# FINNISCH-UGRISCHE FORSCHUNGEN

# Inhalt

## Band 68

# Besprechungen

# Berichte und Nekrologe

*Ekaterina Georgieva*
HUN-REN Hungarian Research Centre for Linguistics

# On adverbial clauses in Udmurt: postpositional phrases and the case of the adverbial case

This paper presents three types of non-finite adverbial clauses in Udmurt: the ones encoded with the suffixes *-(e)mja*, *-(o)nja*, and *-(o)ńńa*. I propose that these suffixes should be decomposed morphologically and that these non-finite adverbial clauses are to be analyzed as postpositional phrases. In this way, the paper contributes to the analysis of non-finite adverbial subordination in Udmurt. Moreover, the description of *-(o)ńńa*-clauses in the Middle Cheptsa dialect, which have not been previously described in the literature, also deepens our knowledge of Udmurt dialectal syntax. Additionally, this study has implications for our understanding of the Udmurt case system, as it makes a novel proposal regarding the adverbial case in Udmurt.

## 1. Introduction

In this paper I discuss the Udmurt non-finite adverbial clauses formed with the suffixes *-(e)mja*, *-(o)nja*, and *-(o)ńńa*. These clause types are illustrated in (1)–(3). (For the time being, a morphological decomposition of these suffixes is not provided, as this will be one of the main questions addressed in this paper.)

(1)     Udmurt Corpus (*Udmurt duńńe*, 2010.04.21)

     *No* [*mon tod-emja*], *soļị siźem koncert-jos vań.*
     but 1SG know-*emja* 3SG.DAT dedicated concert-PL COP
     'But as far as I know, there are concerts dedicated to him/her.'

(2)     (Winkler 2011: 53)

     [*Oźị mịn-onja-z*] *metro-je vu-i-z.*
     this.way go-*onja*-POSS.3SG metro-ILL arrive-PST-3SG
     'As he went like this, he arrived at the metro (station).'

(3)     fieldwork recording, 2014_08_11, TS, Balezino district, Udmurtia

     *Muš-jos-mị pegǯi-ľľam=ńi=no*
     bee-PL-POSS.1PL escape-EVID.3PL=already=ADD

     [*baba-jeṇị-mị vịr-ońńa-mị*].
     grandmother-INS-POSS.1PL be.busy-*ońńa*-POSS.1PL
     'Our bees had (long) flown away while we/me and our grandmother were busy (taking care of the chicks).'

### 1.1. Previous research

The clauses in (1)–(3) have not received much attention in descriptive studies of Udmurt, but some observations are made in Fokos-Fuchs (1958), Edygarova (2010), and Winkler (2011). All three examples feature a non-finite clause, one that is encoded with the suffixes *-(e)m* or *-(o)n*, which are independently attested in the language, plus the adverbial case *-ja*. The combination of the non-finite suffix and the case suffix is generally not treated as a converb suffix. The adverbial case occupies a special place in the Udmurt case system, as it is the only case that can either precede or follow the possessive suffixes. This property of the adverbial case is also observed with the clauses under consideration: for instance, in (2), the

possessive morphology comes after the case suffix. The morpheme order in -*(e)mja*- and -*(o)nja*-clauses has been discussed by Edygarova (2010), who argues that it depends on the function of the adverbial clause.

The suffix -*(o)ńńa* is used in the Middle Cheptsa dialect and is considered to differ only morphophonologically from -*(o)nja*. Together with the Upper Cheptsa and the Lower Cheptsa dialects, the Middle Cheptsa dialect constitutes the group of Northern Udmurt dialects (see Kel'makov 1998 on the dialectal division of Udmurt and Karpova 2005 for a general description of the Middle Cheptsa dialect). The Middle Cheptsa dialect is spoken in five municipalities of Udmurtia: Glazov, Yukamensk, Yarsk, Balezino, and Krasnogorsk. To the best of my knowledge, -*(o)ńńa*-clauses have not been studied in detail in the previous literature. Beserman Udmurt, which is also spoken in the northern part of Udmurtia, utilizes a similar clause type, formed with the suffix -*(o)ńńiga* (see Usacheva & Serdobolskaya 2015; forthcoming).

## 1.2. Preview of the proposed analysis

The paper makes a contribution from both empirical and theoretical perspectives. On the empirical level, I show that there are two types of -*(e)mja*- and -*(o)nja*-clauses. The first type has a temporal meaning ('while') and can only feature an event nominalization. Here possessive morphology follows the adverbial case suffix. The second type has an oblique meaning ('according to' or 'based on') and must contain a non-event nominalization. In this second type, possessive morphology precedes the adverbial case. As for -*(o)ńńa*-clauses, I show that they are not simply temporal adjuncts but have a locative meaning as well. I also draw a comparison with Beserman Udmurt -*(o)ńńiga*-clauses.

The theoretical analysis of these empirical findings relies on the understanding of postpositional phrases in generative syntactic terms. Under the proposed analysis, the clauses in question are postpositional phrases (PPs). PPs can be headed by adpositions or semantic cases.[1] The Udmurt adverbial case, being a semantic case, is also a P head. I will argue that we

---

1.  In the literature a distinction is made between *abstract cases*, i.e. those expressing grammatical relations like subject or object, and *semantic cases*, i.e. those encoding semantic roles, such as spatial relations (Blake 1994; on the terminology used for the two classes of cases see Haspelmath 2009).

need to distinguish between two different types of PPs with the adverbial case, and thus we can account for the two types of *-(e)mja-* and *-(o)nja-* clauses. In other words, the proposal is that there are in fact two adverbial cases in present-day Udmurt. Moreover, I will argue that these two adverbial cases are diachronically related.

I argue that the locative meaning of *-(o)ńńa-*clauses comes from the so-called DOMUS suffix *-ń*, a spatial suffix independently attested in the Middle Cheptsa dialect, combined with the exponent *-a* that marks the inessive or illative case in the possessive declension. In light of the PP-analysis proposed, this means that we are dealing with a complex PP, and the meaning of these clauses is compositionally derived from the subparts of this complex PP. This proposal also implies that *-(o)ńńa* is not a morphophonological variant of *-(o)nja* because it does not feature the adverbial case (*pace* the standard analysis).

The PP-analysis that I put forward for *-(e)mja-*, *-(o)nja-*, and *-(o)ńńa-* clauses agrees with the existing descriptive studies that the three suffixes should be decomposed morphologically. Thus, we are not dealing with converb suffixes. However, it supersedes the previous analyses because it does not only derive the meaning of these clauses in a compositional way but also explains their morphosyntax. Specifically, it accounts for the morpheme order of the adverbial case and the possessive suffixes.

## 1.3. Data

The data used in this paper come from various sources. I provide examples from my own fieldwork conducted between 2013 and 2016. These examples are listed as follows: *fieldwork recording*, *date of recording*, *(filename)*, *speaker's initials*, *collection point*. I also use examples obtained from elicitation tasks (listed as *elicited*); those provide crucial negative evidence. The Middle Cheptsa data presented in Sections 3.2 and 4.3 were collected in Spring 2023 from two Udmurt dominant speakers of the Middle Cheptsa dialect, who were born and raised in the village of Isak (Russian: Исаково), Balezino district (this village marks the eastern border of the Middle Cheptsa dialect according to Karpova 2005: 16). The data were obtained through elicitation sessions, which targeted the meaning of the *-(o)ńńa-*clauses, including their temporal interpretation, as well as the possibility of having locative adverbials in them and how this affects the intended meaning of the clause. Furthermore, I also use corpus data from

the Udmurt Corpus, the Udmurt Social Media Corpus, and the Turku–Izhevsk Corpus, as well as from other descriptive studies.[2]

## 1.4. Structure of the paper

This paper is organized as follows: In Section 2, I provide the relevant background information on how non-finite adverbial subordination is expressed in Udmurt. I also summarize the main functions of the Udmurt adverbial case. This will be relevant in order to compare the three clause types in question to the functions of the adverbial case in general. In Section 3, I present the new empirical findings regarding the *-(e)mja-*, *-(o)nja-*, and *-(o)ńńa*-clauses. In Section 4, I provide a theoretical account in a generative syntactic framework (the relevant theoretical assumptions are summarized in the beginning of this section). In Section 5, I offer conclusions.

## 2. Background

This section first gives a general background on non-finite adverbial subordination in Udmurt. It then provides an overview of how the adverbial case is used in Udmurt.

## 2.1. Non-finite adverbial clauses in Udmurt

Descriptively, non-finite adverbial clauses in Udmurt can be encoded in two ways (Winkler 2011: 110–121, 173–175; Georgieva 2018: Ch. 3). One option is to use a non-finite clause selected by a postposition or a semantic

---

2. The Udmurt Corpus, which is available online at http://udmurt.web-corpora. net, currently contains 9.57 million words of mostly newspaper texts published between 2007 and 2018; these texts represent standard Udmurt. The Udmurt Social Media Corpus is available online at http://udmurt.web-corpora.net and it contains 2.66 million words; it features texts coming from open posts and comments by Udmurt-speaking vKontakte users (up to February 2018). The Turku–Izhevsk Corpus, which is available at http://volga.utu.fi/portal/cgi-bin/ login.cgi, contains approx. 11,000 texts from newspapers published between 1997 and 2002. The searches were carried out in May–June 2018 and February–July 2023. In some, but not all cases, the results were manually disambiguated. The source of each example is listed next to it. The English translations are mine throughout the paper. The glossing and/or transcription of examples from other sources was slightly modified for consistency.

case. The other possibility is by using a converb.[3] The former strategy is illustrated with *ber-e* [back-ILL] 'after (temporal, causal)' in (4) and with *aź-in* [front-INE] 'before' in (5). Other postpositions that select for a non-finite clause are *vil-iś* [top-ELA] 'because', *dir-ja* [time-ADV] 'during', *inti-je* [place-ILL] 'instead', etc.

(4)  Turku–Izhevsk Corpus (Vordskem kyl/I/4.txt:110)

*žitaže,*　　　　[*sobrani-len*　　*ortć-em-ez*]
in.the.evening　　convention-GEN　be.over-NMLZ-POSS.3SG

*bere, kino lu-o-z.*
back-ILL movie be-FUT-3SG

'In the evening, after the convention is over, there will be a movie.'

(5)  Udmurt Corpus (*Udmurt duńńe*, 2013.06.14)

[*Viľ už bordi kutsk-on*] *aź-in　　niriś*
new work to start-VN front-INE first

*vań-ze*　　　　　　　　*radjale,*　　　　*ćotale,*
everything-POSS.3SG.ACC　organize.IMP.2PL　count.IMP.2PL

*mertale.*
measure.IMP.2PL

'Before starting a new project, first consider every detail
(lit. organize, count, and measure everything).'

The non-finite clauses selected by these postpositions are formed with the suffixes *-(e)m* and *-(o)n*. These nominalizations have a very wide distribution: they occur as non-finite relative and argument clauses.[4] In addition, they can be selected by postpositions or semantic cases, and as a result, can be used as adverbial clauses. In what follows the suffixes *-(e)m* and *-(o)n* used in non-finite adverbial clauses will be glossed as NMLZ and VN, respectively. The different glosses, which are adopted from earlier studies,

---

3.  The terms *gerund*, *verbal adverb*, and the Russian *деепричастие* are used for *converb* in the descriptive literature (see Fokos-Fuchs 1958; Perevoshchikov 1962: 255–283; Kel'makov & Hännikäinen 1999: 206–209, 213–216, 218–219, 224–233; Bartens 2000: 228–265; Winkler 2001: 56–61; 2011: 110–121). For a detailed description of converb clauses see Perevoshchikov (1959), Perevoshchikov (1962: 269–293), and Georgieva (2018: Ch. 3).

4.  The question of whether non-finite relative and argument clauses can or should be unified has sparked debate in the literature on Udmurt (see Georgieva 2018: 46–68 for an overview and Dékány & Georgieva 2020 for a theoretical analysis). In this paper I focus on the adverbial clauses with *-(e)m* and *-(o)n*.

are meant to indicate that *-(o)n*-nominalizations have more nominal properties than *-(e)m*-nominalizations (see Serdobolskaya et al. 2012; Georgieva 2018), although these differences will not play a role in the description of the adverbial clauses under consideration. What will be important is the distinction between event and non-event nominalizations for both *-(e)m* and *-(o)n* (see Section 3.1).

The literature agrees that the same nominalizations, *-(e)m* and *-(o)n*, are found in the suffixes *-(e)mja*, *-(o)nja*, and *-(o)ńńa*, which were illustrated in (1)–(3). In these clauses, the nominalization is said to combine with the adverbial case *-ja* (Fokos-Fuchs 1958; Edygarova 2010; Winkler 2011). The two nominalizations can also be selected by other semantic cases, e.g. by the instrumental and the elative, the former being illustrated in (6). This adverbial clause functions as a cause/reason clause.

(6)    Turku–Izhevsk Corpus (Kenesh/D/5:783)

*Tolon*     [*kuaź*     *zor-em-en*]     *busɨ-je*     *ɛz*
yesterday   weather    rain-NMLZ-INS    field-ILL    NEG.PST.3

*vetle=no* […]
go.CNG.PL=ADD

'Yesterday they did not go to the field because it was raining
(lit. with the weather raining).'

The non-finite form *-(e)men* is listed as a converb, i.e. a non-finite form used to express adverbial subordination, in some grammars of Udmurt (Kel'makov & Hännikäinen 1999; Winkler 2001; 2011). This means that instead of decomposing it morphologically and treating it as a case-marked form of the *-(e)m*-nominalization, a separate converb suffix *-(e)men* is postulated. The criteria for distinguishing converbs in Udmurt are discussed by Fokos-Fuchs (1958).[5] He argues that converb suffixes are not simply a combination of a nominalization and a case suffix, because the converb suffix is no longer segmentable and/or is semantically opaque. Thus, his main criteria are related to the morphological segmentability and semantic transparency. In his view, the non-finites *-(e)men* in (6) are a borderline case: they are segmentable, but their meaning is not transparent. He argues that *-(e)men*-clauses are translated into German with *als* '(causal)

---

5.  For typological definitions of the notion of converbs, see the contributions in Haspelmath & König (1995). Converbs in Uralic have been extensively discussed within a typological framework in Ylikoski (2003).

since' rather than with *mit* 'with', which he takes to be indicative of the grammaticalization of this suffix into a converb (Fokos-Fuchs 1958: 287; Winkler 2011: 115–116 shares this opinion, but without discussing these clauses in detail). In contrast, in my previous work, I have argued that all segmentable "converbs" are in fact case-marked nominalizations and that the semantic transparency can be derived from morphological transparency: morphologically segmentable suffixes are semantically transparent, and vice versa; this applies to *-(e)men*-clauses, too (see Georgieva 2018: Ch. 3 for extensive discussion).

The parallel with *-(e)men*-clauses is relevant, as the *-(e)mja-*, *-(o)nja-*, and *-(o)ńńa*-clauses are also built on the two nominalizations in combination with a semantic case. Importantly, as already stated, the three suffixes under investigation are not analyzed as converbs in the descriptive literature (Fokos-Fuchs 1958; Edygarova 2010; Winkler 2011; Georgieva 2018). One piece of support in favor of this comes from the fact that these suffixes are morphologically segmentable. Fokos-Fuchs (1958) mentions an additional argument that concerns morpheme order. The *-(e)mja-*, *-(o)nja-*, and *-(o)ńńa*-clauses can show possessive morphology, as shown in (2). In this example, the possessive morphology comes after the nominalization suffix and the case suffix. The possessive morphology can also precede the adverbial case, as in (7). In Fokos-Fuchs's view, the fact that both morpheme orders are attested indicates that the combination of the nominalization and the adverbial case should not be treated as a single, fully grammaticalized converb suffix.

(7)　　Udmurt Corpus (*Udmurt duńńe*, 2008.05.28)

| *Mi* | *um* | *ľukiśke* | *aďami-jez* |
|---|---|---|---|
| 1PL.EXCL | NEG.1PL | separate.PRS.CNG.PL | person-ACC |

| *viʒi̯-jez-ja,* | *osk-on-ez-ja* […] |
|---|---|
| root-POSS.3SG-ADV | believe-VN-POSS.3SG-ADV |

'We don't separate people based on their origin and/or religion.'

