

TOWARDS A PHONOLOGICAL TYPOLOGY OF URALIC LANGUAGES

**Karl Pajusalu¹, Kristel Uihoaed¹, Péter Pomozi²,
Endre Németh³, and Tibor Fehér³**

¹University of Tartu, ²Eötvös Loránd University, Budapest, and

³Institute of Forensic Medicine, Budapest

Abstract. The paper focuses on phonological similarities between Uralic languages. The study is based on a dataset which includes 33 word-prosodic and segmental features of 28 Uralic languages or main dialects, including all traditional subgroups of the language family. In statistical analysis clustering and dimension reduction techniques such as multidimensional scaling are applied. This methodology enables to explore distinctive subgroups of languages as well as calculate distances between languages and language groups. As a result we present a quantitative phonological typology. The main division appears between western and central-eastern phonological types of the Uralic languages. The detected phonological subgroups coincide with the traditional ones, i.e. Finnic, Saami, Mordvin, Mari, Permian, Hungarian, Ob-Ugric and Samoyedic. The Hungarian subgroup (Standard Hungarian, Csángó Hungarian) and the Ob-Ugric subgroup (Northern Mansi, Eastern Mansi, Northern Khanty, Eastern Khanty) are internally stable. However, their interrelation and relationship with other groups is ambiguous; according to our results, Hungarian is typologically closer to the Western Uralic language groups (i.e. Finnic and Saami), whereas Ob-Ugric languages form a distinct branch of Central-Eastern Uralic. In general, the results reveal a significant influence of multiple areal connections on the phonological formation of Uralic languages.

Keywords: Uralic languages, Proto-Uralic, phonology, typology, word prosody, language areals

DOI: <https://doi.org/10.12697/jeful.2018.9.1.08>

1. Introduction

The Uralic phonological typology has so far been a rather understudied field in linguistics. Phonology of the Uralic languages has traditionally been studied from the historical point of view, i.e. the aim of the phonological comparison of Uralic languages has been reconstruction of the Proto-Uralic sound system (e.g. Sammallahti 1988). Despite their

shared genetic origin, the Uralic languages are typologically diverse (Comrie 1988). Spreading over a huge territory, the languages have been in long-term and complex contact situations with neighboring languages from other language families (e.g. Finnic languages in the Circum-Baltic language area and Ob-Ugric languages in the Western-Siberian language area, see Helimski 2003, Laakso 2013) which has resulted in the emergence of typologically new traits. Therefore an areal-typological approach, as applied, for example, to the study of suprasegmental phonology of Circum-Baltic languages (e.g. Koptjevskaja-Tamm and Wälchli 2001: 637–646) is relevant in explaining the phonological typology of Uralic languages.

The starting point of this study is synchronic and quantitative. 33 prosodic and segmental features of 28 Uralic languages or main dialects, including all groups and sub-groups of the language family, are synchronically identified and statistically analyzed using clustering and dimension reduction techniques. As a result we present a quantitative phonological typology of Uralic languages. The results of the analyses provide us a basis for defining main phonological areas of the Uralic language family.

The dataset of the studied phonological features was compiled by using relevant descriptions of the Uralic languages (e.g. Sinor 1988, Abondolo 1998, and a number of studies on certain languages and phonological phenomena, see Section 2). The occurrence of phonological features were also checked by experts of various Uralic languages¹.

In our article we'll give at first an overview of studied languages and their phonological descriptions, then we describe the phonological features and after that we introduce the applied research methods and results. Finally, we discuss the results of the statistical analyses and characterize main phonological types of the Uralic languages.

2. An overview of the studied languages and their descriptions

Our dataset contains phonological data of all main branches of Uralic languages: Finnic, Saami, Mordvin, Mari, Permic, Ugric (Ob-Ugric and Hungarian), and Samoyed. To improve premises of quantitative

¹ We are especially grateful to Prof. Riho Grünthal for consulting us with Finnic, Saami and Mordvin data, Prof. Juha Janhunen for checking Samoyed data and Prof. Emer. Tiit-Rein Viitso for giving advice in various research questions.

analyses we included at least two languages in each group or sub-group (if a group was divided into sub-groups). Our studied languages are as follows:

2.1. Finnic languages. We involved into the dataset seven Finnic languages considering the primary divisions of the language groups: three northern Finnic languages – (Standard) Finnish, Karelian (proper or Northern Karelian), and Veps (Central dialect); four southern Finnic languages: (Standard) Estonian, Võro (South Estonian), Votic and (Standard) Livonian (based on eastern Courland Livonian). Historical overviews of Finnic phonology are presented by Viitso (1998) and Laakso (2001); Pajusalu (2012) observes phonological innovations of southern Finnic languages, Kallio (2016) historical phonology from Proto-Finnic to Proto-Livonian. Suomi et al. (2008) explain the Finnish sound structure, Asu et al. (2016) provide a broad treatment of the Estonian sound system. All Finnic languages are phonologically relatively close, e.g. they reveal primary word-initial stress, trochaic stress patterning; the distinction between short and long vowels and consonants is phonological in most languages. However, there are some significant differences between northern and southern Finnic sound systems (Pajusalu 2012).

2.2. Saami languages. The dataset includes two western Saami languages – Ume Saami and North Saami, and three eastern Saami languages – Inari Saami, Skolt Saami and Kildin Saami. The development of Saami phonological structures is treated by Korhonen (1981) and Sammallahti (1998a, 1998b). There are also available a number of studies on phonology of different Saami languages, such as Itkonen (1984), Bye et al. (2009), Türk et al. (2016) on Inari Saami, McRobbie-Utasi (1999, 2007) on Skolt Saami, and Sammallahti (2012) on Ume Saami. The main phonological system of Saami languages is mostly similar to Finnic but they have even more complex durational distinctions (Korhonen 1981).

2.3. Mordvin languages. From the central area of Uralic languages both Erzya and Moksha are included to the dataset. The phonological traits of Erzya are characterized by Lehiste et al. (2003) and Aasmäe (2006); an overview of Moksha phonology is presented by Aasmäe et al. (2013). The sound structures of Standard Erzya and Moksha are described by Devaev and Cygankin (1970). The sound structure of Mordvin languages and dialects is largely variable; they lack as a rule phonological distinction of segmental duration; vowel duration can be used as a signal of word stress (Lehiste et al. 2003).

