Approaches to Women's Studies, or vice versa* (Wien: Finno-Ugrian Studies in Austria, „Schriftenreihe für die Forschungsbereiche der Abteilung Finno-Ugristik“, 2005).
- LAANEST (Arvo), *Einführung in die ostseefinnische Sprachen* (Hamburg: Buske Verlag, 1982).
- LAGERCANTZ (Eliel), *Lappischer Wortschatz I-II* (Helsinki: Société Finno-Ougrienne, 1939).
- LALLUKKA (Seppo), *The East Fennic Minorities in the Soviet Union* (Helsinki: Finnish Academy of Sciences, 1990).
- LALLUKKA (Seppo), „Finno-Ugrians of Russia: Vanishing Cultural Communities?“, *Nationalities Papers*, 29 (2001), pp. 9-39.
- LEHTIRANTA (Juhani), *Yhteisaamelainen sanasto*. (Helsinki: Société Finno-Ougrienne, „Mémoires de la Société Finno-Ougriennes 200“, 1989).
- MAITINSKAJA (Klara E.), *Istoriko-sopostavitel'naia morfologia Finno-ugorskikh yazikov* (Moscow: Nauka, 1979).
- MIKOLA (Tibor), *Studien zur Geschichte der samjedischen Sprachen* (Szeged: SzTE Finnisch-Ugrisches Institut, „Studia uralo-altaica 45“, 2004).
- MOSELEY (Christopher), *Livonian* (München: Lincom Europa, 2006).
- MUNKÁCSI (Bernát), KÁLMÁN (Béla), *Wogulisches Wörterbuch* (Budapest: Akadémiai Kiadó, 1986).
- MUNKÁCSI (Bernát), FUCHS (David R.), *1952 Volksbräuche und Volksdichtung der Wotjaken* (Helsinki: Société Finno-Ougrienne, „Mémoires de la Société Finno-Ougrienne 102“, 1952).
- NIELSEN (Konrad), *Lappisk ordbok-Lapp Dictionary I-IV* (Oslo: Instituttet for sammenlignende kulturforskning, 1933-1962).
- PENTIKÄINEN (Juha), ed. *Shamanism and Northern Ecology* (Berlin-New York: Mouton, 1996).
- PIROTTI (Giuliano), *Grammatica vogula* (Mantu: Associazione culturale Italo-Ungherese Taddeo Ugoletto da Parma, 1972).
- RÉDEI (Károly), *Die syrjänischen Lehnwörter im Wogulischen* (Budapest-Bloomington-The Hague: Akadémiai Kiadó-Bloomington-The Hague, 1970).
- RÉDEI (Károly), *Syrjänische Chrestomathie* (Wien: Finno-Ugric Institute of the University of Vienna, 1978).
- RÉDEI (Károly), *Uralisches etymologisches Wörterbuch I-III* (Wiesbaden-Budapest: Harrassowitz Verlag-Akadémiai Kiadó, 1988-1991).
- RIESE (Timothy), *Vogul (Mansi)* (München: Lincom Europa, 2003).
- ROMBANDEJEVA (Evdokija I.), *Mansijskij (vogulskij) jazyk*. (Moscow: Nauka, 1973).

- SALMINEN (Tapani), „Classification of the Uralic languages”, in GRÜNTAL (Riho), PENTTINEN (Sirpa), SALMINEN (Tapani), eds. *IFUSCO 1988. Proceedings of the Fifth International Finno-Ugrist Students' Conference* (Helsinki: Snellman-Instituutti, 1989).
- SALMINEN (Tapani), ed. *Uralilaiset kielet tänään* (Kuopio: Snellman-Instituutti, 1993).
- SAMMALLAHTI (Pekka), „A linguist looks at Saami Prehistory”, *Acta Borealia*, 2 (1989), pp. 3-11.
- SEBEOK (Thomas A.), INGEMANN (Frances), *An Eastern Cheremis Manual* (Bloomington: Indiana University, 1961).
- SINOR (Denis), ed. *The Uralic Languages. Description, History and Foreign Influences* (Leiden: Brill, 1988).
- SOKOLOVA (Soja), *Das Land Jugorien* (Moskau-Leipzig: Verlag Progress-VEB F. A. Brockhaus Verlag, 1976).
- SKRIBNIK (Elena), *Bibliographia Sibirica. Finno-Ugristics and Samoyedology in scientific centres of Siberia* (Savariae, „Specimina Sibirica IX”, 1994).
- SKRIBNIK (Elena), „The Research into Siberian Languages and Educational Programs for Students Representatives of Siberian Indigenous Peoples at the Novosibirsk University”, in *Peuples des Grands Nords: Traditions et Transitions* (Paris: Presses de la Sorbonne Nouvelle, INALCO, 1995), pp. 57-78.
- STEINITZ (Wolfgang), *Ostjakologische Arbeiten* (Budapest-Berlin: Akadémiai Kiadó-Akademie Verlag, 1980).
- TAAGEPERA (Rein), *The Finno-Ugric Republics and the Russian State*. (London-New York: Hurst and Routledge, 1999).
- TAAGEPERA (Rein), „Eastern Finno-Ugric cooperation and foreign relations”, *Nationalities Papers* 29 (2001), pp. 181-199.
- TEREŠČENKO (Natalia M.), „Enetskij jazyk” in LYTKIN (Vasilij I.), et al., red. *Jazyki narodov SSSR III. Finno-ugorskie i samodijskie jazyki* (Moscow: Nauka, 1966), pp. 438-457.
- TSIGANKIN (Dmitrij V.), red., *Grammatika mordvskikh jazykov* (Saransk: Mordovian State University, 1980).
- VIKÁR (László), BERECSKI (Gábor), *Votyak Folksongs* (Budapest: Akadémiai Kiadó, 1989).
- VUORELA (Toivo), *Finno-Ugric Peoples* (Bloomington: Indiana University, 1997).
- WAGNER-NAGY (Beáta), ed. *Chrestomathia Nganasanica* (Szeged-Budapest: SzTE Finnugor Tanszék-MTA Nyelvtudományi Intézet, „Studia uraloaltaica, Supplementum 10”, 2002).
- WINSA (Birger), ed. *Finno-ugric People in the Nordic Countries. Roots V: The Roots of Peoples and Languages of Northern Eurasia* (Hedenäset, 2005).