Furthermore, Winkler (2011: 116) mentions these clauses *passim* and suggests that their meaning can be compositionally derived from the non-finite suffixes: *-(e)m* or *-(o)n* plus the adverbial case. The example (2) is listed in the section dealing with the adverbial case; this also suggests that Winkler considers the non-finite verb form to be morphologically decomposable.

These empirical findings discussed in the earlier studies already highlight the main issues that will be addressed in the present paper: the order

of the adverbial case and the possessive suffix in these adverbial clauses, i.e. the difference between (2) and (7), as well as the semantics of these clauses, i.e. how their meaning relates to that of the adverbial case. New empirical findings regarding these issues will be presented in Section 3. However, in order to investigate these clause types in greater detail, we first need to get acquainted with the main properties of the adverbial case. This will serve as a baseline for the discussion in Section 3.

## 2.2. The Udmurt adverbial case

The descriptive studies distinguish between two functions of the adverbial case in Udmurt: (i) it derives adnominal modifiers, as in (8); (ii) it encodes adverbial modifiers: with the meaning 'according to' or 'based on', as in (9a, b), but a purely spatial meaning 'along' is also possible, albeit rarely mentioned in the literature (9c) (Perevoshchikov 1962: 100–101; Kel'makov & Hännikäinen 1999: 188; Bartens 2000: 89, 103; Winkler 2001: 24; 2011: 53; Edygarova 2017).

(8)     (Winkler 2001: 24)

     *udmurt   kı̣l-ja        dišetiś*
     Udmurt   language-ADV  teacher
     'teacher of Udmurt'

(9) a. *Plan-ja    uža-j.*
     plan-ADV   work-PST.1SG
     'I worked according to the plan.' (Georgieva 2018: 81)

   b. *Diśkut-ez-ja          todma-j.*
     clothes-POSS.3SG-ADV  recognize-PST.1SG
     'I recognized [him/her] based on his/her clothes.' (Georgieva 2018: 81)

   c. *Kuar  tḛl-ja     košk-i-z.*
     leaf    wind-ADV   leave-PST.3SG
     'The leaf flew away along/with the wind.' (Edygarova 2017: 78)

The use exemplified in (8) is discussed in detail in Edygarova (2017), who argues that forming adnominal modifiers with the adverbial case is particularly productive in the literary variety of modern Udmurt. The use illustrated in (9) is of interest in this paper, as it shows up in the adverbial clauses with *-(e)mja*, *-(o)nja* and, according to the standard analysis, with

*-(o)ńńa* as well. In the descriptive studies, these clauses are subsumed under the use illustrated in (9) (see Winkler 2001: 24; 2011: 53; Edygarova 2010).

The peculiar property of the adverbial case in Udmurt is that it can either precede or follow the possessive suffixes. Case suffixes in Udmurt generally have a fixed position: some (e.g. inessive, illative, elative) precede the possessive markers, while others (e.g. genitive, ablative, abessive) follow them; in the tradition of Finno-Ugric linguistics, the two morpheme orders are referred to as *Cx-Px* and *Px-Cx*, respectively (*Px* stands for possessive suffix, *Cx* for case suffix). Crucially, the adverbial is the only one in modern Udmurt that displays both orders (Edygarova 2010: 109–111). Edygarova (2010: 110) notes that Px-Cx is the general pattern for the adverbial, as in (9b); the Cx-Px order is rare, but she reports a few examples from dialectal texts: *ʒek śurel-len jugit-ja-z* [rye pollen-GEN light-ADV-POSS.3SG] '(we walked) by the light of the rye pollen'. Emelyanov (1927: 135) argues that the Px-Cx order is a new development and that the Cx-Px order is found in spoken language and folklore texts, e.g. *sojos-len mịnon-ja-zị* [they-GEN going-ADV-POSS.3PL] 'as they were going'.

In my view there are at least two factors that complicate the analysis of the adverbial case in Udmurt. As far as its functions are concerned, we see that this case has several seemingly unrelated functions.[6] As for its morphology, the varying order of the possessive suffixes and the adverbial case calls for an explanation, and this will be addressed in Section 3.1 and 4.2. Another issue regarding its morphology is that it is formally similar to the possessive declension of the inessive and illative cases. Generally, the inessive is expressed with the suffix *-ịn* (e.g. *gurt-ịn* [village-INE] 'in the village') and the illative is expressed by *-(j)e* (e.g. *gurt-e* [village-ILL] 'to the village'). In the presence of possessive suffixes, these two cases are marked in the same way: instead of *-ịn* or *-(j)e*, we find *-a*, which precedes the Px, e.g., *gurt-a-mị* [village-INE-POSS.1PL] or [village-ILL-POSS.1PL] 'in or to our village' (Perevoshchikov 1962: 88; Winkler

---

6. Note also that the term *adverbial case* is rather unfortunate from a comparative perspective. In other languages that employ a marker labeled as *adverbial case*, e.g. Georgian and Adyghe, this suffix has various functions: it is used to derive adverbs as well as to mark secondary predicates and certain non-finite clauses (see Hewitt 1995: 534–535; Serdobolskaya 2016). These functions may seem similar to the one illustrated in (9), but the nominal modifier function shown in (8) clearly does not fit the label *adverbial*.

2001: 29; 2011: 62–64; Edygarova 2010).[7] When the stem ends in a vowel, we find *-ja*, e.g. *busi̯-ja-mi̯* [field-INE-POSS.1PL] or [field-ILL-POSS.1PL] 'in or to our field'. This means that when the stem ends in a vowel, the adverbial case is identical in form to the suffix encoding the inessive or illative cases before possessive suffixes. The presence of the glide in the exponent marking the inessive or illative cases is due to epenthesis (Edygarova 2010: 107). Epenthetic *-j* occurs in various contexts in Udmurt, although differences between the standard language and the dialects are observed, cf. standard *kniga-jez* [book-POSS.3SG] vs. dialectal *kniga-ez* [book-POSS.3SG] and standard *karta-os* [map-PL] vs. dialectal *karta-jos* [map-PL] (Perevoshchikov 1962: 45–46). I will argue below that the key to understanding the adverbial case is the spatial meaning illustrated in (9c) and, more generally, the morphological similarity of the exponent of the adverbial case (*-ja*) to the one that marks the inessive or illative before possessive suffixes (*-(j)a*). Moreover, in my view, the *-(e)mja-* and *-(o)nja*-clauses present the crucial piece of evidence here (see Section 3.1). More generally, the proposal regarding the adverbial case will be fed into the general analysis of postpositional phrases in Udmurt (see Section 4.2).

With this in mind, let us turn to the detailed description of *-(e)mja-*, *-(o)nja-*, and *-(o)ńńa*-clauses.

## 3. New empirical findings

In this section, I present new empirical findings regarding the adverbial clauses expressed with the suffixes *-(e)mja*, *-(o)nja* and *-(o)ńńa*. As stated in the Introduction, according to the standard analysis of these clauses, a nominalization combines with the adverbial case; moreover, the suffix *-(o)ńńa* is considered to be a dialectal variant of *-(o)nja*. In this section, I will present empirical arguments that refine or even challenge these assumptions and then in Section 4, I will propose an alternative analysis

---

7. This segmentation follows Winkler (2001: 29; 2011: 63–64) and Usacheva (2012), that is, synchronically, *-(j)a* can be considered to be the exponent of the inessive or illative cases used before possessive morphemes. The historical development of these forms has been debated, however (see Serebrennikov 1963: 112–115; Csúcs 2005: 205; see also Edygarova 2010: 108). Nevertheless, there does not seem to be a consensus on how to gloss *-(j)a*; Winkler (2001) uses either INE or ILL, depending on the meaning of the datum, and I will follow this convention.

that successfully captures the semantic and morphosyntactic properties of these clauses.

I first carefully examine *-(e)mja-* and *-(o)nja-*clauses. In accordance with the existing literature, I propose that they indeed feature the adverbial case (and are glossed accordingly). However, I will present new data based on which I will claim that there are two types of *-(e)mja-* and *-(o)nja-*clauses, both semantically and morphosyntactically.

Secondly, I shall investigate the adverbial clauses formed with the suffix *-(o)ńńa* used in the Middle Cheptsa dialect. What I will show is that these clauses are not simply temporal ones; rather, they also have a locative component in their meaning. I will also draw a parallel with Beserman Udmurt, which utilizes a similar clause type. In order to account for the locative semantics, I will later argue in Section 4 that the suffix *-(o)ńńa* is not to be decomposed morphologically the same way as *-(o)nja* (*pace* the standard analysis). The alternative morphological decomposition requires some theoretical background, which will only be introduced in Section 4; for this reason in this section I do not segment the suffix *-(o)ńńa* in the glosses.[8]

## 3.1. Two types of adverbial clauses with the adverbial case

This section deals with the adverbial clauses formed with suffixes *-(e)mja* and *-(o)nja*. I capitalize on an observation made by Edygarova (2010) regarding the morpheme order in these clauses, by linking morpheme order to the meaning of the adverbial clause. Furthermore, I present new findings regarding the distribution of the Cx-Px and Px-Cx orders with the adverbial case based on corpus data. In addition, I present new data regarding the type of the non-finite clause involved (event or non-event

---

8. A remark is in order regarding these clauses. Based on corpus data from the Udmurt Social Media Corpus (which contains spoken/dialectal texts), it can be shown that several suffixes are in use, and alongside the standard Udmurt *-(o)nja*, we also find *-(o)ńńa* and *-(o)nna*. The examples presented in Fokos-Fuchs (1958) also contain different forms. This might suggest that there is dialectal variation with respect to the form of the suffix. It has been reported that *-(o)nja*-clauses are far less frequent than other types of temporal clauses, e.g. the 'when'-clauses encoded with the converb *-ku* (see Georgieva 2018). Section 3.1 focuses on Standard Udmurt *-(o)nja*-clauses, while Section 3.2 zooms in on *-(o)ńńa*-clauses in the Middle Cheptsa dialect; all further questions regarding *-(o)nja*-clauses in Standard Udmurt and across the Udmurt dialects will be left for future research.

nominalization, see below). Thus, based on their semantic and morpho-syntactic properties, I will argue that two types of *-(e)mja-* and *-(o)nja-* clauses are to be distinguished.

Edygarova (2010: 109–111) observes the following regarding the Cx-Px and Px-Cx order when the adverbial case is used with *-(e)m-* and *-(o)n-*nominalizations. The Cx-Px order occurs when the nominalization is used as an adverbial clause; here the possessive suffixes mark agreement with the subject of the non-finite clause. The Px-Cx order, on the other hand, is preferred when the nominalization is used as a manner or circum-stantial adverbial. This is shown for *-(e)mja-*clauses in (10) and (11), respec-tively. Edygarova does not provide a minimal pair for *-(o)nja-*clauses, but we find one in (2) and (7), repeated below in (12) and (13) for the reader's convenience.

(10)    (Edygarova 2010: 110)

[*Uža-nӭ     bӭgat-em-ja-s*]              *śud-em,*
work-INF    be.able-NMLZ-ADV-POSS.3SG    feed-EVID.3SG

*pe,       soostӭ     so.*
QUOT      3PL.ACC    3SG
'While he was able to work, he was feeding them, they say.'

(11)    (Edygarova 2010: 111)

[*Bӣgat-em-e-ja=no*
be.able-NMLZ-POSS.1SG-ADV=ADD

*vala-m-e-ja*]                     *uža-śko.*
understand-NMLZ-POSS.1SG-ADV       work-PRS.1SG
'I work according to my (own) abilities and understanding.'

(12) =(2) (Winkler 2011: 53)

[*Oźị     mịn-on-ja-z*]       *metro-je     vu-i-z.*
this.way  go-VN-ADV-POSS.3SG   metro-ILL    arrive-PST-3SG
'As he went like this, he arrived at the metro (station).'

(13) =(7) Udmurt Corpus (*Udmurt duńńe*, 2008.05.28)

*Mi         um        ľukiśke               aďami-jez*
1PL.EXCL    NEG.1PL   separate.PRS.CNG.PL   person-ACC

*vịžị-jez-ja,            osk-on-ez-ja* […]
root-POSS.3SG-ADV       believe-VN-POSS.3SG-ADV
'We don't separate people based on their origin and/or religion.'

Although Edygarova's generalization seems to be on the right track, it can be further qualified. What I would like to point out is that the morpheme order correlates with the meaning of the adverbial clause. In (11) and (13), the non-finite clause means 'according to' or 'based on', as expected with the adverbial case. In (10) and (12), however, the meaning of the adverbial clause is not 'according to' or 'based on' but rather 'while'.[9] This holds for all examples with the Cx-Px order presented by Edygarova (2010). The same pattern is found in the examples presented by Fokos-Fuchs (1958) as well as in the data from the Udmurt Corpus. Recall that Winkler (2011: 116) has noted *passim* that the meaning of these clauses is compositional, i.e. derivable from the meaning of the adverbial case. However, this does not predict the correlation between morpheme order and semantics, nor does it explain why the adverbial clauses in (11) and (13) have temporal semantics – which is not the typical use of the adverbial case in general, cf. its functions as summarized in Section 2.2.

The second empirical observation concerns the distribution of the two morpheme orders with the adverbial case. Recall from Section 2.2 that according to Edygarova (2010: 110), Px-Cx is the general pattern for the adverbial case and the Cx-Px order is rare. This is indeed confirmed by corpus data. In the Udmurt Corpus (9.57 million words), the Px-Cx pattern has 16,129 hits, whereas the Cx-Px order has only 675 hits. But the corpus data allow to make new observations about the type of nouns the adverbial case combines with. Specifically, it can be observed that the corpus hits for the Cx-Px order with the adverbial case feature *only -(e)m* and *-(o)n* forms. The Px-Cx order, on the other hand, is attested with non-derived nouns (cf. *diśkut-ez-ja* [clothes-POSS.3SG-ADV] 'based on his clothes' in (9b)) and nominalizations, as in (11) and (13).

Thus, the puzzle is why nominalizations show "dual" behavior, unlike non-derived nouns. I argue that this is because nominalizations in Udmurt fall into two types and one of them patterns with non-derived nouns. In my previous work, I have argued that *-(e)m-* and *-(o)n-*nonfinites come in two guises: *event* and *non-event nominalizations* (Georgieva 2018: 48–57, see also Serdobolskaya et al. 2012 and Dékány & Georgieva 2020). The latter may denote result nouns, manner nominalizations (in the sense of Comrie & Thompson 2007) or object nominalizations (e.g.

---

9.   The English translation of (10) follows Edygarova's translations in which she uses *poka* 'while'.

instrument). Consider *žećira-n* [swing-vn], which can be (i) an event nominalization ('swinging'), (ii) a manner nominalization ('the way of swinging') or (iii) an instrument nominalization '(a) swing'. In (14), *žećira-m* [swing-NMLZ] can have either a manner or an event reading.

(14)  (Georgieva 2018: 53)

Ivan-len       *žećira-m-ez*
Ivan-GEN     swing-NMLZ-POSS.3SG

*anaj-ataj-os-se    pajmit-i-z.*
mother-father-PL-POSS.3SG.ACC    amaze-PST-3SG
'The way in which Ivan was swinging amazed his parents.'
(manner nominalization)
'Ivan's swinging amazed his parents.' (event nominalization)

The two types of nominalizations differ not only in their semantics but also in their grammatical properties. For example, only non-event nominalizations can be pluralized, as shown in (15) for *-(e)m*-nominalizations (pluralization of *-(o)n*-nominalizations patterns alike, see Georgieva 2018). This supports the idea that non-event nominalizations behave like garden-variety noun phrases.