2.4. Mari languages. Meadow Mari and Hill Mari are representatives of this language group in the dataset. A general characterization of Mari sound system is presented by Alhoniemi (1985) and Bereczki (1990); the phonological features of Meadow Mari are described by Lehiste et al. (2005), the Hill Mari sound system by Vikström and Zorina (2007). The Mari languages do not have duration oppositions of vowels, consonant geminates occur only at morpheme boundaries; they show extensive variation in their accentuation systems which is determined by various phonological and morphological factors (Rozhanskiy 2013).

2.5. Permic languages. Udmurt and Komi-Zyryan are selected from the Permic group. Overviews of Permic phonological structures are given by Riese (1998) and Geisler (2005). Winkler (2011) presents the principal traits of Udmurt phonology and Hausenberg (1998) introduces the Komi sound system. There are no duration distinctions in Permic languages but they differ largely in their prosodic structure (Riese 1998).

2.6. The Hungarian sub-group is represented by Standard Hungarian and eastern, or Csángó Hungarian. Siptár and Törkenczy (2000) introduce Standard Hungarian phonology. The phonological peculiarities of Csángó are observed in studies on Hungarian dialectology (e.g. Kiss 2001). The phonological systems of Standard Hungarian and Csángó are mostly similar, they have fixed word-initial stress, round vowel harmony, etc.; at the same time they reveal some phonological differences in their phonemic systems, for example, unlike Standard Hungarian, Csángó has diphthongs.

2.7. The Ob-Ugric languages are represented by two main varieties of Mansi and Khanty: Northern Mansi, Eastern Mansi, Northern Khanty and Eastern Khanty. Honti (1998) characterized the general structure of Ob-Ugric; Keresztes (1998) gives an overview of the Mansi sound system, Nikolaeva (1999) and Filchenko (2007) present the main traits of Khanty phonology. Among Ob-Ugric languages, Mansi varieties are phonologically closer to Hungarian than Khanty. In Khanty varieties word stress is often related to phrase-level rhythmic patterns (Filchenko 2007).

2.8. The Samoyed branch of the Uralic languages is divided into northern and southern groups. In our dataset representatives of Northern Samoyed are Tundra Nenets and Nganasan, the Southern Samoyed languages are Selkup (Northern dialect) and Kamas. Main characteristics of Samoyed phonology are introduced by Janhunen (1998), and the word-prosodic developments are observed by Helimski (1978). The sound structure of Tundra Nenets is described by Salminen (1998,

2007), Nganasan by Helimski (1998a) and Várnai (2012), Selkup by Helimski (1998b), and Kamas by Simoncsics (1998). The Samoyed languages show extensive variation in their sound systems and prosody. Especially in northern Samoyed languages word stress is not always prominent and some morphophonological factors can affect stress location (Rozhanskiy 2013).

Thus, the database includes data from all main groups of the Uralic languages. On the level of sub-groups the classification of the Uralic languages have remained several open questions because of the extensive variation of languages spoken in huge disconnected territories. Therefore it is justified to treat some Uralic languages as a sub-group of languages and their main dialects as languages, for example, in the case of Mansi and Khanty. In our study, such tentative grouping of these languages is not an obstacle for applied quantitative study, quite the contrary – all the closely related languages or main dialects are explored as distinct entities in statistical analyses where they are likely grouped together. A sufficiently detailed and balanced composition of the dataset is essential for the output of quantitative research.

3. Overview of Uralic phonology and selected features

The studied dataset contains phonological features which have probably categorical values, it means that it is possible to estimate their occurrence or non-occurrence in a certain language. Such features are, for example, phonologically contrastive vowel length, consonant clusters in syllable onset or occurrence of a close rounded vowel *ü* [IPA: y]. Only the existence of a feature in the basic vocabulary of a language has been taken into account, marginal occurrences in loanwords have not been considered. For example, if in a studied language consonant clusters occur only in syllable onsets of loanwords, it is not regarded as an occurrence of the feature. We have tried to select the independent features which are not interrelated, but in some cases the results show still frequent co-occurrences. For example, non-initial heavy syllables tend to attract word stress in such Uralic languages which do not have as a rule trochaic feet. Such outcomes are part of the research results and they have been treated in further steps of the study.

The selection of phonological features for this study has been influenced by outcomes of a larger project of Finno-Ugric prosody (see Lehiste and Pajusalu 2010); that is why our dataset contains most

systematically prosodic and phonotactic features (14 features from 33). Concerning segmental traits we observe 13 features of consonant systems and 6 features of vowels.

3.1. Prosodic features

Our data set includes properties of word stress, temporal and tonal alternations. Besides these, general constraints of syllable and foot structures and vowel harmony are investigated. Thus, the main analyzed prosodic features are as follows:

The studied properties of word stress are lexical significance (is there lexical stress?), position (is there non-initial stress?) and relationship between stress placement and syllable structure (i.e. stress in heavy syllables). Prominent lexical stress is characteristic for Finnic, Saami, Hungarian, Mari, Ob-Ugric, also for Moksha and Udmurt. Non-initial stress that is untypical for the most Uralic languages appears in Erzya, Udmurt, and variously in Mari and Samoyed language groups. Heavy syllables attract stress in Meadow Mari, Komi(-Permyak), Northern Mansi and Nganasan.

To study temporal contrasts we observe duration of consonants and vowels. Contrastive length of consonants occurs in Finnic, Saami, Hungarian and Northern Selkup. Contrastive length of vowels occurs in Finnic, excluding Veps, Saami, Hungarian, Mansi, and Northern Selkup. In unstressed syllables distinction of short and long vowels is developed in Finnish, Karelian, some Votic varieties and Hungarian.

Tonal contrasts are not widespread in the Uralic languages. They appear only in some Southern Finnic languages: as contrast of broken and plain tone in Livonian, and in relation to quantity degrees in Estonian and Võro.