(15)  (Georgieva 2018: 53)

Ivan-len       *žećira-m-jos-iz*
Ivan-GEN     swing-NMLZ-PL-POSS.3SG

*anaj-ataj-os-se    pajmit-i-z.*
mother-father-PL-POSS.3SG.ACC    amaze-PST-3SG
'The ways in which Ivan was swinging amazed his parents.'
(manner nominalization)
*'Ivan's swingings amazed his parents.' (event nominalization)

The distinction between event and non-event nominalizations is relevant for the morpheme order with the adverbial case in the following way. Above, I argued based on corpus data that the Cx-Px order with the adverbial occurs only with nominalizations. I also argued that the Px-Cx order is attested with both non-derived nouns and nominalizations. Here, I would like to further specify these claims: the Cx-Px order occurs when the adverbial case combines with event nominalizations, whereas the Px-Cx order is found with non-derived nouns, including non-event nominalizations.

Support for these claims comes from corpus data. Based on (impressionistic) observations it seems that the *-(o)n*-nominalizations showing the Px-Cx order with the adverbial case have the semantics of non-events. We have already seen an example: the nominalization *osk-on* [believe-vn] in (13) does not have an event reading but encodes an abstract noun related to the event: it means 'religion, belief'. Other corpus examples are also non-events: *ivorton* 'notification', *kuron* 'request', *kiržan* 'song', *ulon* 'life', *koson* 'order', etc.

More importantly, the corpus data allow us to check for the compatibility of the two morpheme orders in combination with the plural. The searches in the Udmurt Corpus revealed that when the nominalization is pluralized, the adverbial case is attested *only* in the Px-Cx order (79 hits for *-(e)m*-nominalizations and 511 hits for *-(o)n*-nominalizations), e.g. *kos-em-jos-iz-ja* [order-NMLZ-PL-POSS.3SG-ADV] 'according to/based on his/her orders' and *kur-on-jos-si-ja* [ask-vn-PL-POSS.3PL-ADV] 'according to/based on their requests'. This provides strong support for the idea that the Px-Cx order with the adverbial case combines with non-event nominalizations – as was shown above, they are pluralizable, unlike event nominalizations. The clauses attested in the corpus are translatable with 'according to' or 'based on'.

The Cx-Px order, on the other hand, is not attested with pluralized nominalizations in the Udmurt Corpus. In fact, based on native speakers' judgments, this is ungrammatical, as shown in (16). In this example, the pluractionality is both lexically encoded (with the adverbial 'many times') and also pragmatically plausible (mountaineers go on multiple hikes); nevertheless, the plural marking is disallowed, as with event nominalizations in general. These facts support the present proposal according to which the Cx-Px order with the adverbial case is possible only with event nominalizations. The adverbial clause in (16) has temporal semantics.

(16)  elicited

[*Gureź-e*          (*tros*     *pol*)     *tuba-m-ja-z* /
mountain-ILL    many    times    climb-NMLZ-ADV-POSS.3SG

*\*tuba-m-jos-ja-z*]          *aľpińist*          *odig*
climb-NMLZ-PL-ADV-POSS.3SG    mountain.climber    one

*pol=no*        *usi-mte.*
time=ADD    fall-NEG.EVID.3SG

'The mountain climber didn't fall a single time while climbing the mountains (many times).'

To recap, the findings in this section allow us to state that there are two types of *-(e)mja-* and *-(o)nja-*clauses, which differ both semantically and morphosyntactically. Their properties are summarized in Table 1 and will be accounted for in Section 4 by rethinking the status of the adverbial case in Udmurt.

Table 1: The two types of adverbial clauses with the adverbial case

|        | Meaning                                      | Morpheme order | Nominalization type |
|--------|----------------------------------------------|----------------|---------------------|
| Type 1 | temporal ('while')                           | Cx-Px          | event               |
| Type 2 | oblique ('according to' or 'based on')       | Px-Cx          | non-event           |

## 3.2. Temporal-locative adverbial clauses with *-(o)ńńa*

In this subsection I discuss *-(o)ńńa-*clauses in the Middle Cheptsa dialect. Since they have not been previously described, I start with some general observations regarding their temporal interpretation. Then, I show that these clauses have a locative component in their meaning.

Based on my data, I argue that *-(o)ńńa-*clauses express an adverbial clause that denotes a time interval. The event of the main clause can overlap with or take place within that time interval, see (17) and (18), respectively. My consultants often paraphrase the former with the converb *-ku* 'when' and the latter with the converb *-ććoź* 'while' (standard Udmurt *-toź*, see below in (24)).

(17)   elicited (Middle Cheptsa dialect)

[*Isak-ịn    dịšetiś    luị-sa    uža-ńńa-m*]
Isak-INE    teacher    be-CVB    work-*ońńa*-POSS.1SG

*umoj    uli-śko        val.*
well    live-PRS.1SG    COP.PST
'While I was working as a teacher in Isak, I was living well.'

(18)   elicited (Middle Cheptsa dialect)

[*Isak-ịn    dịšetiś    luị-sa    uža-ńńa-m*]        *kuiń    pol*
Isak-INE    teacher    be-CVB    work-*ońńa*-POSS.1SG    three    times

*už-me            danjazị.*
work-POSS.1SG.ACC    award.PST.3PL
'While I was working as a teacher in Isak, my work was awarded three times (lit. they awarded my work three times).'

As an introduction to the locative semantics of the clause type illustrated in (17) and (18), let us consider *-(o)ńńiga-*clauses in Beserman Udmurt, as described by Usacheva & Serdobolskaya (2015). According to them, the temporal orientation of *-(o)ńńiga-*clauses in Beserman Udmurt is similar to the Middle Cheptsa facts. What they also claim is that the event expressed by the adverbial clause and the event of the main clause should take place in the same location. To support this they provide (19): the example was rejected by their consultants, as using an *-(o)ńńiga-*clause implies that the teacher delivered her babies at work. To express the intended meaning, the speakers used the converb *-idćoź* 'while, until, as long as' (which corresponds to *-toź* in standard Udmurt and *-ććoź* in Middle Cheptsa dialect, see below in (24)). Importantly, this is the only example presented in their study in support of their claim.

(19)    (Usacheva & Serdobolskaya 2015: 386) (Beserman Udmurt)

*[*Vorća-jə̑n=no    Šamardan-ə̑n    učiťeľ-ə̑n*
Vortsa-INE=ADD   Shamardan-INE   teacher-INE

*uža-ńńiga-m*]      *mon   kwiń   pińal    vaj-i.*
work-*ońńiga*-POSS.1SG    1SG   three   child.ACC   bring-PST.1SG
Intended: 'While I was working as a teacher in Vortsa and Shamardan, I gave birth to three children.'

I tested (20), which was modeled after the Beserman Udmurt (19). My consultants found (20) semantically/pragmatically odd (as indicated by the # sign), as it implies delivering the babies at work.

(20)    elicited (Middle Cheptsa dialect)

#[*Isak-i̯n   di̯šetiś   lu̯i-sa    uža-ńńa-m*]
Isak-INE   teacher   be-CVB   work-*ońńa*-POSS.1SG

*kuiń   pinal    vaj-i.*
three   child.ACC   bring-PST.1SG
'While I was working as a teacher in Isak, I gave birth to three children.'

Thus, at first sight these clauses in Beserman Udmurt and the Middle Cheptsa dialect show a parallel behavior: they are not simply temporal clauses but imply that the two events, the one of the main clause and the one of the adverbial clause, should take place in the same location.

This initial hypothesis should be refined for the Middle Cheptsa data, however.[10] One might ask whether adding two different locative adverbials or having different implicit locations would be grammatical. In (21), the adverbial clause contains a locative that is different from the one in the main clause (the example was modeled after (3)). In (22), the two events are expected to take place in different locations (for pragmatic reasons), although the locations are left implicit.[11] In both cases it is possible to use *-(o)ńńa*-clauses.

(21)  elicited (Middle Cheptsa dialect)

> *Muš-jos-mi̧*      *bakča-i̧n*     *si̧li̧ś*     *umorto-i̧ś*
> bee-PL-POSS.1PL   garden-INE    standing    beehive-ELA
>
> *pegǯi-ľľam=ńi=no*              [*baba-jeni̧-mi̧*
> escape-EVID.3PL=already=ADD    grandmother-INS-POSS.1PL
>
> *azbar-i̧n*    *vi̧r-ońńa-mi̧*].
> yard-INE     be.busy-*ońńa*-POSS.1PL
> 'Our bees had (long) flown away from the beehive in the garden while we/me and our grandmother were busy (taking care of the chicks) in the yard.'

(22)  elicited (Middle Cheptsa dialect)

> [*Gubija-ni̧*           *vetl-ońńa-m*]    *baba-je*
> mushroom.pick-INF     go-*ońńa*-POSS.1SG   grandmother-POSS.1SG
>
> *skal-jos-mes*          *ki̧sk-em.*
> cow-PL-POSS.1PL.ACC    milk-EVID.3SG
> 'While I was picking mushrooms, my grandmother milked our cows.'

Thus, it seems that the semantic/pragmatic oddity of (20) cannot be directly explained in terms of the location of the two events. I believe that the correct explanation is related to the two events, the one expressed by

---

10. In a more recent study, Usacheva & Serdobolskaya (forthcoming) claim that the requirement for the two locations to be identical is not that strict in Beserman Udmurt. They provide one example for which they argue that the partial overlap between the locations makes the sentence felicitous. Since there are only two examples presented for Beserman Udmurt, it is difficult to make a comparison. It would be interesting to find out whether the restrictions are similar to what I show below for the Middle Cheptsa dialect. Hopefully, this question will be addressed in future studies.

11. I thank an anonymous reviewer for raising the issue of implicit locations.

the main clause and the one expressed by the adverbial clause, being interpreted as describing a *single situation*. That is, they are subparts of one macro-event. This requirement goes hand in hand with a spatiotemporal match between the events – but this is more complex than simply a ban on different locative adverbials. In (17) and (18), the adverbial clause and the main clause are viewed as subparts of a global situation: a description of a teacher's professional life. In contrast, construing working as a teacher and giving birth to three children as one macro-event yields the semantically odd reading of giving birth at school. Example (23) also supports this line of thinking: it is minimally different from (20), and importantly, the main clause allows for a construal according to which the two events form a single situation.

(23)  elicited (Middle Cheptsa dialect)

[*Isak-ịn  dịšetiś  luị-sa  uža-ńńa-m*]      *kuiń  pol*
Isak-INE   teacher  be-CVB  work-*ońńa*-POSS.1SG  three  times

*praktika-je  vetl-i    Iževsk-e.*
training-ILL  go-PST.1SG  Izhevsk-ILL

'While I was working as a teacher in Isak, I went to three trainings in Izhevsk.'

It seems much more difficult to ensure that the two events cannot be interpreted as a single situation when the main and the adverbial clauses have the same subject. Sentence (20) is the perfect example for such a construal, since under the intended reading, the adverbial clause has a scene-setting function: it sets a general background for the event of the main clause. But as I argue, -*(o)ńńa*-clauses must be interpreted as subparts of a macro-event together with the main clause. As a result of this, (20) is rendered the semantically odd reading of delivering babies at work. Examples (17), (18), and (23) can be interpreted as part of a macro-event together with the main clause. They also allow for a scene-setting reading of the adverbial clause, and under such a scenario, the speakers prefer using an alternative type of non-finite clause (-*ku* 'when' or -*ććoź* 'while'). Example (20) stands out because it allows only for a scene-setting reading.

When the two clauses have different subjects, it is possible to have construals such that the two events are viewed as a single situation. I argue that this is what we observe in (21) and (22): both can be perceived as descriptions of one larger event, and despite the fact that the two clauses contain two locative adverbials, the subevents "revolve" around the same

location (i.e. they are descriptions of what happened at home). This part of the proposal may sound slightly unconvincing, as such semantic contrasts, i.e. what counts as a macro-event, can be quite subtle and hard to capture. Consider the next two examples, however: the two events cannot possibly be construed as a single situation. Rather, they are about contrasting two events that take place at different locations. The context of (24) is different time zones, and the two events are simply contrasted with each other, without being included in a single situation. Example (25) is about the differences between Southern and Northern Udmurtia, and again, the main and the adverbial clause cannot be interpreted as subevents of one macro-event. As a consequence, in both contexts, -(o)ńńa-clauses are not acceptable; my consultants suggested using a different non-finite clause: -ććoź 'while' in (24) and -ku 'when' in (25).

(24)   elicited (Middle Cheptsa dialect)

[*Moskva-ịn   iźị-ććoźa-zị* /       **iź-ońńa-zị*],
Moscow-INE   sleep-CVB-POSS.3PL   sleep-ońńa-POSS.3PL

*Vladivostok-ịn   už-ịś   bertị-nị   poto=ńi.*
Vladivostok-INE   work-ELA   go.home-INF   exit.PRS.3PL

'While people in Moscow are (still) sleeping, people in Vladivostok are already going home from work.'

(25)   elicited (Middle Cheptsa dialect)

[*Lịmšor-ịn   kartoška   merttị-ku* /   **mertt-ońńa-zị*],
south-INE   potato.ACC   plant-CVB   plant-ońńa-POSS.3PL

*ujpal-ịn   lịmị   suźalo=na=uk!*
north-INE   snow.ACC   clean.PRS.3PL=still=EMPH

'While in the south [Udmurt] people are planting potatoes, in the north people are still shoveling show!'

Based on the presented evidence, I conclude that the -(o)ńńa-clauses used in the Middle Cheptsa dialect do not simply encode an event simultaneous with the matrix event (as 'when' and 'while'-clauses do). I showed that using -(o)ńńa-clauses is only possible when they can be construed as belonging to one macro-event together with the main clause. In my view, this requires or, rather, results in a spatiotemporal match between the two events. In Section 4, I will present an account of these properties of -(o)ńńa-clauses.

## 4. Analysis

In this section I put forward a theoretical analysis couched in a generative syntactic framework. The proposed analysis of *-(e)mja*, *-(o)nja*, and *-(o)ńńa* is similar to the traditional ones in treating these suffixes as morphologically decomposable, but it differs from them in several important respects. First, the proposed morphological decomposition of *-(o)ńńa* is crucially different from the standard one. Second, a novel analysis is put forward for the adverbial case in *-(e)mja* and *-(o)nja*.

Apart from decomposing the suffixes morphologically, the present proposal also states that these adverbial clauses are postpositional phrases (PPs). In Section 2.1, it was already demonstrated that non-finite adverbial subordination in Udmurt is typically encoded by using a non-finite clause selected by a postposition or a semantic case. Since postpositions and semantic cases are treated as exponents of the same syntactic head (P) in this framework (see below), this means that these non-finite adverbial clauses in Udmurt are PPs. Thus, *-(e)m bere* 'after' in (4), *-(e)men* 'by (doing)' or because' in (6) and *-(e)mja* 'as' in (1) are all PPs (see Georgieva 2018: Ch. 4 for further discussion). In the adopted framework, PPs are argued to have internally complex structure (the relevant theoretical assumptions are summarized in Section 4.1). The internal complexity of PPs will make it possible to account for the differences between the two types of *-(e)mja*- and *-(o)nja*-clauses (Section 4.2) and for the temporal-locative semantics of *-(o)ńńa*-clauses (Section 4.3).

### 4.1. The internal structure of postpositional phrases with special reference to Udmurt

In this section I first summarize the main assumptions regarding the structure of postpositional phrases made in the generative syntactic tradition. I then provide an overview of the existing studies dealing with PPs in Udmurt in this framework.

It is received wisdom in the literature that the heads of PPs, Ps, can be adpositions or semantic cases. This is supported by their syntactic and semantic similarities (from a typological perspective see Malchukov & Spencer 2009; Moravcsik 2009). Semantically, spatial Ps express how the position of the Figure is related to the Ground; this holds for both adpositions and cases. Syntactically, PPs headed by adpositions have the same distribution as those headed by semantic cases. The cross-linguistic comparison,

e.g. the fact that languages like English employ the preposition *in*, whereas Hungarian uses the inessive case suffix to express the same meaning, also supports the idea that adpositions and semantic cases are exponents of same syntactic head, namely, P. One can provide language-specific arguments in favor of this, too. In several Uralic languages, spatial adpositions and case suffixes show a tripartite division into goal, source, and location (see Kittilä et al. 2022 for general overview; see also below on Udmurt). Consider for instance the triplet of the postpositions *elé* 'to the front', *elől* 'from the front', *előtt* 'in front' and the triplet of the illative, elative, and inessive cases in Hungarian (see Asbury 2008; Dékány 2011; Dékány & Hegedűs 2021, among many others). These studies argue that the difference between postpositions and case suffixes in Hungarian is morphophonological in nature, e.g. suffixes are monosyllabic and most of them show vowel harmony with the word they attach to.

The internal structure of spatial PPs is argued to be complex: it consists of several projections, on top of the nominal complement (a noun phrase) (Jackendoff 1983; van Riemsdijk & Huybregts 2002; Svenonius 2006; the contributions in Asbury et al. 2008 and Cinque & Rizzi 2010, among many others). Firstly, PPs feature projections for place- and path-denoting elements: PlaceP and PathP. Secondly, PPs may also host elements that are at the intermediate stage between relational nouns and adpositions based on their morphosyntactic properties. This can be observed for *front* in *in front of the car*, for example: it can be used as a noun, but in the aforementioned construction, it can be neither pluralized (*in fronts of the car*) nor modified (*in smashed-up front of the car*) (Svenonius 2006). Svenonius (2006) proposes that this kind of elements are hosted in a separate projection, Ax(ial)PartP. Thus, the internal structure of such PPs is internally complex, with both PlaceP and AxPartP being projected: [$_{PlaceP}$ *in* [$_{AxPartP}$ *front* [ *of the car*]]].

It has been observed that Ps often grammaticalize from nominal elements. This diachronic change involves filling the AxPart head; later this element may lose its nominal properties completely, which results into the development of a new Place or Path head (see Waters 2009 on English; Hegedűs 2014 on Hungarian; Grünthal 2022 on Uralic in general). Another diachronic change in the PP domain involves morphologization: a syntactically independent adposition may turn into a case suffix (see for example Hegedűs 2014 on the history of spatial cases in Hungarian). But as stated above, both syntactically independent adpositions and morphologically bound cases are treated as Ps in this framework.

This short overview of the generative literature on PPs was meant to provide the underlying assumptions of the proposed analysis, namely, that PPs may have a complex internal structure consisting of several projections. As will be shown below, PPs in Udmurt have also been analyzed as internally complex. The grammaticalization of nouns into AxPart heads and the morphologization of Ps will also be important when discussing the Udmurt data, to which I turn next.

Postpositional phrases in Udmurt have been analyzed in a generative syntactic framework by Simonenko & Leontyev (2012) and Usacheva (2012). Usacheva (2012) shows that most postpositions in Udmurt are in fact nominal: *vi̧l* 'top', *aź* 'front', *ul* 'bottom', etc. Thus, they are similar to the AxPart heads discussed above; Usacheva uses the label $N_{PLACE}$. She argues that these Ps express how the Figure is located with respect to the Ground, e.g. *vi̧l-i̧n* in *korka vi̧l-i̧n* [house top-INE] 'on top of the house' expresses that the Figure is located on top of the Ground (the house). The place semantics, i.e. that the Figure is stationary, comes from the inessive case. These nominal Ps can combine with various semantic cases, thus forming *series*: *vi̧l-i̧n* [top-INE] 'on top', *vi̧l-e* [top-ILL] 'onto the top', *vi̧l-i̧ś* [top-ELA] 'from the top', etc. In Usacheva's work, place- and path-denoting Ps are hosted in a dedicated locative K[ase] projection, $K_{LOC}P$. This gives the structure of [$_{KlocP}$ *-i̧n* [$_{NplaceP}$ *vi̧l* [*korka*]]] for *korka vi̧l-i̧n* [house top-INE] 'on top of the house'. Usacheva (2012) mentions that there are also a few *non-serial postpositions* in Udmurt, e.g. *vamen* 'across', *kuźa* 'along', *ponna* 'for'. These are analyzed as heads of simple PPs (without $K_{LOC}P$ and $N_{PLACE}P$), as they do not combine with semantic cases: [$_{PP}$ *ponna* [$_{NP}$]].[12]

Usacheva also discusses what she calls *serial spatial cases* in the Permic languages. An example of this is the spatial suffix *-ń*, the so-called DOMUS suffix, used in Beserman Udmurt and in the Middle Cheptsa dialect (see also Teplyashina 1970; Karpova 2005: 85–89). It locates the Figure with respect to the Ground's place (home). The DOMUS suffix also combines with semantic cases, just like *vi̧l* 'top', as shown for the Middle Cheptsa dialect in (26).[13] This suffix will be important for the discussion of *-(o)ńńa*-clauses (Section 4.3).

---

12. A similar distinction is made by Winkler (2011: 133–136): the two groups are referred to as *inflecting* and *non-inflecting* postpositions.

13. The so-called *familial local cases* used in certain dialects of Hungarian are similar (see Kittilä et al. 2022: 888).

(26) elicited (Middle Cheptsa dialect)

   a. *baba-ń-e*
      grandma-DOMUS-ILL

   b. *baba-ń-i̯n*
      grandma-DOMUS-INE

   c. *baba-ń-i̯ś*
      grandma-DOMUS-ELA
      'to, at, from grandma's place'

The DOMUS suffix is said to originate from the postposition *dińe* 'at, around, next to' (Teplyashina 1970: 169) or the noun *iń* 'place' (Emelyanov 1927: 123). Thus, this can be argued to be another instance of the morphologization path mentioned above: the nominal P became suffixal. Importantly, in Usacheva's analysis, both nominal Ps like *vi̯l* 'top' and serial spatial cases like the DOMUS suffix are analyzed as $N_{PLACE}$ heads, and thus the syntactic structure of *korka vi̯l-in* [house top-INE] 'on top of the house' and *baba-ń-i̯n* [grandma-DOMUS-INE] 'at grandma's place' is identical: $[_{KlocP} -in [_{NplaceP} vi̯l / -ń-]]$. Additionally, possessive agreement may be present in the Udmurt PPs, as in *vi̯l-a-z* [top-INE-POSS.3SG] 'on top of it'.

Usacheva also discusses how motion with respect to the Ground is expressed. This is done with the help of the semantic cases, hosted in the $K_{LOC}P$ projection. This was already shown for the inessive, but other semantic cases also belong here: the illative, elative, prolative,[14] terminative, egressive, and approximative cases. This means that these cases are also Ps in this syntactic framework. Simonenko & Leontjev (2012) extend this line of analysis to the instrumental case as well. The adverbial case, which lies at the heart of the present study, as it is argued to be found in the suffixes *-(e)mja*, *-(o)nja*, and *-(o)ńńa*, is not discussed by Simonenko & Leontjev (2012) nor by Usacheva (2012), even though it is a semantic case, which can have a spatial meaning (as in (9c)).

In sum, the relevant point from this subsection is that PPs in Udmurt are argued to be of two types: simple and complex PPs. The latter feature a $K_{LOC}P$ and an $N_{PLACE}P$.[15] As for their morphological boundedness, both $K_{LOC}$ and $N_{PLACE}$ heads can be suffixal in Udmurt.

---

14. This case is generally termed *prolative* in most of the grammars, with the exception of Winkler (2001) who uses the term *transitive*.

15. In what follows I will continue using $K_{LOC}P$ and $N_{PLACE}P$, following Usacheva. As noted above, the former corresponds to Place or PathP and the latter to AxPartP.

In light of this discussion, the structure of the 'after'- and 'before'-clause in (4) and (5) is identical to the postpositional phrases discussed above: it is internally complex, consisting of a nominal P and a path- or place-denoting P. The difference is that the complement of these Ps is a non-finite clause. The fact that these spatial Ps have a temporal meaning in (4) and (5) is not surprising: it is a cross-linguistic tendency that spatial elements may acquire temporal meanings (see Haspelmath 1997); in Section 4.2 we will also discuss how temporal Ps may develop non-temporal meanings and how this is related to the internal complexity of the postpositional phrase.

## 4.2. The two types of *-(e)mja-* and *-(o)nja*-clauses

In Section 3.1, I argued that there are two types of *-(e)mja-* and *-(o)nja*-clauses: one of them has a temporal meaning, shows the Cx-Px order, and features an event nominalization, while the other has an oblique meaning, displays the Px-Cx order, and contains a non-event nominalization.

These new empirical findings will lead to a newly proposed analysis of the adverbial case, different from the standard one. Below I will argue that what traditional grammars have labeled as *adverbial case* should be analyzed as two types of postpositional phrases. The temporal one (ADV-1 henceforth) is a complex P, whereas the oblique one (ADV-2 henceforth) is a simple P.[16] This means that it would be more accurate to say that there are two adverbial cases in present-day Udmurt. Furthermore, I will argue that they are diachronically related: ADV-1 gave rise to ADV-2. This will provide an elegant and explanatorily powerful account of the morpheme order. In addition, it will derive the semantics of *-(e)mja-* and *-(o)nja*-clauses.

## 4.2.1. The spatiotemporal adverbial case: a complex PP

Recall from Section 3.2 that Type 1 clauses with the adverbial case have a temporal meaning (see (10) and (12)). Thus, their meaning resembles the inessive case, i.e. temporal 'in'. Moreover, Type 1 clauses illustrated in (10) and (12) show the Cx-Px order, which is found with the inessive cases. Recall also from Section 2.2 that the suffix of the adverbial case (*-ja*) is formally similar to the possessive declension of the inessive and illative cases (*-(j)a*).

---

16. These two labels are used primarily for presentational purposes, i.e. to disambiguate which adverbial case I am talking about in the text.

Two hypotheses can be entertained in order to account for connection between the inessive and the adverbial cases in examples like (10) and (12). The first hypothesis is that examples like (10) and (12) contain the inessive rather than the adverbial case (as I have proposed in Georgieva 2018). This is supported by the semantics of the adverbial clause as well as by the Cx-Px order. In this way, we also account for the varying order of the adverbial case and the possessive suffixes – there is no variation since the Cx-Px order involves the inessive, and not the adverbial case.

There are two complications with this hypothesis. Firstly, if what we find in examples like (10) and (12) is the exponent of the inessive used before possessive morphemes rather than the adverbial case, the question is why there is always a glide, even though the stem does not end in a vowel. In Section 2.2, I pointed out that there is variation with respect to epenthesis across the varieties of Udmurt (standard vs. dialectal), and thus one may speculate that the adverbial was codified in the standard language in a morphophonologically exceptional form. Another issue is that the inessive or illative *-(j)a* is used only with possessive suffixes, whereas the adverbial suffix *-ja* can be used without possessive suffixes, cf. (1), (8), and (9a). In principle, this problem can also receive an explanation: Serebrennikov (1963: 117) points out that the illative *-a* is found in certain postpositions, e.g. *pala* 'towards, in the direction of', which can be used without possessive suffixes. Even though this does not seem to be productive in modern Udmurt, one might speculate that the suffix of the inessive and illative cases *-(j)a* can be used without possessive markers, at least with nominalizations. Because of these complications, a second hypothesis can be put forward.

The second hypothesis is a modified version of the first one: the adverbial case is not *identical to* but *contains* the inessive. This is the line of analysis I will pursue here. This proposal also implies that the glide is not simply epenthetic, and thus gives a more convincing explanation of the formal differences between the adverbial and inessive than the mere reference to epenthesis; the function of *-j* will be discussed below. The main motivation for this proposal comes from the meaning of the adverbial case. Above I pointed out *-(e)mja*-clauses like (10) comparable to 'while', which can be derived from the meaning of the inessive case (temporal 'in'). However, the meaning of *tel-ja* [wind-ADV] 'along/with the wind' in (9c) is a spatial one, but it is not identical to the inessive. I would like to argue that this spatial meaning of the adverbial case is the original one. Although this is not productive in modern Udmurt in comparison with the

adnominal modifier use of the adverbial case (Edygarova 2017), the spatial meaning is attested in other forms, e.g. (*so vetliz*) *kįrja* '(s/he walked) along the field(s)' (Aminoff 1896: 26).[17] In Section 2.2, I mentioned the example given by Edygarova (2010: 110) *źek śurel-len jugit-ja-z* [rye pollen-GEN light-ADV-POSS.3SG] '(we walked) by the light of the rye pollen'. The same meaning is also found with *-(e)mja*-clauses and *-(o)nja*-clauses like (10) and in (12): 'as long as' being the temporal equivalent of the spatial 'along/ with/by'. This is also found in the postposition *dįr-ja* [time-ADV] 'during'. As was shown in Section 3.1, these non-productive forms aside, the original spatiotemporal meaning of the adverbial is found only with nominalizations in modern Udmurt.

In order to account for the meaning 'along', I propose that the adverbial case in (10) and (12), ADV-1, is a complex postpositional phrase that consists of the *-j* element, a reduced $N_{PLACE}$ head, that brings in the 'along' meaning, and the inessive. In the framework adopted here, this means that the PP of ADV-1 is structurally equivalent to the internally complex PPs discussed in Section 4.1: [$_{KlocP}$ *-a* [$_{NplaceP}$ *-j* [ ]]]. In terms of its morphophonology, the $N_{PLACE}$ *-j* is similar to the ones of the serial spatial cases, e.g. the DOMUS suffix *-ń*, because of its phonological reduction. It differs from those, however, in that it does not participate in a series, i.e. it does not combine with other place or path-denoting Ps.

In my view, it is precisely the phonological reduction of *-j* that has led descriptive grammars to classify *-ja* as a (single, non-decomposable) suffix, namely, the adverbial case. However, diachronic studies provide support for the decomposition analysis of the adverbial case. For example, according to Rédei (1988: 383), the adverbial suffix goes back to a lative/prolative *-j* and lative *-a* (the latter goes back to *\*-k*). This proposal is not identical to mine, but what is common is that the adverbial case is argued to be composed of two spatial elements. I argue that the internally complex PP is still found synchronically when the adverbial case has spatiotemporal semantics. That is, in examples like (10) and (12), the non-finite clause is embedded under a postpositional phrase consisting of *-j* plus the inessive. Hence, these clauses show agreement morphology that follows the inessive, similarly to postpositional phrases like *vįl-a-z* [top-INE-3SG] 'on top of it'.

---

17. The original translation to Finnish is *hän kulki aromaita myöten*. Edygarova (2017: 79) translates it with *он(a) ходил(a) по лугам*.

### 4.2.2. The oblique adverbial case: a simple PP

The next question is how the temporal clauses in (10) and (12) relate to those in (11) and (13), i.e. the ones that have oblique semantics and show the Px-Cx order. I propose that the two pairs are indeed related: the structure of the internally complex postpositional phrase was reanalyzed as a simple PP. Thus, ADV-1 gave rise to ADV-2. The structure of the two PPs is given in (27).

(27) a. ADV-1

$[_{\text{KlocP}} \text{-}a \, [_{\text{NplaceP}} \text{-}j \, [ \quad ]]]$

b. ADV-2

$[_{\text{PP}} \text{-}ja \, [ \quad ]]$

On the semantic side, I argue that the original spatial meaning of changed into a more abstract oblique meaning ('according to' or 'based on'). On the formal side, the reanalysis led to the Px-Cx order: the inessive is no longer "visible", thus the Px-Cx order is found when the new P, ADV-2, selects for a possessed noun phrase. Let me elaborate on both aspects of the change.

It is a well-known cross-linguistic observation that diachronically temporal clauses may develop non-temporal meanings, as for example English *since* (temporal > causal), *while* (temporal > concessive), and *rather* (temporal > preference) (Traugott & König 1991; from a typological perspective see Kortmann 1996: 89–94, Ch. 7). I propose that the more abstract oblique meaning 'according to' or 'based on' arose from the temporal 'while'. Consider examples where we potentially face ambiguity: 'in my thinking', which can have a temporal ('in the time of my thinking') or an oblique ('in my opinion') reading. In fact, *-(e)mja*-clauses are very often used with the verb 'think' in the corpora, which might have facilitated the semantic change. Thus, one can hypothesize a semantic change of a temporal relation into a more abstract oblique adverbial relation along the lines of *in (the time of) my thinking > in my opinion* or *I recognized him while he was walking > I recognized him based on the way he was walking*. Hence, it can be proposed that the meaning of the new ADV-2, i.e. 'according to' or 'based on', has developed from the spatiotemporal meaning of ADV-1. Importantly, the semantic change accompanied the structural reanalysis of the postpositional phrase of ADV-2, which became a simple P.