Concerning syllable structures limited occurrence of consonant clusters in syllable onset or coda is typical of many Uralic languages. In syllable onset, consonant clusters are common only in Southern and Eastern Finnic languages, i.e. in Veps, Estonian, Livonian, and Võro. They are also adapted for Mordvin and Mari languages. In syllable coda consonant clusters are much more common: they occur in Finnic, Saami, Mordvin, Mari, Hungarian and Ob-Ugric languages.

Some prosodic constraints apply to the whole foot structure. Alternation of long and overlong quantity degrees that is characteristic of Southern Finnic (Estonian, Võro, Livonian) and Saami languages is related as a rule to special foot types (see Markus et al. 2014). Trochaic

feet are typical of many Uralic languages: Finnic, Saami, Mordvin, Hungarian, Ob-Ugric, and some Samoyed languages (Tundra Nenets). Consonant gradation is also a feature which is constrained by foot structures. It appears in most Finnic languages (Finnish, Karelian, Estonian, Votic, Võro), Saami, and some Samoyed languages, such as Nganasan and Selkup.

3.2. Vowel system

General properties of vowel systems are occurrence of diphthongs, distinction between reduced and full vowels and qualitative difference of short and long vowels. Monosyllabic diphthongs occur in Finnic, Saami, Mansi and Csángó Hungarian. Distinction between reduced and full vowels has phonological significance in Mordvin, Mari, Permian, Ob-Ugric and Samoyed. The qualitative difference between some short and long vowels appears in Hungarian.

There are also several differences in the phonemic composition of vowel systems. Unrounded central or back vowels (e.g. Võro *y* [IPA: u], Estonian *õ* [IPA: ɤ]) occur in Southern Finnic (Estonian, Votic, Livonian, Võro), Mordvin, Eastern Saami (Inari, Skolt, Kildin Saami), Mari, Permian, Eastern Mansi and Khanty, and Csángó Hungarian. Rounded front vowels (e.g. Estonian *ü* [IPA: y], *ö* [IPA: ø]) occur in Finnic (excl. Livonian), Inari Saami, Mari, Hungarian, Eastern Mansi and Khanty and in some Samoyed languages (Nganasan, Selkup). Distinction of open front *ä* [IPA: æ] and mid front *e* lacks in North and Kildin Saami, Erzya, Meadow Mari, Permian, Hungarian and Mansi.

Front-back vowel harmony occurs both in western and eastern Uralic languages, such as Finnish, Karelian, Votic, Võro, the Saami languages, Hill Mari, Hungarian, limitedly in Mordvin, Ob-Ugric and in Samoyed language groups. Round vowel harmony occurs only in Hungarian and Meadow Mari.

3.3. Consonant system

A number of features distinguish consonant systems of Uralic languages. Affricates are common in most Uralic languages but they lack in Finnish, Estonian, and some varieties of Mansi and Khanty. Retroflex consonants occur only in Komi, Ob-Ugric, and Nganasan. Palatal consonants occur in Mari, Hungarian, Moksha, Komi, Nganasan

and Kamas. Palatalized consonants are widespread in Finnic, excluding Finnish, Inari and Kildin Saami, Mordvin, Mari, Permic, Ob-Ugric, and several Samoyed languages. Glottal stop [IPA: ʔ] occurs in some Finnic languages (Finnish, Karelian, Võro) and in Samoyed (excluding Selkup). Glottal or velar fricatives [IPA: h, x] are spread in most modern Uralic languages, they lack only in Livonian, Mari varieties, Northern Mansi and Khanty. Voiceless labiodental fricative *f* occur in Saami languages, Moksha, Hungarian and Kamas.

Eastern Khanty distinguish between voiceless velar and uvular stops (*k* vs. *q*), Mansi varieties between velar plain and labialized stops (*k* vs. *kʷ*). Voiced stops occur in some Finnic (Karelian, Veps, Votic, Livonian) and Saami (Ume, Skolt, Kildin Saami) languages, in Mordvin, Mari, Permic, Hungarian and several Samoyed languages (Tundra Nenets, Northern Selkup, Kamas). Distinction between alveolar and palate-alveolar sibilants (*s* vs. *š* [IPA ʃ]) appears in some peripheral Finnic languages (Karelian, Veps, Livonian), in Saami, Mordvin, Mari and Permic languages, in Standard Hungarian, Northern Mansi and Kamas.

Word-initial *r* does not appear in the core vocabulary of Udmurt and Samoyed languages. Velar nasal *ŋ* is a phoneme in several Finnic languages (Finnish, Karelian, Veps), in Saami, Mari, Ob-Ugric and Samoyed languages.

4. Statistical research methods

In statistical analysis we used clustering (Everitt et al. 2011) and dimension reduction techniques, such as multidimensional scaling (Kruskal and Wish 1978). This methodology enabled us to explore distinctive sub-groups of languages as well as calculate distances between languages and language groups.

We applied hierarchical agglomerative bootstrap clustering on our data in order to detect language groups, which could be formed based on observed linguistic features. Hierarchical method is more suitable for detailed data analysis and on smaller data set as the present one. Agglomerative clustering means that the tree is grown from the leaves to roots, i.e. the method begins with the assumption that each item (language in our case) forms a separate group. It continues with merging most similar objects into a group, and similarity here means objects for which the distance is the smallest. This process is continued until the objects form one single tree. We used binary distance in order

to calculate distances between observed languages, because our data consists only of information which indicates presence or absence of linguistic feature in a given language. The applied clustering algorithm is Ward method, which tries to minimize the increase in the variance in the distances between the members of clusters. This method usually produces compact clusters.

The higher the merge has happened, the more different objects or groups there are. For example, most different are two main groups as they are merged higher than other objects. Most similar are Ume Saami and Skolt Saami as they are merged on the lowest level.

Bootstrap clustering takes iteratively subsamples of the data and the final result is produced as the summary of the iterations. Bootstrapping means that many subsamples of the initial data are generated with replacement, i.e. some data is removed from the initial dataset and these values are replaced with random values while the size of the data remains the same with the initial dataset. This approach also enables to estimate the certainty of clusters. More iterations are always better although it can be computationally demanding. We performed 10,000 iterations on our data, i.e. the data was sampled 10,000 times so that on every iteration some data was removed and replaced with new random data; clustering was performed every time on such subsample. Each group has been assigned a *p*-value (calculated over 10,000 iterations), which is the indication how strongly an observed cluster is supported by the data. The *pvc* package (Suzuki and Shimodaira 2015) provides two values: Approximately Unbiased (au) and Bootstrap Probability (bp) *p*-values. The AU *p*-value is considered to be a better approximation to an unbiased *p*-value than BP, which is normally calculated by bootstrap resampling.

5. Results

The outcome of multidimensional scaling shows some expected and some surprising relations, see Figure 1. At first, the closeness of Finnic and Saami languages is predictable but also Csángó and Hungarian are located close to them. Ob-Ugric languages differ significantly from Hungarian and all others. The third main group contains Mordvin, Mari, Permic and Samoyed languages. Coordinate values on axes are arbitrary and have no exact meanings. The method only presents relationships between observed items. Two-dimensional figure visualizes two most

explanatory dimensions of high-dimensional data. The first dimension runs along horizontal and the second along vertical axis.

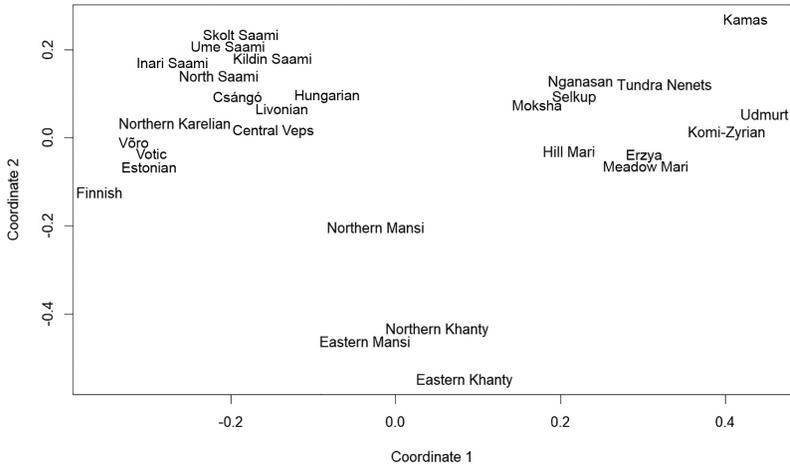


Figure 1. Multidimensional scaling of Uralic languages

According to these results we can synchronically define a western Uralic area which consists of Finnic, Saami and also Hungarian languages, secondly there is an area of Ob-Ugric languages, and thirdly a large central-eastern area of other Uralic language groups. In the western area the Finno-Ugric languages have been historically in intense contacts with Germanic and Slavic languages, for the eastern groups the contacts with Turkic and several Western Siberian languages have had consequences. In comparison to Proto-Uralic the western group is the most innovative one (17 innovations in Hungarian, 16 innovations in Northern and Ume Saami, 14 innovations in Livonian) and Ob-Ugric and Samoyed languages are the most conservative (7 innovations in Eastern Khanty, 8 in Northern Selkup, 9 in Northern Khanty, Tundra Nenets and Kamas).

The results of the binary clustering presented in Figure 2 are mostly in line with outcomes of multidimensional scaling. Approximately Unbiased (au) values present the percentage of how many times the group was clustered together during the iteration process of all 10,000 bootstrapped samples. Because of the binary grouping method, the two main groups are the western and the central-eastern group. Ob-Ugric languages belong as a distinct sub-group to the central-eastern group. In the western group Hungarian and Saami languages constitute stable

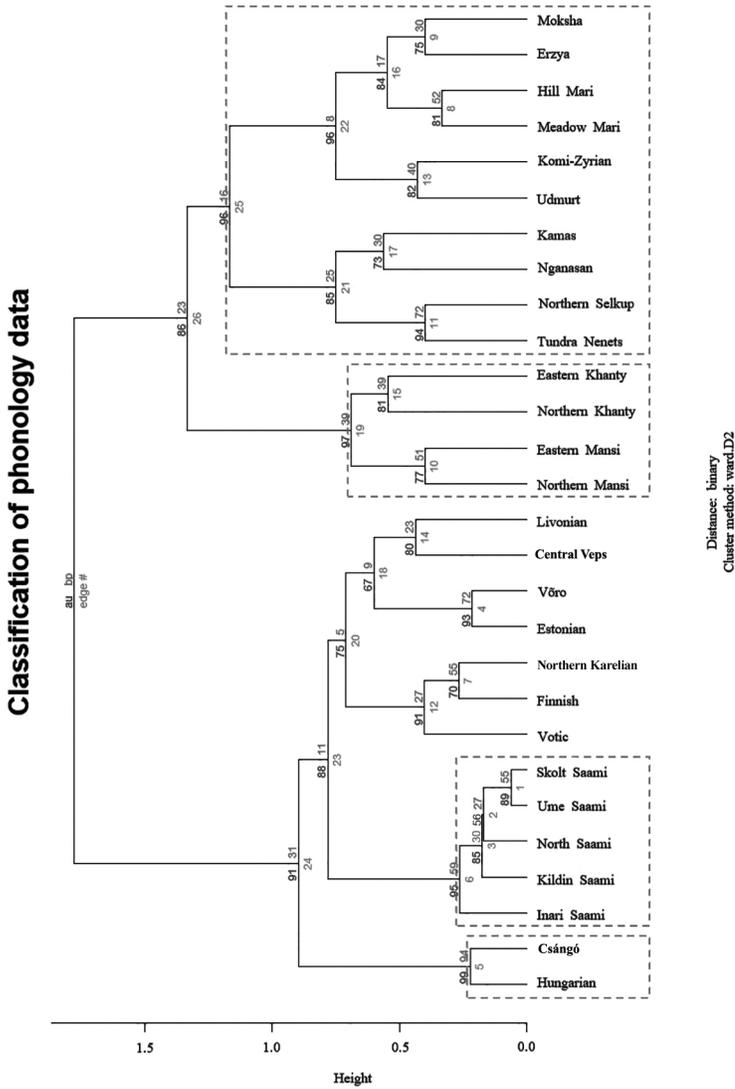


Figure 2. Hierarchical agglomerative bootstrap clustering

sub-groups but the Finnic sub-group is more diverse and peripheral Veps and Livonian are grouped together. The expected sub-groups are Finnish-Karelian and Estonian-Võro. In the central-eastern group besides Ob-Ugric the second principal sub-group is Samoyed, thirdly Permic languages diverse from Mari and Mordvin languages. In the Samoyed group Tundra Nenets and Northern Selkup form one sub-

group and Nganasan and Kamas another sub-group. It is noticeable that in some cases peripheral languages from remote areas of bigger languages groups are linked into same sub-groups, such as Ume and Skolt Saami, Veps and Livonian, or Nganasan and Kamas. However, the principal language groups are established in line with traditional grouping of Uralic languages. Only the position of Hungarian would seem to be unexpected but areal-typological factors can explain it.