Regarding the morpheme orders possible with the adverbial case, I propose the following. In Section 3.2, I showed that the newly developed P,

ADV-2, takes *noun phrases* (including non-event nominalizations) as its complement. Here, I would like to propose that the restriction on why ADV-2 selects for noun phrases is semantically motivated: only referential nominals such as 'fact' but not event-denoting nominalizations are compatible with the new P because of its meaning, 'according to' or 'based on'. When the noun phrase is possessed, this gives rise to a Px-Cx order. The Cx-Px order is only possible with ADV-1: due to the presence of the inessive in this complex PP (cf. (27a)), the possessive morphology appears on top of the PP. The Cx-Px order is impossible with ADV-2, just like with simple Ps in general (see Arkhangelskiy & Usacheva 2015).

The present proposal also implies that the Cx-Px order is the original order for the adverbial case – as the complex PP in (27a) is the source from which ADV-2 developed. This is similar to what Emelyanov (1927: 135) claimed: he argued that the Px-Cx order is a new development (see Section 2.2). However, Emelyanov does not provide any arguments for this claim.[18] In my analysis, the Px-Cx order became possible as the result of the structural reanalysis of the postpositional phrase into a simple PP.

In sum, the theoretical analysis of the two types of *-(e)mja-* and *-(o)nja-* clauses discussed in Section 3.1 states the following: (i) the temporal ones feature the original adverbial case, which is an internally complex PP that includes the inessive, and thus show a Cx-Px order; (ii) the oblique ones feature the newly-developed adverbial case: a simple P, the complement of which is a noun phrase (which itself might be possessed, yielding a Px-Cx order). I argued that the latter emerged from the former. It should be emphasized that although this reanalysis is explained in light of the structure of postpositional phrases in the chosen framework, similar historical changes have been proposed in more traditional studies as well. For example, Serebrennikov (1963: 12) and Bartens (2000: 84) argue that the instrumental and the inessive cases have arisen as allomorphs of the Proto-Uralic locative case. Thus, such "splits" of the original case are not unheard of in the Permic languages.[19]

In the next subsection, I turn to *-(o)ńńa-*-clauses, which also tackle the problem of the adverbial case.

---

18. As far as I can tell, the discussion of morpheme order in his work is set in the long-standing debate in Finno-Ugristics regarding the order of possessive suffixes and case. The main question in this debate is which order, i.e. Px-Cx or Cx-Px, is to be reconstructed for Proto-Uralic.
19. I thank Arja Hamari for the discussion of this issue.

## 4.3. The morphological decomposition of *-(o)ńńa*

In Section 3.2, I showed that *-(o)ńńa*-clauses in the Middle Cheptsa dialect are not simply temporal clauses but have locative semantics: the events expressed by the main and adverbial clauses must be construed as one event, and thus they must match spatiotemporally. These findings are not predicted from the standard morphological decomposition of the suffix, according to which it contains the *-(o)n*-nominalization and the adverbial case, i.e. *-on-ja*; the surface form *-(o)ńńa* being the result of an assimilation rule of the glide with the preceding consonant (similar assimilation is found with other suffixes as well, see Karpova 2005). Although this segmentation is plausible from a morphophonological point of view, the alternative proposed here is to segment the suffix as *-(o)n-ń-a* [VN-DOMUS-INE], with the DOMUS suffix discussed in Section 4.1. Below I will argue that this morphological decomposition correctly derives the semantics of these clauses and the requirement for a spatiotemporal match between the two events.

Before I go into the details of the new analysis, let me briefly discuss the proposal of Usacheva & Serdobolskaya (2015) for Beserman *-(o)ńńiga*-clauses. Recall from Section 3.2 that those authors argued that the event expressed by the *-(o)ńńiga*-clause and the one expressed by the main clause must take place in the same location (cf. (19)). Usacheva & Serdobolskaya (2015) explain these facts in the following way. Beserman Udmurt employs locative nominalizations in the sense of Comrie & Thompson (2007: 340) (glossed as NLOC below); they are formed with the suffix *-(o)ńńig* (cf. locative nominalizations with the suffix *-(o)ńńi* used in standard Udmurt: *dugd-ońńi* [stop:V-*(o)ńńi*] '(bus/tram) stop'). These locative nominalizations are fully nominal: they can be pluralized, can stand in argument position, etc. They can also be complements of semantic cases: in (28) they are used with the suffix *-(j)a* that expresses the inessive or the illative cases before possessive suffixes. The resulting form is identical to the temporal-locative clauses with the suffix *-(o)ńńiga* in (19). Observe the ambiguity in (28): locative nominalization ('to my cow pasture') vs. clause ('where I herd my cows').[20]

(28)    (Usacheva & Serdobolskaya 2015: 371) (Beserman Udmurt)

    [*Skal voźma-ń-ńig-a-m*]         *lɜkt-em*      *kijon.*
    cow    herd:V-VN-NLOC-ILL-POSS.1SG      come-EVID.3SG   wolf
    'A wolf came to the place where I herd (the) cows / to my cow pasture.'

---

20. The English translation follows the original Russian translation.

Usacheva & Serdobolskaya (2015) propose that *-(o)ńńiga* is based on the locative nominalization and that it is currently on its way to grammaticalization into a converb suffix. They also discuss the morphological decomposition of the suffix: it is more complex than what I propose for Middle Cheptsa Udmurt (because of the *-ig* element), but crucially it features the deverbal nominalizer *-(o)n* as well as the DOMUS suffix *-ń*.

My analysis is partly similar to Usacheva & Serdobolskaya's (2015), but I treat the suffix *-(o)ńńa* as morphologically decomposable synchronically. Specifically, I propose that the suffix *-(o)ńńa* is composed of the following parts. First, we find the nominalization *-(o)n*, which brings the event semantics. Its presence is indisputable (cf. also the standard segmentation *-on-ja* [VN-ADV]).

Second, in accordance with Usacheva & Serdobolskaya (2015), I propose that the requirement for a spatiotemporal match between the two events is due to the presence of the the DOMUS suffix *-ń*. The DOMUS suffix today expresses the location of the Figure with respect to the Ground's home (Section 4.1), but recall also that the origin of this suffix was argued to be either the postposition *dińe* 'at, around, next to' (Teplyashina 1970: 169) or the noun *iń* 'place' (Emelyanov 1927: 123). Hence, I propose that in *-(o)ńńa*-clauses we find this more general 'place' meaning of the DOMUS suffix.

Third, unlike the standard segmentation according to which the suffix *-(o)ńńa* contains the adverbial case, I argue that it features the inessive *-(j)a* used before possessive suffixes. This is not only supported by the semantics (the inessive brings in the meaning of a temporal 'in') but also by the fact *-(o)ńńa*-clauses in the Middle Cheptsa dialect never occur without possessive suffixes, just like the inessive *-(j)a*. Furthermore, the Px-Cx order, as with the adverbial case, is impossible. This is shown in (29) (which was modeled after (3)).

(29)   elicited (Middle Cheptsa dialect)

    *Muš-jos-mi̯*        *pegǯi-ľľam=ńi=no*
    bee-PL-POSS.1PL    escape-EVID.3PL=already=ADD

    [*baba-jeni̯-mi̯*           *azbar-i̯n*
    grandmother-INS-POSS.1PL    yard-INE

    *vi̯r-oń-ń-a-mi̯* /              *\*vi̯r-on-mi̯-ja* /
    be.busy-VN-DOMUS-INE-POSS.1PL    be.busy-VN-POSS.1PL-ADV

    *\*vi̯r-on-ja*].
    be.busy-VN-ADV

    'Our bees had (long) flown away while we/me and our grandmother were busy (taking care of the chicks) in the yard.'

The data in (29) provide support against the standard morphological decomposition which postulates the adverbial case. But recall from the previous subsection that I proposed that synchronically we need to distinguish two adverbial cases. Thus, my proposal regarding the morphological decomposition of *-(o)ńńa* should be evaluated not only against the standard understanding of what the Udmurt adverbial case is but also against the predictions of the proposal in Section 4.2. There I argued that ADV-2 is a simple P, which has oblique meaning ('according to' or 'based on') and displays the Px-Cx order. Given that *-(o)ńńa*-clauses have temporal semantics and disallow this morpheme order, we can rule out ADV-2. More interesting, however, is whether ADV-1 is a possibility because, as shown in Section 4.2.1, the PPs headed by it have temporal semantics, similarly to *-(o)ńńa*-clauses. Despite this fact, I argue that the suffix *-(o)ńńa* cannot be decomposed as containing ADV-1. One piece of evidence comes from the obligatoriness of possessive marking: the exponent of the inessive, unlike the adverbial case (both ADV-1 and ADV-2), must appear with possessive markers. The other piece of evidence comes from semantics: both ADV-1-clauses and *-(o)ńńa*-clauses have temporal semantics, but only the latter have an additional locative component in their meaning. I argue that the locative meaning results from the presence of the DOMUS suffix *-ń*, an $N_{PLACE}$ head, which is independently attested in the Middle Cheptsa dialect. Recall that ADV-1 also contains an $N_{PLACE}$ head, *-j*, which I argued to be a spatial one, with the meaning 'along', and its temporal equivalents. In the theoretical framework adopted here, this means that both ADV-1-clauses and *-(o)ńńa*-clauses are analyzed as complex PPs, with the suffixes *-ń* and *-j* filling in the same syntactic position ($N_{PLACE}$). Given that they are hosted in the same position, they are predicted to be in complementary distribution. This is a theoretical argument against postulating ADV-1 in the morphological decomposition of *-(o)ńńa*.

In sum, the underlying form is *-(o)n-ń-a* [VN-DOMUS-INE]. The structure of this complex PP is [$_{KlocP}$ *-a* [$_{NplaceP}$ *-ń* [ ]]]. Accordingly, the meaning of *-(o)ńńa*-clauses is 'in the time and place of V-ing'. The morphological decomposition allows us to explain the fairly complex meaning of *-(o)ńńa*-clauses (temporal clause with an additional locative meaning) in a natural way: it is derived by the components of the suffix.

## 5. Conclusion

In this paper I examined three types of non-finite adverbial clauses: the ones formed with the suffixes *-(e)mja*, *-(o)nja*, and *-(o)ńńa*. The analysis proposed for the structure of these clauses aligned them with the syntax of postpositional phrases in the language. I argued that these adverbial clauses are comprised of a non-finite clause and a semantic case. This analysis has two advantages: it accounts for the external distribution of these clauses and it also derives their meaning in a compositional manner.

On the more general level, the present analysis also contributed to our understanding of PP syntax in Udmurt, especially with reference to clausal PPs and the distinction between internally complex and simple PPs. This study had implications for our understanding of the case system of Udmurt, as it made a novel proposal regarding the adverbial case – or rather, the adverbial cases. I argued that what is at stake here are two types of PPs: an internally complex PP with spatiotemporal semantics and a simple PP with oblique semantics. This allowed us not only to explain the most puzzling question regarding the morphosyntax of the adverbial case, namely the varying morpheme order Cx-Px or Px-Cx, but also the question of how the morphosyntax of these clauses correlates with their meaning (temporal vs. oblique). The *-(e)mja-* and *-(o)nja-*clauses presented the crucial piece of evidence here.

Furthermore, I described *-(o)ńńa-*clauses in the Middle Cheptsa dialect of Udmurt, for which I argued that they have a locative meaning in addition to the temporal one. I proposed that this spatiotemporal meaning can be derived from the "building blocks" that compose these clauses; the crucial part was the presence of the DOMUS suffix. The description of this clause type contributes to our knowledge of Udmurt dialectal syntax, which has been severely understudied.

## Abbreviations

| | | | |
|---|---|---|---|
| 1 | first person | ILL | illative |
| 2 | second person | IMP | imperative |
| 3 | third person | INE | inessive |
| ACC | accusative | INF | infinitive |
| ADD | additive particle | INS | instrumental |
| ADV | adverbial case | NEG | negation |
| CNG | connegative (verb) | NLOC | locative nominalization |
| COP | copula | NMLZ | nominalization |
| CVB | converb | PL | plural |
| DAT | dative | POSS | possessive |
| DOMUS | locative case | PRS | present |
| ELA | elative | PRT | particle |
| EMPH | emphatic (particle) | PST | past |
| EVID | evidential past tense | QUOT | quotative (particle) |
| EXCL | exclusive | SG | singular |
| FUT | future | V | verb |
| GEN | genitive | VN | verbal noun |

## References

AMINOFF, TORSTEN G. 1896. Votjakin äänne- ja muoto-opin luonnos. *Suomalais-Ugrilaisen Seuran Aikakauskirja* 14(2). 1–48.

ARKHANGELSKIY, TIMOFEY & USACHEVA, MARIA. 2015. Syntactic and morphosyntactic properties of postpositional phrases in Beserman Udmurt as part-of-speech criteria. *SKY Journal of Linguistics* 28. 103–137.

ASBURY, ANNA. 2008. *The morphosyntax of case and adpositions.* Utrecht: Utrecht University. (Doctoral dissertation.)

ASBURY, ANNA & DOTLAČIL, JAKUB & GEHRKE, BERIT & NOUWEN, RICK (eds.). 2008. *Syntax and semantics of spatial P.* Amsterdam & Philadelphia: John Benjamins. https://doi.org/10.1075/la.120

Bartens, Raija. 2000. *Permiläisten kielten rakenne ja kehitys*. Helsinki: Suomalais-Ugrilainen Seura.

Blake, Barry. 1994. *Case*. Cambridge: Cambridge University Press.

Cinque, Guglielmo & Rizzi, Luigi (eds.). 2010. *Mapping spatial PPs: The cartography of syntactic structures, Volume 6*. New York: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195393675.001.0001

Comrie, Bernard & Thompson, Sandra A. 2007. Lexical nominalization. In Timothy Shopen (ed.), *Language typology and syntactic description, Vol. 3: Grammatical categories and the lexicon*. 2nd edn. 334–381. Cambridge: CUP. https://doi.org/10.1017/CBO9780511618437.006

Csúcs, Sándor. 2005. *Die Rekonstruktion der permischen Grundsprache.* Budapest: Akadémiai Kiadó.

Dékány, Éva. 2011. *A profile of Hungarian DP: The interaction of lexicalization, agreement and linearization with the functional sequence*. Tromsø: University of Tromsø. (Doctoral dissertation.)

Dékány, Éva & Georgieva, Ekaterina. 2020. Three ways of unifying participles and nominalizations: The case of Udmurt. In Alexiadou, Artemis & Borer, Hagit (eds.), *Nominalization: 50 years on from Chomsky's Remarks*, 169–202. Oxford: Oxford University Press.
https://doi.org/10.1093/oso/9780198865544.003.0008

Dékány, Éva & Hegedűs, Veronika. 2021. Formal and semantic classification. In É. Kiss, Katalin & Hegedűs, Veronika (eds.), *Syntax of Hungarian: Postpositions and postpositional phrases*, 11–191. Amsterdam: Amsterdam University Press.

Edygarova 2010 = Едыгарова, Светлана. 2010. *Категория посессивности в удмуртском языке*. Tartu: University of Tartu. (Doctoral dissertation.)

Edygarova, Svetlana. 2017. Об одной функции адвербиала *-я* в удмуртском языке (как пример языковой стандартизации). *Suomalais-Ugrilaisen Seuran Aikakauskirja* 96. 75–96. https://doi.org/10.33340/susa.70221

Emelyanov 1927 = Емельянов, Аркадий И. 1927. *Грамматика вотяцкого языка*. Ленинград: Издание Ленинградского восточного института им. А. С. Енукидзе.