6. Discussion

These statistical results give a ground to define two primary typological areas of the Uralic languages – the western and the central-eastern. We try to give below a firsthand characterization of their distinctive features.

6.1. Special features of the western Uralic area

Prosodic features are most essential in defining the western Uralic phonological area. They are contrastive length of consonants and vowels, lexical stress which is not dependent on syllable weight, dynamic word-initial syllable stress, trochaic feet and complex quantity alternations in prosodic feet. An important common feature of the western Uralic group is the formation of extra-long syllables where a long nucleus can be followed by complex coda.

Western Uralic languages usually do not reveal phonologically significant differences between full and reduced vowels. In most cases they have diphthongs (cf. Csángó vs. Standard Hungarian) and in some cases long vowels occur in a weak (unstressed) part of a prosodic foot. These languages distinguish as a rule between an open front vowel *ä* [IPA: æ], mid front vowel *e* and open back vowel *a*, and they lack central or back unrounded vowels (excluding some Saami and Southern Finnic languages, Csángó). The western Uralic languages have consonant clusters in syllable coda and the glottal fricative *h* (excluding Livonian), but the labiodental fricative *f* and palatal consonants are typical only of Saami and Hungarian.

A number of these features are similar to Germanic and southern Slavic languages (cf. Helimski 2003) but there are also a lot of peculiarities which are possibly connected to common developments of old Uralic traits.

6.2. Special features of the central-eastern Uralic area

The phonology of central-eastern Uralic languages differs from the western group in many respects. In these languages there is typically no contrastive length of consonants and vowels, they do not have consonant clusters in syllable coda and complex quantity alternations. Vowel sequences tend to be disyllabic and vowel duration or quality can be a cue of stress. The central-eastern Uralic languages have often rhythmic (phrasal) stress, such as in Erzya or Tundra Nenets.

These languages reveal often a difference between full and reduced vowels and they have central or back unrounded vowels [ɤ, ʉ]; in several cases they do not distinguish between open and mid front vowel, as it is in Erzya, Meadow Mari, Permic, and Mansi. In most cases these languages have palatalized consonants and distinguish between *s* vs. *š* (excluding Samoyed).

The most typical languages of the central-eastern type are Mari and Permic languages. The Samoyed and Ob-Ugric languages have a few features of the western type. This is explainable with the peripheral location of these languages in the Uralic language area. Figure 3 illustrates the relationship of Uralic languages according to the biggest commonalities between the main Uralic language groups (numbers show common features between language groups).

Figure 3 gives a figurative generalization of the main results of our statistical analyses. In addition to the division between western and central-eastern phonological areas it is important to bring out the difference between central and lateral positions of certain language groups (cf. Korhonen 1996, Helimski 2003: 162). Within the huge territory of Uralic languages the lateral language groups are Finnic, Saami, Hungarian, Ob-Ugric and Samoyed, and the central language groups are Mordvin, Mari and Permic (taking into account also syntactic features, Mordvin has earlier been treated as a controversial case which does not belong neither to core nor to lateral Uralic).

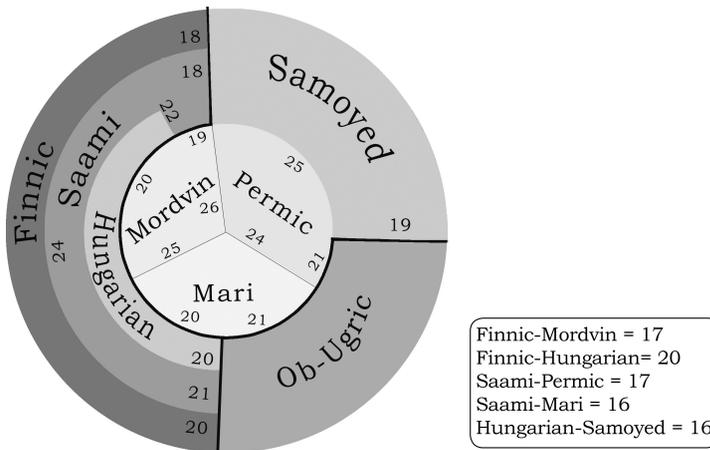


Figure 3. Phonological distances between the branches of the Uralic languages

The position of Hungarian is most complex. According to our study of phonological-typological features Hungarian has even more common traits with Saami than with Ob-Ugric and an equal number of shared features with Finnic, Mordvin and Mari. Therefore, Hungarian is not a typical peripheral member of the typological area but it is in an intermediate position between the western and central-eastern language groups. Ob-Ugric has also a special position among the Uralic languages as our previous analyses indicated.

Western and eastern lateral groups share some phonological features, such as consonant gradation in Finnic, Saami and Samoyed languages. However, central language groups share typically more common features with each other than lateral groups.

7. Conclusion

Our study is an attempt to identify typological groupings of Uralic languages and their stable members on the basis of phonological-typological features. In most cases the detected phonological groups and sub-groups coincide with the traditional ones, i.e. Finnic, Saami, Mordvin, Mari, Permic, Hungarian, Ob-Ugric and Samoyedic. Two main phonological types of the languages were defined as the western and the central-eastern ones. Finnic, Saami, and Hungarian belong typologically to the western type but they all can be treated as lateral groups in some respects.

The central-eastern typological area contains Ob-Ugric, Mordvin, Mari, Permic and Samoyed languages. Ob-Ugric and Samoyed groups are in a lateral position. The interrelationship between Permic and Samoyed groups is noteworthy. Unlike Samoyed languages the Permic group does not share significant phonological features with the western group.

These results are obviously caused by long-term and complex contacts, on the one hand with other language families (e.g. Saami, Finnic and Hungarian languages had contacts in the north-western and central European area and eastern Uralic languages in the western Siberian language area, cf. Helimski 2003) which have resulted in the emergence of typologically new traits. On the other hand, in the lateral areas some typologically rare features have preserved better than in the central areas, and contact influences have often been adapted in periphery and centre in different ways. Our results reveal a significant influence of multiple areal connections – also between different Uralic language groups – on the phonological formation of all Uralic languages.

Acknowledgements

This study was partly supported by the Estonian Research Council project IUT2-37.

Addresses:

Karl Pajusalu
Institute of Estonian and General Linguistics
University of Tartu
Jakobi 2
51014 Tartu, Estonia
E-mail: Karl.Pajusalu@ut.ee

Péter Pomozi
Institute of Hungarian and Finno-Ugric Linguistics
Eötvös Loránd University
Múzeum körút 4
1088 Budapest
E-mail: pomozipeter@btk.elte.hu

Kristel Uiboaed
E-mail: kristel.uiboaed@ut.ee

Endre Németh

E-mail: endre.nemeth@gmail.com

Tibor Fehér

E-mail: jafetie@gmail.com

References

- Aasmäe, Niina (2006) *Stress and quantity in Erzya*. (Dissertationes Philologicae Uralicae Universitatis Tartuenssis, 6.) Tartu: University of Tartu Press.
- Aasmäe, Niina, Pärtel Lippus, Karl Pajusalu, Nele Salveste, Tatjana Zirnask, and Tiit-Rein Viitso (2013) *Moksha prosody*. (Mémoires de la Société Finno-Ougrienne, 268.) Helsinki: Finno-Ugrian Society.
- Abondolo, Daniel, ed. (1998) *The Uralic languages*. London and New York: Routledge.
- Alhoniemi, Alho (1985) *Marin kieliooppi*. Helsinki: Suomalais-Ugrilainen Seura.
- Asu, Eva Liina, Pärtel Lippus, Karl Pajusalu, and Pire Teras (2016) *Eesti keele hääldus*. Tartu: University of Tartu Press.
- Berezcki, Gábor (1990) *Chrestomathia Ceremissica*. Budapest: Tankönyvkiadó.
- Bye, Patrick, Elin Sagulin, and Ida Toivonen (2009) “Phonetic Duration, phonological quantity and prosodic structure in Inari Saami”. *Phonetica* 66:4, 199–221.
- Comrie, Bernard (1988) “General features of the Uralic languages”. In Denis Sinor, ed. *The Uralic languages: description, history, and foreign influences*, 451–477. Leiden: Brill.
- Devaev, S. Z., and Dmitrij V. Cygankin (1970) *Fonetika mordovskix (mokšanskogo i erzjanskogo) literaturnyx jazykov*. Saransk: Mordovskoe knižnoe izdatel'stvo.
- Everitt, Brian S., Sabine Landau, Morven Leese, and Daniel Stahl (2011) *Cluster analysis*. 5th rev. ed. Chichester: Wiley-Blackwell.
- Filchenko, Andrey Yu. (2007). *A Grammar of Eastern Khanty*. Rice University, unpublished PhD dissertation.
- Geisler, Michael (2005) *Vokal-Null-Alternation: Synkope und Akzent in den permischen Sprachen*. (Veröffentlichungen der Societas Uralo-Altaica, 68.) Wiesbaden: Harrassowitz.
- Hausenberg, Anu-Reet (1998) “Komi.” In Daniel Abondolo, ed. *The Uralic languages*, 305–326. London and New York: Routledge.
- Helimski, Eugene (1978) “Notes on the origin of prosodic features in some Samoyed and Ugric languages”. *Estonian papers in phonetics*, 35–38. Tallinn: Institute of Language and Literature.
- Helimski, Eugene (1998a) “Nganasan” In Daniel Abondolo, ed. *The Uralic languages*, 480–515. London and New York: Routledge.
- Helimski, Eugene (1998b) “Selkup.” In Daniel Abondolo, ed. *The Uralic languages*, 548–579. London and New York: Routledge.