Fokos-Fuchs, Dávid Rafael. 1958. Die Verbaladverbien der permischen Sprachen. *Acta Linguistica Hungarica* 8. 273–342.

Georgieva, Ekaterina. 2018. *Non-finite adverbial clauses in Udmurt*. Szeged: University of Szeged. (Doctoral dissertation.)

Grünthal, Riho. 2022. Adpositions and adpositional phrases. In Bakró-Nagy, Marianne & Laakso, Johanna & Skribnik, Elena (eds.), *The Oxford guide to the Uralic languages*, 961–969. Oxford: Oxford University Press.
https://doi.org/10.1093/oso/9780198767664.003.0050

Haspelmath, Martin. 1997. *From space to time: Temporal adverbials in the world's languages*. München & Newcastle: Lincom Europa.

Haspelmath, Martin. 2009. Terminology of case. In Malchukov, Andrej L. & Spencer, Andrew (eds.), *The Oxford handbook of case*, 505–517. Oxford: Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199206476.013.0034

Haspelmath, Martin & König, Ekkehard (eds.). 1995. *Converbs in cross-linguistic perspective* (Empirical Approaches to Language Typology 13). Berlin: Mouton De Gruyter. https://doi.org/10.1515/9783110884463

Hegedűs, Veronika. 2014. The cyclical development of Ps in Hungarian. In É. Kiss, Katalin (ed.), *The evolution of functional left peripheries in Hungarian*, 122–147. Oxford: Oxford University Press.
https://doi.org/10.1093/acprof:oso/9780198709855.003.0005

Hewitt, George. 1995. *Georgian: A structural reference grammar*. Amsterdam & Philadelphia: John Benjamins.

Jackendoff, Ray. 1983. *Semantics and cognition*. Cambridge, MA: MIT Press.

Karpova 2005 = Карпова, Людмила Л. 2005. *Среднечепецкий диалект удмуртского языка*. Ижевск: Рос. акад. наук, Урал. отд-ние, Удмурт. ин-т истории, яз. и лит. УрО РАН.

Kelʹmakov 1998 = Кельмаков, Валентин К. 1998. *Краткий курс удмуртской диалектологий*. Ижевск: Издательство Удмуртского университета.

Kelʹmakov, Valentin & Hännikäinen, Sara. 1999. *Udmurtin kielioppia ja harjoituksia*. Helsinki: Suomalais-Ugrilainen Seura.

Kittilä, Seppo & Laakso, Johanna & Ylikoski, Jussi. 2022. Case. In Bakró-Nagy, Marianne & Laakso, Johanna & Skribnik, Elena (eds.), *The Oxford guide to the Uralic languages*, 879–893. Oxford: Oxford University Press.
https://doi.org/10.1093/oso/9780198767664.003.0044

Kortmann, Bernd. 1996. *Adverbial subordination: A typology and history of adverbial subordinators based on European languages* (Empirical Approaches to Language Typology 18). Berlin & Boston: Mouton de Gruyter.
https://doi.org/10.1515/9783110812428

Malchukov, Andrej L. & Spencer, Andrew. 2009. A typology of case systems: Parameters of variation. In Malchukov, Andrej L. & Spencer, Andrew (eds.), *The Oxford handbook of case*, 651–667. Oxford: Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199206476.013.0046

Moravcsik, Edith. 2009. The distribution of case. In Malchukov, Andrej L. & Spencer, Andrew (eds.), *The Oxford handbook of case*, 231–245. Oxford: Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199206476.013.0016

Perevoshchikov 1959 = Перевощиков, Петр Н. 1959. *Деепричастия и деепричастные конструкции в удмуртском языке*. Ижевск: Удмуртское книжное издательство.

Perevoshchikov 1962 = Перевощиков, П. Н. (отв. ред.). 1962. *Грамматика современного удмуртского языка: Фонетика и морфология*. Ижевск: Удмуртское книжное издательство.

Rédei, Károly. 1988. Geschichte der permischen Sprachen. In Sinor, Denis (ed.), *The Uralic languages: Description, history and foreign influences*, 351–394. Leiden & New York: Brill.

Serdobolskaya, Natalia. 2016. Semantics of complementation in Adyghe. In Boye, Kasper & Kehayov, Petar (eds.), *Complementizer semantics in European languages* (Empirical Approaches to Language Typology 57), 691–744. Berlin & Boston: Mouton de Gruyter. https://doi.org/10.1515/9783110416619-020

SERDOBOLSKAYA et al. 2012 = Сердобольская, Наталья В. & Ильевская, Анфиса А. & Минор, Сергей А. & Митева, Полина С. & Файнвейц, Александра В. & Матвеева, Наталья С. 2012. Конструкции с сентенциальными актантами в финно-угорских языках. In Кузнецова, Ариадна И. (отв. ред.), *Финно-угорские языки: Фрагменты грамматического описания, Формальный и функциональный подходы*, 382–476. Москва: Рукописные памятники Древней Руси.

SEREBRENNIKOV 1963 = Серебренников, Борис А. 1963. *Историческая морфология пермских языков*. Москва: Издательство АН СССР.

SIMONENKO & LEONTYEV 2012 = Симоненко, Александра П. & Леонтьев, Алексей П. 2012. Морфосинтаксис именного комплекса в финно-пермских языках: анализ в рамках программы минимализма. In Кузнецова, Ариадна И. (отв. ред.), *Финно-угорские языки: Фрагменты грамматического описания, Формальный и функциональный подходы*, 262–339. Москва: Рукописные памятники Древней Руси.

SVENONIUS, PETER. 2006. The emergence of Axial Parts. *Nordlyd* 33(1). 49–77. https://doi.org/10.7557/12.85

TEPLYASHINA 1970 = Тепляшина, Тамара И. 1970. *Язык бесермян*. Москва: Наука.

TRAUGOTT, ELIZABETH C. & KÖNIG, EKKERHARD. 1991. The semantics-pragmatics of grammaticalization revisited. In Traugott, Elizabeth C. & Heine, Bernd (eds.), *Approaches to grammaticalization I*, 189–218. Amsterdam: John Benjamins. https://doi.org/10.1075/tsl.19.1.10clo

USACHEVA 2012 = Усачева, Мария Н. 2012. Локативные падежи в составе групп с пространственным значением в пермских языках In Кузнецова, Ариадна И. (отв. ред.), *Финно-угорские языки: Фрагменты грамматического описания, Формальный и функциональный подходы*, 141–217. Москва: Рукописные памятники Древней Руси.

USACHEVA & SERDOBOLSKAYA 2015 = Усачева, Мария Н. & Сердобольская, Наталья В. 2015. Грамматикализация новой деепричастной формы на *-оннига-* в бесермянском удмуртском. In Лютикова, Е. А. & Циммерлинг, А. В. & Коношенко, М. Б. (ред.), *Типология морфосинтаксических параметров 2015*, 366–392. Москва: МГПУ.

USACHEVA, MARIIA & SERDOBOLSKAYA, NATALIA. (Forthcoming.) A new converb originating from a locative noun in Beserman. (*Studies in Language*.)

VAN RIEMSDIJK, HENK C. & HUYBREGTS, RINY. 2002. Location and locality. In van Oostendorp, Marc & Anagnostopoulou, Elena (eds.), *Progress in grammar: Articles at the 20th anniversary of the Comparison of Grammatical Models Group in Tilburg*, 1–23. Amsterdam: Meertens Instituut.

WATERS, CATHLEEN. 2009. The preposition cycle in English. In van Gelderen, Elly (ed.), *Cyclical change* (Linguistik Aktuell / Linguistics Today 146), 285–300. Amsterdam: John Benjamins. https://doi.org/10.1075/la.146.16wat

WINKLER, EBERHARD. 2001. *Udmurt*. München & Newcastle: Lincom Europa.

WINKLER, EBERHARD. 2011. *Udmurtische Grammatik*. Wiesbaden: Harrassowitz.

YLIKOSKI, JUSSI. 2003. Defining non-finites: Action nominals, converbs and infinitives. *SKY Journal of Linguistics* 16. 185–237.

*Jaakko Häkkinen*
Sipoo

# On locating Proto-Uralic

In recent years the debate regarding the Proto-Uralic homeland has again intensified. However, not all the relevant arguments have been considered thoroughly. Therefore, in the present article their validity and weight are evaluated. The article also develops further concepts and methodology for reconstruction of stages of Uralic, making it possible to compare Uralic stages to the Indo-Iranian loanword layers with higher resolution than before. As a result, the paper locates Late Proto-Uralic and successive stages in the Central Ural Region, matching the Koptyaki Culture (dated to the early 2nd millennium BCE) and its local predecessor.

## 1. Introduction

The debate on the Proto-Uralic homeland goes back a long way (for earlier research history, see K. Häkkinen 1996: 65–76) and it is still ongoing, the strongest candidates being located on either side of the Ural Mountains, either a European homeland in the Volga-Ural Region or a Siberian homeland in West Siberia (up to the Yenisei). Recent articles supporting the Volga-Ural homeland are Kallio (2006; 2015b), J. Häkkinen (2009), and Parpola (2012b; 2017; 2022). Recent articles supporting the West Siberian homeland are Janhunen (2009), Nichols & Rhodes (2018), Nichols (2021), Grünthal et al. (2022), and Saarikivi (2022). In practice, the actual distance between the candidates for the homeland is sometimes minimal; compare the homeland in the Kama-Ural Region in Parpola (2022: 270, 264) to the homeland reaching to the east from the Central Trans-Urals in Saarikivi (2022: 56).

In this article I intend to go through all the relevant arguments for locating the Late Proto-Uralic homeland. Due to ongoing advances in the fields of Uralic etymology and historical phonology, the criteria are stricter in this critical examination than in many earlier articles – including my own previous attempt on the topic (J. Häkkinen 2009).

It is crucial to define what is relevant evidence and what is not. First, many earlier pieces of evidence have been discarded due to flaws which weaken their evidentiary value. If a word has too narrow a distribution, too irregular sound correspondences between cognates, or the original meaning cannot be reconstructed reliably, it has no value for locating the homeland.

This leads to the second point: the only relevant chronological stage here is Late Proto-Uralic – the moment in time right before the disintegration. Not only are pieces of evidence that are too late discarded, but so are pieces of evidence that are too early. The suggested Ural-Altaic features of an areal-typological nature and possible distant contacts or even relatedness with Indo-European, Yukaghir, or Eskimo-Aleut would in any case precede Late Proto-Uralic by several millennia. These phenomena are just as irrelevant for locating Late Proto-Uralic as the location of Late Proto-Uralic is for locating Late Proto-Finnic (in Estonia) or Late Proto-Samoyedic (in South Siberia).

Arguments for locating the homeland can be divided into two categories: compelling arguments and suggesting arguments. Compelling arguments are undeniable, and only if they cannot help to locate the homeland must we turn to suggesting arguments. Among the compelling arguments are the loanword layers from identifiable and locatable donor languages (Section 2) and the paleolinguistic evidence (Section 3). There are several suggesting arguments, but those deserve a shorter treatment (Section 4).

As byproducts of this critical examination, a new model for the disintegration of Late Proto-Uralic will be proposed (Subsection 5.1), as well as the dating of Late Proto-Uralic and the successive stages of reconstruction (Subsection 5.2). After that, a conclusion is drawn from the locating arguments (Subsection 5.3), followed by a brief review of possible archaeological counterparts (Subsections 5.4 and 5.5).

The new model for disintegration has consequences for the question of which words should be counted as Proto-Uralic. When is the distribution of a word wide enough to be counted as Proto-Uralic? In the present critical examination I apply a criterion that in the west, a cognate must be found in at least Mordvin, Finnic, or Saami, and in the east, a cognate must be found at least in Mansi or Khanty. I will acknowledge the special status of Samoyedic, yet I will argue in Subsection 5.3 for why the presented disintegration model allows a word to not be present in Samoyedic, although in that case a regular cognate in Mansi or Khanty is required. For example, the famous words for 'bee' and 'honey' do not have cognates beyond Hungarian in the east, making them too suspect as Late Proto-Uralic words.

As will be demonstrated later, there is probably more than half a millennium (but less than a full millennium) between the first regional division and the first (macro-)branch-specific sound changes. Such circumstances naturally lead to the existence of a multitude of words which appeared after Late Proto-Uralic but still cannot be distinguished from the Late Proto-Uralic words by the phonological criteria. Such circumstances require applying quite a strict distributional criterion.[1]

---

## 1.1. On notations and labels

In the first syllable, there are eight generally accepted vowel qualities in Proto-Uralic: the front vowels *ä, *e, *i, *ü and the back vowels *a, *o, *u, *e̯ (*i̯) (Aikio 2022: 5). There were only two certain vowel qualities beyond the first syllable, namely *a and *ə, the primary distinction probably being a full vowel vs. a reduced vowel (Kallio 2012: 163–165). Possibly a phonetically front allophone appeared after a front vowel in the first syllable, and a phonetically back allophone after a back vowel. There is also some evidence supporting the possible existence of *o in non-initial syllables (Aikio 2015: 37–38; 2022: 9).

In the Uralic Phonetic Alphabet (UPA), the symbol /ə/ denotes a phonetic value different from the International Phonetic Alphabet (IPA), that is, it is usually a front counterpart of the back /ə̂/, and both of those UPA symbols are semilabial and mid-high (the three vowel heights in the UPA do not neatly correspond to the four vowel heights in the IPA; see Iivonen 2012: 18).

The reconstruction stages following Late Proto-Uralic, as well as other related concepts, will be labeled and explained when they are encountered in the present critical examination. Labels for units at different stages of disintegration following the uniform proto-language are: (1) center = regionally separated unit > (2) pre-dialect = substitutionally differing unit > (3) proto-dialect = phonologically differing unit (see Subsection 5.1).

Datings are specified with the conventional marking "BCE", denoting calendar years before the common calendrical starting point. Datings are rough approximations, which is shown by the use of even centuries. Abbreviations are explained when they first appear, and a list of abbreviations can be found at the end of the article.

The Ural Mountains are generally divided into five regions in Russia and one in Kazakhstan. From the north, the regions are the Polar Urals, the Sub-Polar Urals (the Nether Urals), the Northern Urals, the Central Urals (the Middle Urals), and the Southern Urals. The southernmost region in Kazakhstan is called the Mughalzhar Hills. The most relevant regions for this article are the Central Urals (parallel to the Lower and Middle Kama up to Perm) and the Southern Urals (parallel to the Middle Volga from the Kama fork down to the Kazakhstan border). The label "Trans-Urals" denotes the eastern slopes of the Urals and a narrow strand of lowland adjacent to them.

## 1.2. Different types of borrowing

As loanwords occupy a great share of the following critical scrutiny, an important topic must be clarified first. The distinction between borrowing into the uniform proto-language and into later stages has been generally acknowledged, but it is necessary to increase the resolution here. There are several types of borrowing concerning the situation after the disintegration of the proto-language, and the evidentiary value varies between them. I illustrate the different types with examples from the Indo-Iranian loanwords:

1. **Parallel borrowing**: Difference in sound substitutions between recipient languages, while the original word in the donor language is the same. U *počaw 'reindeer' (> Saami, Finnic?, Mari, Permic) represents substitution with o and U *päčVw/γ 'reindeer calf' (> Mansi, Khanty) substitution with ä of the very same Proto-Iranian *patsu- 'cattle' (Holopainen 2019: 184–185, 196).
2. **Variant borrowing**: Difference between variants of the original word in the donor language. U *s/šȩrńa 'gold' (> Hungarian, Mansi, Khanty) vs. U *serńa 'gold' (> Mordvin) belong to this type, as the latter was borrowed from a different grade of the same Proto-Iranian word, where there was a syllabic resonant instead of a vowel: PIr *dzəranya- vs. PIr *dzr̥Hnya- 'gold' (Holopainen 2019: 232–234).
3. **Separate borrowing**: Different donor languages or different chronological stages of the same donor lineage. U *sa/ora 'lake' (> Hungarian, Mansi, Khanty) vs. Proto-Permic *sariďz 'sea' (Holopainen 2019: 217–219) must represent two donor languages/stages separated by a great gap in time, because the original U *a has changed into the PPe *u (Metsäranta 2020: 94).
4. **Irregular developments** in a recipient language, which cannot be entirely explained by differences between sound substitutions, donor variants, or donor languages, for example U *ćarwi 'horn' vs. PKh *ćerpa 'horn' (Holopainen 2019: 220–222).

Considering the evidentiary value of these types, (1) *parallel borrowing* is solid evidence for the disintegration of the proto-language: if the very same word has been borrowed in parallel, showing different sound substitutions between branches, then we can be certain that the proto-language indeed

disintegrated during the borrowing. However, there is a distributional re-striction for this rule: if one of these substitutions has a wide enough dis-tribution in Uralic, it still can be reconstructed for Late Proto-Uralic, as the other substitution with narrow distribution could be later. In such case we would actually be dealing with separate borrowings.

(2) *Variant borrowing* is almost as strong evidence for the disintegra-tion of Late Proto-Uralic, because it requires at least two centers within the proto-language speech community. This type of borrowing can be consid-ered the second strongest evidence pointing to the already disintegrated proto-language. The same distributional restriction applies here as in the previous type.

In the case of (3) *separate borrowing*, when the words are from different points in time, we cannot exclude the possibility that the older or more widespread loanword was earlier known also in the other language, until it became replaced by the younger rather similar-looking loanword. There-fore, this type is somewhat weaker evidence for the disintegration of Late Proto-Uralic.

(4) *Irregular developments* occurring in an individual language or branch after the borrowing event cannot testify against the status of the word in the uniform proto-language, although it certainly can distort our attempts to reconstruct the word in the proto-language. Irregular cognates in some branches cannot diminish the value of regular cognates in other branches: if those branches are distant enough in the taxonomic model of the language family, the word can be reconstructed for the proto-language.

## 2. Early Indo-European loanword layers

Within the Early Indo-European loanword layers, I include early stages of two separate donor lineages, which descend from Late Proto-Indo-Euro-pean (LPIE): the Indo-Iranian lineage consisting of Early (EPIIr) > Middle (MPIIr) > Late Proto-Indo-Iranian (LPIIr) > Proto-Iranian (PIr), and the Northwest Indo-European lineage consisting of Archaic Indo-European (AIE) > Northwest Indo-European (NwIE). At the moment the number of convincing loanwords in the Early Proto-Indo-Iranian layer is about a dozen, and greater still in the Late Proto-Indo-Iranian and Proto-Iranian layers (Grünthal et al. 2022: Appendix 2 "Indo-Iranian loans in Uralic"). There are several dozen proposed Archaic and Northwest Indo-European loanwords, but their reassessment is still an ongoing process.

## 2.1. Indo-European and Uralic

Since the generally accepted rejuvenation of Late Proto-Uralic (Kallio 2006; J. Häkkinen 2009), this stage is no longer contemporaneous with Late Proto-Indo-European. That does not disqualify the earliest proposed Indo-European loanwords in Uralic, it only requires a new label for them. I use the label Archaic Indo-European as an umbrella term for the loanwords resembling LPIE with wide distribution in Uralic, and the label Northwest Indo-European for the (usually likewise LPIE resembling) loanwords, which have a western (Finno-Permic) distribution in Uralic and cognates mainly in the Northwest Indo-European branches. Most if not all of the Archaic Indo-European loanwords could probably be explained as borrowed from the ancestor of Northwest Indo-European.

Northwest Indo-European is not a proto-dialect in the same sense as Late Proto-Indo-Iranian, but rather a continuum of phonologically conservative Indo-European varieties roughly corresponding to the wide area of the Corded Ware Cultures. Northwest Indo-European branches (at least Balto-Slavic, Germanic, Celtic, and Italic) share quite a lot of words that lack cognates in more distant branches. Many such words also show phonological or phonotactic features foreign to Late Proto-Indo-European and tend to denote local flora, fauna, and livelihoods, which points towards a substrate origin from unknown ancient languages (Mallory & Adams 2006: 78–79).

It has also been noted that while Anatolian, Indo-Iranian, and Greek must be assumed to have gone through some branch-specific sound changes already during the 3rd millennium BCE, in the northwest, the phonological distinctions occur only during the 2nd millennium BCE (Mallory & Adams 2006: 103–104). Even the loanwords borrowed into Uralic branches seem to testify that distinguishably Balto-Slavic and Germanic phonological features appear later than the recognizable Indo-Iranian features (Kallio 2009: 38–40; forthcoming).

In principle, both of these Early Indo-European donor lineages weigh heavily on locating Proto-Uralic. In practice, however, we suffer from limitations in the resolution between the reconstruction stages and in the quality of the loanwords. These topics are considered in Subsections 2.3 and 2.4.

## 2.2. Changing views in the Uralic studies

During the late 20th century, in the Uralic studies there prevailed the so-called Moderate Continuity Theory, in which the arrival of Finnic and Saami in the Baltic Sea Region was connected to the spread of the Typical Combed Ware ca. 4000 BCE, and (Late) Proto-Indo-European was seen as roughly contemporaneous with (Late) Proto-Uralic. In that framework, the Indo-Iranian loanword layers were seen as clearly later than Proto-Uralic (on the research history, see Aikio & Aikio 2001).

At the beginning of the current century, the accumulating evidence especially from Germanic loanwords led to a later dating for the phonological divergence between Finnic and Saami (Koivulehto 2002; Aikio 2006; Kallio 2009; 2015a). At the same time, discontent towards the traditional Uralic taxonomic model also grew, i.e. the family tree in which Samoyedic was the first branch to split away (K. Häkkinen 1984; Salminen 1989; 2002; J. Häkkinen 2007).

It was also more pronouncedly emphasized that linguistic continuity could not be reliably tracked from archaeological continuity (Aikio & Aikio 2001; Mallory 2001; J. Häkkinen 2010). The utter unreliability of the continuity argument was finally revealed by gravely contradicting datings achieved by the very same method from the very same data, when Proto-Uralic was claimed to have been spoken in Finland already right after the Ice Age (on the history of this scientific debate in Finland, see Tirkkonen 2012).

Together all these factors led to a paradigm shift. When Samoyedic no longer had veto power to dismiss words from being Late Proto-Uralic, it was possible to consider even the Late Proto-Indo-Iranian loanwords as being borrowed into Late Proto-Uralic. Consequently, earlier datings were rejected and later datings around 2000 BCE were accepted for Late Proto-Uralic (Kallio 2006; J. Häkkinen 2009; Parpola 2012b). The level of phonology (enough regular cognates in the Uralic branches) subdued the level of distribution as the paramount dating criterion.

However, it has been recently demonstrated that distribution cannot be so easily overruled when the number of convincing loanwords in a layer is high enough. That is, when there appear to be more than a dozen Early Proto-Indo-Iranian loanwords in Uralic and none of these has a cognate in Samoyedic, it becomes highly improbable that all the words could have simply disappeared since Late Proto-Uralic (Grünthal et al. 2022: 10). At the same time, the stage where a shift occurs from regular cognates to

parallel borrowings appears to be much later, occurring only around the Proto-Iranian stage (Holopainen 2019: 343).

Neither of these arguments can be explained away or ignored. The reason for the apparently contradictory results must lie in the fundamental difference between the lexical and phonological levels. Therefore, based on the evidence from the Indo-Iranian loanword layers, I will construct a new model for the disintegration of Late Proto-Uralic, presented in Subsection 5.1.

## 2.3. Resolution of the donor lineages

By resolution I mean the density of successive distinguishable reconstruction stages in a donor lineage, seen in the loanwords through the Uralic filter. In a donor lineage, there might have occurred changes concerning word-initial consonant clusters or voiced obstruents, but many such changes would remain invisible due to the restrictions caused by Uralic phonology and phonotactics: an initial cluster would have been substituted by a single consonant, and a voiced obstruent would have been substituted by a voiceless obstruent. Similarly, the presence of the Indo-European palatovelars and labiovelars in a donor lineage would be difficult to recognize reliably (Holopainen 2021).

For the taxonomy of Proto-Indo-European, I follow the well-argued consensus view, according to which Anatolian was the first branch to split away, followed by Tocharian (Jasanoff 2017; Ringe 2017). The remaining core is here called Late Proto-Indo-European (LPIE), following Anthony & Ringe (2015: 201).

Concerning the disintegration of Late Proto-Indo-European, centumization and satemization are no longer considered clade-defining changes: the first occurred independently across different branches, and the second has spread secondarily (Ringe 2017). There are more exceptions to satemization in Balto-Slavic than in Indo-Iranian, which points to its secondary spread (Kim 2018: 1975). In the position after *s LPIE *$\hat{k}$ was depalatalized to *k before satemization in Balto-Slavic (Matasović 2005: 148), showing that satemization was not the earliest Balto-Slavic sound change.

The *ruki*-rule in Indo-Iranian was triggered also by the secondary *i developing from an earlier syllabic laryngeal and the secondary *r from *l (Lubotsky 2018: 1881), so it was not among the earliest Indo-Iranian sound changes. Also, the results in Nuristani differ from those in Iranian and Indic (Hegedűs 2012). In Balto-Slavic the *ruki*-rule is regular only in Slavic, while

in Baltic it is more restricted (Kim 2018: 1976), and therefore it seems to have spread there only after the disintegration of Proto-Balto-Slavic. In Balto-Slavic the *ruki*-rule is also later than satemization (Matasović 2005: 148).

Interestingly, some Baltic loanwords in Finnic show the *ruki* reflex even though the modern East Baltic languages do not, e.g. LPFi *\*laiha* 'thin, slender' < MPFi *\*lajša* ← dialectal Balto-Slavic *\*laišas* ~ Lithuanian *líesas* 'thin' < *\*laisas* (Kallio 2008: 267). Due to the secondary nature of this change, it is natural to assume that *\*laišas* would represent some eastern Balto-Slavic dialect spoken closest to Indo-Iranian, while *\*laisas* could represent a more central Balto-Slavic dialect (Proto-Latvo-Lithuanian?). There are possible traces of an even more diverse continuum of Balto-Slavic varieties than has been recognized thus far, based on recurring irregularities in the loanwords borrowed into the West Uralic branches (J. Häkkinen 2022).

Archaic Indo-European and Northwest Indo-European were still phonologically very similar to Late Proto-Indo-European, at least as far as we can see through the Uralic filter. However, there seem to be no certain examples of preserved palatovelars in the loanwords borrowed into Uralic (Holopainen 2021: 199). This could point to post-Proto-Indo-European donor languages, as the palatovelars either merged with the plain velars (in *centum*-dialects) or changed to palatalized affricates or sibilants (in *satem*-dialects). As Germanic has gone through centumization and Balto-Slavic through satemization, there could be ancient loanwords from both types of dialects in Uralic. However, it is questionable whether we could distinguish even the centumized or satemized consonants from the Late Proto-Indo-European consonants through the Uralic filter:

1. LPIE *\*ḱ*, *\*ǵ*, and *\*ǵʰ* could have been substituted by either U *\*k* or U *\*ć*, and LPIE *\*kʷ*, *\*gʷ*, and *\*gʷʰ* could have been substituted by U *\*k* or *\*ku*.
2. The centumized *\*k*, *\*g*, and *\*gʰ* (< *\*ḱ*, *\*ǵ*, and *\*ǵʰ*) would have been substituted by U *\*k*, and the preserved *\*kʷ*, *\*gʷ*, and *\*gʷʰ* could have been substituted by U *\*k* or *\*ku*.
3. The satemized *\*ć*, *\*ź*, and *\*źʰ* would have been substituted by U *\*ć*, and the *\*k*, *\*g*, and *\*gʰ* (< *\*kʷ*, *\*gʷ*, and *\*gʷʰ*) would have been substituted by U *\*k*.

Consequently, even if there were centumization- or satemization-related sound changes in an Indo-European donor language, we could not, in the absence of other branch-specific sound changes to guide our interpretation, reliably distinguish them from the Late Proto-Indo-European stage.

Moreover, even if there were loanwords from some early *centum-* or *satem-* dialects adopted into Uralic, their accurate dating would still be practically impossible to determine: the centumization- and satemization-related sound changes could have occurred in different Indo-European branches at different times, from right after the Late Proto-Indo-European stage (around or before 3000 BCE) to much later dates (around or after 2000 BCE).

Because this time span also includes the disintegration of Late Proto-Uralic, such a lack of resolution prevents us from estimating whether those kinds of loanwords were borrowed into Late Proto-Uralic or already into separate Uralic pre-dialects, and therefore their value for locating Proto-Uralic would be gravely diminished. Temporally relevant would be only those loanwords which have been borrowed from a datable reconstruction stage, and among those, spatially relevant would be only those loanwords which would immediately precede the dispersal of Late Proto-Uralic.

The chronological resolution is high only in the Indo-Iranian lineage (see Subsection 2.5), while in the Northwest Indo-European lineage the lack of distinguishable sound changes (visible through the Uralic filter) continues up to the 2nd millennium BCE, making that lineage practically worthless for locating Late Proto-Uralic (Figure 1; for the chronology, see Subsection 5.2).

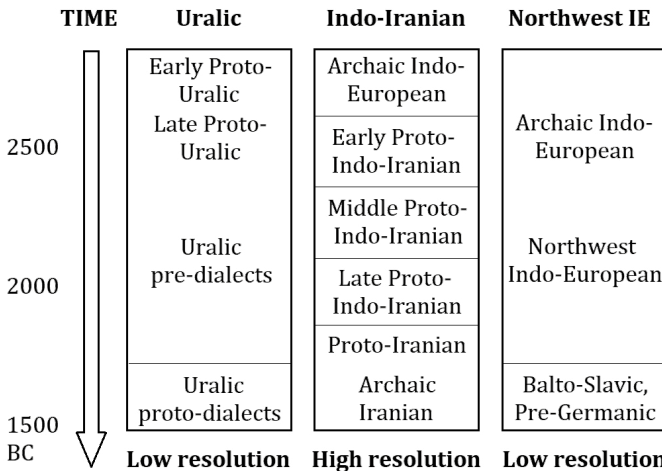| TIME | Uralic | Indo-Iranian | Northwest IE |
|---|---|---|---|
| | Early Proto-Uralic | Archaic Indo-European | |
| | Late Proto-Uralic | Early Proto-Indo-Iranian | Archaic Indo-European |
| 2500 | | | |
| | | Middle Proto-Indo-Iranian | |
| | Uralic pre-dialects | Late Proto-Indo-Iranian | Northwest Indo-European |
| 2000 | | | |
| | | Proto-Iranian | |
| | Uralic proto-dialects | Archaic Iranian | Balto-Slavic, Pre-Germanic |
| 1500 BC | | | |
| | Low resolution | High resolution | Low resolution |

Figure 1: Successive reconstruction stages are not always phonologically distinguishable; only those which are separated by a thin horizontal line are. In the Indo-Iranian lineage the resolution is highest, but between the Uralic and the Northwest Indo-European lineages it is difficult to date the loanwords precisely.

At the moment, there is no reliable method to distinguish loanwords adopted into Late Proto-Uralic from loanwords adopted earlier or later, and consequently I am forced to dismiss the Northwest Indo-European lineage from this scrutiny.

## 2.4. Quality of loanwords

There are several dozen proposed loanwords which could be borrowed from Late Proto-Indo-European, Archaic Indo-European, or Northwest Indo-European (e.g., Koivulehto 1991; 2001), so the existence of such loanword layers does not suffer from the lack of quantity but from the possible lack of quality. There are recent critical assessments about the words with Uralic *$\check{s}$ as the assumed substitute for the Indo-European laryngeals (Hyllested 2014), about the most widespread Indo-European loanwords in Uralic (Simon 2020), and about the words containing alleged Late Proto-Indo-European palatovelars (Holopainen 2021). Many of the assumed early Indo-European loanwords have already been proven to be improbable, but there are still plenty of proposed loanwords waiting to be assessed more thoroughly.