- Helimski, Eugene (2003) "Areal groupings (Sprachbünde) within and across the borders of the Uralic language family: a survey". *Nyelvtudományi Közlemények* 100, 156–167.
- Honti, László (1998) "Ob-Ugrian". In Daniel Abondolo, ed. *The Uralic languages*, 327–357. London and New York: Routledge.
- Itkonen, Erkki (1984) "Ehdotus inarilapin fonemaattiseksi transkriptioksi." In Erkki Itkonen, Terhi Itkonen, Mikko Korhonen, and Pekka Sammallahti, eds. *Lapin murteiden fonologiaa*, 43–68. (Castrenianumin toimitteita, 1.) Helsinki: Helsingin Yliopisto.
- Janhunen, Juha (1998) "Samoyedic." In Daniel Abondolo, ed. *The Uralic languages*, 457–479. London and New York: Routledge.
- Kallio, Petri (2016) "Historical phonology from Proto-Finnic to Proto-Livonian." In Valts Ernštreits and Karl Pajusalu, eds. *Studies on Livonian II*, 39–65. (Journal of Estonian and Finno-Ugric Linguistics 7, 1.) Tartu: Tartu Ülikool.
- Keresztes, László (1998) "Mansi". In Daniel Abondolo, ed. *The Uralic languages*, 387–427. London and New York: Routledge.
- Kiss, Jenő (2001) *Magyar dialektológia*. Budapest: Osiris Kiadó.
- Korhonen, Mikko (1981) *Johdatus lapin kielen historiaan*. Helsinki: Suomalaisen Kirjallisuuden Seura.
- Korhonen, Mikko (1996) *Typological and historical studies in language*. (Mémoires de la Société Finno-Ougrienne, 223.) Helsinki: Finno-Ugrian Society.
- Koptjevskaja-Tamm, Maria, and Bernhard Wälchli (2001) "The Circum-Baltic languages: an areal-typological approach." In Östen Dahl and Maria Koptjevskaja-Tamm, eds. *The Circum-Baltic languages*. Vol. 2: *Grammar and typology*, 615–750. Amsterdam and Philadelphia: John Benjamins.
- Kruskal, Joseph B. and Myron Wish (1978) *Multidimensional scaling*. (Sage University Paper Series on Quantitative Applications in the Social Sciences, 07–011.) Newbury Park: Sage Publications.
- Laakso, Johanna (2001) "The Finnic languages." In Östen Dahl and Maria Koptjevskaja-Tamm, eds. *The Circum-Baltic languages*. Vol. 1: *Past and present*, 179–212. Amsterdam and Philadelphia: John Benjamins.
- Laakso, Johanna (2013) "Contact and the Finno-Ugric Languages". In Raymond Hickey, ed. *The handbook of language contact*, 598–617. Chichester: Blackwell Publishing.
- Lehiste, Ilse, Niina Aasmäe, Einar Meister, Karl Pajusalu, Pire Teras, and Tiit-Rein Viitso (2003) *Erzya prosody*. (Mémoires de la Société Finno-Ougrienne, 245.) Helsinki: Finno-Ugrian Society.
- Lehiste, Ilse, Pire Teras, Toomas Help, Pärtel Lippus, Einar Meister, Karl Pajusalu, and Tiit-Rein Viitso (2005) *Meadow Mari prosody*. (Linguistica Uralica Supplementary Series, 2.) Tallinn: Estonian Academy Publishers.
- Lehiste, Ilse and Karl Pajusalu (2010) "Experimental study of prosody in Finno-Ugric languages". In Sándor Csúcs, ed. *Congressus XI Internationalis Fenno-Ugristarum, Piliscsaba 2010*, Pars I, 225–245. Piliscsaba: Reguly Társaság.
- Markus, Elena, Pärtel Lippus, Karl Pajusalu, and Pire Teras (2013) "Three-way opposition of consonant quantity in Finnic and Saamic languages." In Eva Liina Asu and

- Pärtel Lippus, eds. *Nordic Prosody. Proceedings of the XIth Conference, Tartu 2012*, 225–234. Frankfurt am Main: Peter Lang.
- McRobbie-Utasi, Zita (1999) *Quantity in the Skolt (Lappish) Saami language: an acoustic analysis*. (Indiana University Uralic and Altaic Series, 165.) Bloomington: Indiana University.
- McRobbie-Utasi, Zita (2007) “The instability of systems with ternary length distinctions: the Skolt Saami evidence.” In Ida Toivonen and Diane Nelson, eds. *Saami Linguistics*, 167–206. Amsterdam and Philadelphia: John Benjamins.
- Nikolaeva, Irina (1999) *Ostyak*. (Languages of the World/Materials, 305.) München: Lincom Europa.
- Pajusalu, Karl (2012) “Phonological innovations of the Southern Finnic languages”. In Riho Grünthal and Petri Kallio, eds. *Linguistic map of prehistoric North Europe*, 201–224. (Mémoires de la Société Finno-Ougrienne, 266.) Helsinki: Finno-Ugrian Society.
- Riese, Timothy (1998) “Permian”. In Daniel Abondolo, ed. *The Uralic languages*, 249–275. London and New York: Routledge.
- Rozhanskiy, Fedor (2013) “Morphophonological nature of Mari accentuation as viewed from the Uralic perspective”. *Linguistica Uralica* 49, 184–207.
- Salminen, Tapani (1998) “Nenets”. In Daniel Abondolo, ed. *The Uralic languages*, 327–357. London and New York: Routledge.
- Salminen, Tapani (2007) “Notes on Tundra Nenets phonology.” In Jussi Ylikoski and Ante Aikio, eds. *Sámit, sánit, sátnehámit. Riepmočála Pekka Sammallahtii miessemánu 21. beaivve 2007*, 349–372. (Mémoires de la Société Finno-Ougrienne, 253.) Helsinki: Finno-Ugrian Society.
- Sammallahti, Pekka (1988) “Historical phonology of the Uralic languages with special reference to Samoyed, Ugric, and Permic”, in Denis Sinor, ed. *The Uralic languages: Description, history, and foreign influences*, 478–554. Leiden: Brill.
- Sammallahti, Pekka (1998a) *The Saami languages: an introduction*. Kárášjohka: Davvi Girji.
- Sammallahti, Pekka (1998b) “Saamic”. In Daniel Abondolo, ed. *The Uralic languages*, 43–95. London and New York: Routledge.
- Sammallahti, Pekka (2012) “Remarks on Ume Saami phonology”. In Eberhard Winkler, Hans-Hermann Bartens, and Cornelius Hasselblatt, eds. *Laponicae investigationes et uralicae. Festschrift zum 65. Geburtstag von Lars-Gunnar Larsson*, 177–201. (Veröffentlichungen der Societas Uralo-Altaica, 82.) Wiesbaden: Harrassowitz.
- Simoncsics, Péter (1998) “Kamassian.” In Daniel Abondolo, ed. *The Uralic languages*, 580–601. London and New York: Routledge.
- Sinor, Denis, ed. (1988) *The Uralic languages: description, history, and foreign influences*. Leiden: Brill.
- Siptár, Péter, and Törkenczy, Miklós (2000) *The phonology of Hungarian*. Oxford: Oxford University Press.
- Suomi, Kari, Juhani Toivanen, and Riikka Ylitalo (2008) *Finnish sound structure: phonetics, phonology, phonotactics and prosody*. (Studia Humaniora Ouluensia, 9.) Oulu: University of Oulu.

- Suzuki, Ryota, and Hidetoshi Shimodaira (2015) *pvclust: Hierarchical Clustering with P-Values via Multiscale Bootstrap Resampling*. R package version 2.0-0. Available online at <<https://CRAN.R-project.org/package=pvclust>>. Accessed on 15.02.2018.
- Türk, Helen, Pärtel Lippus, Karl Pajusalu, and Pire Teras (2016) “Quantity contrast in Inari Saami: the role of pitch and intensity.” *Speech Prosody 2016*, 1090–1094.
- Várnai, Zsuzsa (2012) “Hangtan”. In Beáta Wagner-Nagy, ed. *Chrestomathia Nganasanica*. Szeged, Budapest: SZTE Finnugor Tanszék, MTA Nyelvtudományi Intézet.
- Viitso, Tiit-Rein (1998) “Fennic”. In Daniel Abondolo, ed. *The Uralic languages*, 96–114. London and New York: Routledge.
- Vikström, O.V. and Zoja G. Zorina (2007) *Zvukovoj stroj sovremennogo gornomarijskogo jazyka*. Joškar-Ola: Izd-vo Marijsk. gosud. universiteta.
- Winkler, Eberhard (2011) *Udmurtische Grammatik*. (Veröffentlichungen der Societas Uralo-Altaica, 81.) Wiesbaden: Harrassowitz.

Аннотация. Карл Паюсалу, Кристел Уйбоад, Петер Помози, Эндре Немет и Тибор Фехер: К фонологической типологии уральских языков. В статье рассматриваются фонологические сходства между уральскими языками. Исследование основывается на выборке данных, которая включает 33 просодических и сегментных признака, которые применяются к 28 уральским языкам или их основным диалектам. Эта выборка покрывает все традиционно выделяемые группы уральской семьи. Для статистической обработки данных используются методы кластерного анализа и многомерного шкалирования. Такой подход позволяет исследовать выявленные подгруппы языков и рассчитывать расстояния между языками и языковыми группами. Результатом анализа является количественная фонологическая типология. Основная граница разделяет западный и центрально-восточный типы уральских языков. Выявленные фонологические подгруппы языков совпадают с традиционными: прибалтийско-финская, саамская, мордовская, марийская, пермская, венгерская, обско-угорская и самодийская. Венгерская подгруппа (стандартный венгерский и чангошский диалект) и обско-угорская подгруппа (северный мансийский, восточный мансийский, северный хантыйский, восточный хантыйский) демонстрируют внутреннюю стабильность. Однако их собственные взаимоотношения и отношение к другим подгруппам неоднозначны: наши результаты показывают, что венгерский язык типологически ближе к западно-уральской группе (т. е. прибалтийско-финским и саамским языкам), тогда как обско-угорские языки оказываются четко выделяемой ветвью центрально-восточных уральских языков. В целом, результаты раскрывают существенное влияние разнообразных ареальных связей на становление фонологии уральских языков.

Ключевые слова: уральские языки, прауральский язык, фонология, типология, просодия слова, лингвистические ареалы

Kokkuvõte. Karl Pajusalu, Kristel Uiboaed, Péter Pomozi, Endre Németh ja Tibor Fehér: Uurali keelte fonoloogilisest tüpoloogiast. Artikkel keskendub uurali keelte fonoloogiliste sarnasuste võrdlemisele. Analüüsitav andmestik hõlmab 33 sõnaprosoodilist ja segmentaalset tunnusjoont 28-st uurali keelest või põhimurdest, mille hulgas on keeli kõigist traditsioonilistest allrühmadest. Statistilises uuringus rakendatakse klasteranalüüsi ja multidimensionaalsest skaleerimist. Nii eristatakse olulised allrühmad ja arvutatakse kaugused keelte ja keelerühmade vahel, saades uurali keelte kvantitatiivse fonoloogilise tüpoloogia. Esmane lahknemine ilmneb geograafiliselt läänepoolsete ning kesk- ja idapoolsete uurali keelte vahel. Nende sees tulevad esile traditsioonilised allrühmad: läänemeresoome, saami, mordva, mari, permi, ungari, obiugri ja samojeedi. Ungari (ungari kirjakeel, csángó) ja obiugri (põhjamansi, idamansi, põhjahandi, idahandi) on sisemiselt stabiilsed, kuid suhe nende kahe rühma vahel on ambivalentne. Meie tulemuste põhjal on ungari tüpoloogiliselt lähedasem läänepoolsetele uurali keelele (st läänemeresoomele ja saamile), obiugri keeled moodustavad aga eraldi allrühma kesk- ja idauurali keelte hulgas. Saadud tulemused näitavad areaalsete kontaktide tähtsust uurali keelte fonoloogilise eripära kujunemises.

Märksõnad: uurali keeled, proto-uurali, fonoloogia, tüpoloogia, sõnaprosoodia, keeleareaalid

Appendix I: studied languages

FINNIC:	Estonian, Finnish, Northern Karelian, Livonian, Central Veps, Votic, Võro (South Estonian);
SAAMI:	Inari Saami, Kildin Saami, North Saami, Skolt Saami, Ume Saami;
MORDVIN:	Erzya, Moksha;
MARI:	Hill Mari, Meadow Mari;
PERMIC:	Komi-Zyryan, Udmurt;
HUNGARIAN:	Hungarian, Csángó;
OB-UGRIC:	Northern Mansi, Eastern Mansi, Northern Khanty, Eastern Khanty;
SAMOYED:	Kamas, Nganasan, Northern Selkup, Tundra Nenets.

Appendix II: studied features

WORD PROSODY AND PHONOTACTICS: lexical stress, non-initial stress, stress in heavy syllables; contrastive length of consonants, contrastive length of vowels, length opposition in unstressed syllables; tonal contrast; distinction of long and overlong quantities; consonant clusters in syllable onset; consonant clusters in syllable coda; trochaic feet; consonant gradation;

VOWELS: diphthongs, reduced vowels, qualitative difference of short and long vowels, unrounded central or back vowels, rounded front vowels, distinction of open and mid front vowels; front-back vowel harmony, round vowel harmony;

CONSONANTS: affricates, retroflex consonants, palatal consonants, palatalized consonants, glottal stop, glottal or velar fricatives, voiceless labiodental fricative; distinction between voiceless velar and uvular stops, distinction between velar plain and labialized stops, voiced stops, distinction between alveolar and palate-alveolar sibilants, word-initial trill, velar nasal.