We must bear in mind that even if some loanwords showed unexpected sound correspondences, it would not automatically make them false. Some phonological mismatches could be caused by phonological developments occurring already before Late Proto-Indo-European or Late Proto-Uralic, so they could reflect very ancient contacts. Other mismatches could reflect a lost dialect of some Indo-European branch or just an arbitrary and unexpected sporadic sound substitution. If there are multiple occurrences of such an unexpected sound correspondence (becoming a recurring irregularity), it becomes more probable that there is some real phenomenon behind them. As an example, there are at least two occasions showing unexpected U *$\acute{c}$ reflecting LPIE/LPIIr *$s$, which I here suggest possibly belong to the Late Proto-Indo-Iranian loanword layer:

U *poćə 'penis' >
    Saami *puoćę ~ Hungarian *fasz* (Sammallahti 1988: 548)
    ← LPIIr *pásas (> Sanskrit *pásas-*; Holopainen 2019: 185) < LPIE *péses 'penis' (Mallory & Adams 2006: 183–184)

U *moćkə- 'wash' > Finnic, Mordvin, Mari, Permic, Hungarian, Samoyedic (Sammallahti 1988: 538)

   ← LPIIr *mazg- (> Sanskrit májjati 'sinks') < LPIE *mesg- 'dip underwater, dive' (Mallory & Adams 2006: 403; Pokorny 2007: 2107); cf. Balto-Slavic *mazgo- 'wash' < *mosg- (Derksen 2015: 308)

Both of these words could represent the well-known *o-substitution for the Late Proto-Indo-Iranian *a (see Holopainen 2019: 49–50). The second word has earlier been considered a Proto-Indo-European loanword, but the Uralic *ć was problematic without any known further examples (Simon 2020: 248). This word could also come from Balto-Slavic, but it is more probable that both of these words come from the same donor language, namely Late Proto-Indo-Iranian. There are also other loanwords in this layer for which a cognate is present in Samoyedic (see Subsection 2.6.2).

One possible explanation for the unexpected substitution with *ć is connected to the fact that Uralic *ć appears frequently in the Late Proto-Indo-Iranian loanwords as the substitute for *ć, *ȝ, and *ȝʰ. Perhaps that sound was therefore associated with that particular donor language and hypercorrectly appeared even in some words where the donor language had plain *s?

My original intention was to include in this scrutiny also loanwords from the Northwest Indo-European donor lineage, but after consulting with Luobbal Sámmol Sámmol Ánte (Aikio, personal communication), I concluded that a critical reassessment of the quality of these loanwords has only begun. Another reason for omitting these loanwords was the low resolution between the Uralic and the Northwest Indo-European donor lineages (see Subsection 2.3).

Moreover, the evidentiary value of Archaic Indo-European loanwords would in any case be weak. Even though Late Proto-Indo-European and Northwest Indo-European were spoken in Europe, the picture is complicated by the location of Pre-Proto-Tocharian. This eastward expansion is generally connected to the movement from the European steppe to South Siberia, where the Afanasyevo Culture was formed (Anthony & Ringe 2015: 209). As this movement is dated already to the 4th millennium BCE, the Archaic Indo-European spoken there could easily explain loanwords into Early or Late Proto-Uralic, if these stages were spoken in South Siberia.

Therefore, no decisive evidence could be gained from the Archaic Indo-European loanwords as long as we cannot reliably demonstrate that they belong to the Northwest Indo-European lineage instead of the Tocharian lineage, or at least demonstrate a temporal gap between the AIE and the NwIE loanword layers being so short that it would require also a regional vicinity and make it probable that they represent successive stages of the very same NwIE lineage.

## 2.5. Reconstruction stages of the Indo-Iranian lineage

The early part of the Indo-Iranian lineage is here divided into four re-construction stages: Early (EPIIr), Middle (MPIIr), and Late Proto-Indo-Iranian (LPIIr), as well as Proto-Iranian (PIr). The following list contains mainly sound changes distinguishable through the Uralic filter, as seen in the Indo-Iranian loanwords borrowed into Uralic, so the list is not comprehensive. The sound changes are taken from Ollett (2014), Cantera (2017), and Lubotsky (2018):

1. Early Proto-Indo-Iranian:
    1.1. Interconsonantal $*h > *i$
    1.2. Brugmann's Law: $*o > *\bar{o}$ in open syllables
    1.3. Laryngeal coloring: $*e+h_2 > *a+h_2$, $*e+h_3 > *o+h_3$
    1.4. $*l > *r$
2. Middle Proto-Indo-Iranian:
    2.1. Satemization: $*\acute{k}$, $*\acute{g}$, $*\acute{g}^h > *\acute{c}$, $*\acute{\zeta}$, $*\acute{\zeta}^h$, while $*k^w$, $*g^w$, $*g^{wh}$ merge into $*k$, $*g$, $*g^h$
    2.2. The *ruki*-rule: $*s > *\check{s}$ next to $r, u, K, i$ (also the secondary $*i$ and $*r$)
    2.3. Palatalization of the velar stops before the remaining $*e$: $*k, *g, *g^h > *\check{c}, *\check{\zeta}, *\check{\zeta}^h$
3. Late Proto-Indo-Iranian:
    3.1. The merger of non-high vowels and syllabic nasals: $*a, *e, *o, *\mathring{m}$, $*\mathring{n} > *a$
    3.2. The merger of remaining laryngeals into $*H$
4. PIr: The depalatalization of affricates $*\acute{c}$, $*\acute{\zeta} > *ts$, $*dz$

Satemization is rather difficult to date. It must precede the stage 2.3, but it could also be somewhat earlier (but not the earliest; see Subsection 2.3). Even though the laryngeal coloring on an adjacent short $*e$ ($*e+h_1 > *e+h_1$; $*e+h_2 > *a+h_2$; $*e+h_3 > *o+h_3$) occurred in every Indo-European branch and is therefore often considered already a Proto-Indo-European development

(Byrd 2018: 2063–2064), other instances leading to similar results are clearly later. Greek shares with Armenian and probably Phrygian similar coloring of a word-initial laryngeal before a consonant (*$h_1$C > *eC; *$h_2$C > *aC; *$h_3$C > *oC), but only in Greek do we see a similar change concerning the syllabic laryngeal and the word-final laryngeal (*$h_1$ > *e; *$h_2$ > *a; *$h_3$ > *o; Beekes 2011: 146–153). These cases show that such colorings have occurred independently more than once, when the conditions (the coloring quality of the laryngeals) still remained.

Even the oldest type, the laryngeal coloring of short *e, could be later than Late Proto-Indo-European. It has been argued that in Indo-Iranian it seems to be younger than Brugmann's Law (*o > *ō in open syllables; Lubotsky 1990; 2018: 1877). Ollett (2014) has shown that in Indo-Iranian the laryngeal coloring preceded the palatalization of the velar stops, but he could not decisively refute that Brugmann's Law came first. According to Lubotsky (2018: 1877), the change of interconsonantal *H > *i must be even older than Brugmann's Law in Indo-Iranian, at least in the final syllable.

Therefore, I place the laryngeal coloring within the Early Proto-Indo-Iranian stage. This leaves room for the option that also in the Northwest Indo-European lineage the laryngeal coloring could be partially a later phenomenon, which should be taken into consideration when assessing possible Archaic or Northwest Indo-European loanwords in Uralic. If we could stratify the convincing loanwords into pre- and post-coloring stages, the resolution in that donor lineage would increase.

Interestingly, among the loanwords into Uralic, there are no certain examples of the remaining *e with reflexes of the new palatalized affricates (*če-, *ǯe-, *ǯʰe), so there might be a gap in the contacts during the Middle Proto-Indo-Iranian stage. However, there are also only three more or less convincing Early Proto-Indo-Iranian loanwords with *ke- in Uralic, so this gap could be only illusory, caused by the low number of loanwords beginning with these secondary MPIIr affricates in the first place. Another possible explanation is that the Late Proto-Indo-Iranian vowel merger occurred very soon after the palatalization of velars.

Yet there is at least one possible loanword before the vowel merger: Late Proto-Finnic *herä- < PrePFi *šera- 'wake up' ← MPIIr *Hǯer- < LPIE *$h_1$ger- 'awake' (Holopainen 2019: 258). There are competing etymologies, and Uralic *š is not considered the most expected substitute (because Uralic also had a consonant *č), but there are parallel examples for the *š-substitution of an initial Indo-Iranian affricate:

West Uralic *\*šukta* 'swidden field'
    ← PIr *\*tsuxta* 'burned' (Holopainen 2019: 264–265)
LPFi *\*hadas* < PrePFi *\*šatas* 'sprout, germ (of seed)'
    ← PIr *\*dzaHta-*, verbal adjective from the root *\*dzanH-* 'be born, grow'
    (could also be Germanic; Holopainen 2019: 257–258)
Possibly U *šęrńa* 'gold'
    ← PIr *\*dzəranya-* 'gold' (Holopainen 2019: 232–234; other initial sibi-
    lants are possible for this Uralic word, see Subsection 2.5.2).

## 2.6. Connecting the Uralic and the Indo-Iranian reconstruction stages

### 2.6.1. Evidence from the lexical level

It has long been known that Samoyedic seems to be the lexical outlier within the Uralic language family. Samoyedic seemed to share the smallest number of words with all the other branches, and this situation was modeled as a division between Finno-Ugric and Samoyedic, leading to the stratification into the earlier Proto-Uralic (cognate present in Samoyedic) vs. the later Proto-Finno-Ugric layer (no cognate present in Samoyedic) in vocabulary (Sammallahti 1988).

However, the lexical level alone cannot reliably testify to the disintegration of Late Proto-Uralic, because there are other possible explanations as to why some branches appear to share less or more inherited words with other branches than expected. Historical phonology is a more reliable level and leaves less room for different interpretations (J. Häkkinen 2012).

Nevertheless, concerning the Indo-Iranian loanwords, it seems remarkable that in the Early Proto-Indo-Iranian loanword layer (14 words), there are no words with a cognate in Samoyedic (Grünthal et al. 2022: Appendix 2 "Indo-Iranian loans in Uralic"). It is statistically highly improbable that Samoyedic would have first borrowed and later lost all the Early Proto-Indo-Iranian loanwords. It seems necessary to assume that Samoyedic was already outside the contact zone during the earliest distinguishable Indo-Iranian contacts. However, Samoyedic surprisingly participated in contacts during the Late Proto-Indo-Iranian loanword layer (see the following subsection).

### 2.6.2. Evidence from the sound substitutions of loanwords

As was presented in Subsection 1.2, parallel borrowings are the strongest evidence against the uniform proto-language. This works also the other way around: the lack of parallel borrowings bears evidentiary value in the cases where we have several possible sound substitutions. It is well known that there were four possible substitutes for Late Proto-Indo-Iranian *a on the Uralic side: *a, *ä, *o, and *ę (Holopainen 2019: 49–50). No phonological conditions have been found to explain the choice of the vowel – it appears to have been purely arbitrary (Holopainen 2019: 327–328). As LPIIr/PIr *a is also very frequent in loanwords, it works perfectly as a diagnostic vowel. In the LPIIr layer we still see many loanwords which have a wide distribution and in which all the Uralic branches agree with the same vowel substitution – even Samoyedic, when there is a cognate:

U *pęŋka 'psychedelic mushroom' >
    Mordvin *paŋgâ ~ Mari *poŋgâ ~ Mansi *pīŋk ~ Khanty *pāŋk ~ Samoyedic *pęŋkå (Aikio 2015: 59)
    ← LPIIr *b(ʰ)anga- 'narcotic plant' > Middle Persian *bang*, *mang* 'henbane (*Hyoscyamus niger*)' (Holopainen 2019: 186–188)

U *ćęra, *ćęr|kə >
    Saami *ćuorē 'light gray (of reindeer hair)', *ćuorkę-dē 'gray (of human hair)' ~ Samoyedic *si̯rå 'snow', *sęr 'white, ice' (Aikio 2020: 125–126)
    ← LPIIr *ćar- (> Sanskrit śārá- 'colored') < LPIE *k̂er- 'grayish blue/green' (Mallory & Adams 2006: 333; Pokorny 2007: 1547–1548 includes words like Armenian 'snow' and Slavic 'hoarfrost'). New proposition; earlier loan etymologies have been discarded by Holopainen (2018: 157–158; 2019: 231–232).

U *ćęta '100' >
    Saami *ćuotē ~ Finnic *sata ~ Mordvin *śadâ ~ Mari *šüðə ~ Permic *śo ~ Hungarian *száz* ~ Mansi *si̯t ~ Khanty *sāt (Aikio 2015: 60)
    ← LPIIr *ćatá-m '100' (Holopainen 2019: 242–244)

U *kęnta(w) 'log, fallen tree' >
    Saami *kuontī ~ Finnic *kanto ~ Mordvin *kandâ ~ Mansi *ki̯ntā ~ Khanty *kānt (Aikio 2015: 59)
    ← LPIIr *skandʰá- > Old Indic *skándhas-* 'twig, branch' (Holopainen 2019: 120)

U *tora-* 'fight' >
   Saami *toarₔ-* ~ Finnic *tora-* ~ ?Mordvin *t́uŕə-* ~ Samoyedic *tåro-*
   (Aikio 2015: 62)
   ← LPIIr *tara-* 'overcome' (Holopainen 2019: 282–285)

U *ćoma-* 'be starving, get tired' >
   Mordvin *śumₔ-* ~ Mari *šŭma-* ~ Permic *śuma-* ~ Hungarian *szom-*
   (Aikio 2015: 61)
   ← LPIIr *ćamH-* 'be extinguished, die out' (Holopainen 2019: 213)

U *onća* 'part, share' >
   Finnic *osa* ~ PSa *oańćē* ~ Mari *užaš* ~ Hungarian *ágyék* ~ Mansi *ūńć*
   (Aikio 2015: 61)
   ← LPIIr *Hanća-* 'share of a fortune, loot' (Holopainen 2019: 170–171)

U *ora* 'awl' >
   Saami *oarē* ~ Finnic *ora* ~ PMd *urₔ* ~ Hu *ár* (Aikio 2015: 61)
   ← LPIIr *HāŕāH-* 'goad' (Holopainen 2019: 163–164)

U *ćaδa-* 'rain' >
   Finnic *sata-* ~ Samoyedic *sårå-* (Aikio 2015: 56)
   ← LPIIr *ćad-* 'fall' (Holopainen 2019: 224)


It is highly improbable that of the four available substitutions for the Indo-Iranian *a*, all separate and already distinct Uralic branches would have independently chosen the same one in every single loanword. Admittedly, only four of these nine words have cognates in Samoyedic, in which the original *a* and *o* cannot always be distinguished, but we can still distinguish from these sounds the original *e̮* and *ä*. All four of these Samoyedic cognates agree with the rest of Uralic. Consequently, the evidence points to a borrowing situation preceding a wider regional dispersal of Uralic. There must have still existed a narrow Uralic speech community during the borrowing of these Late Proto-Indo-Iranian loanwords.

However, even though parallel borrowing from Indo-Iranian seems improbable, a possibility of borrowing from one Uralic pre-dialect to another should also be considered. Nevertheless, the more words there are with a wide distribution and regular cognates, the more improbable it is that all of them could be due to intra-Uralic borrowing. These nine words cannot be dismissed by this explanation.

During the slightly later Proto-Iranian loanword layer, separate regional pre-dialects already existed. For Proto-Iranian *a, we find different substitutions in different branches, each of them usually having quite a narrow distribution:

U *keɽtańa 'iron' > Mari, Mordvin?
Parallel borrowing U *kärta > Permic, Khanty?
    ← PIr *kártana- 'a cut' or PIr *kr̥tí- 'dagger, knife' (Holopainen 2019: 121–125)

U *počaw 'reindeer' > Saami, Finnic?, Mari, Permic
Parallel borrowing U *päčVw/γ > Mansi, Khanty
    ← PIr *patsuka 'livestock' (Holopainen 2019: 184–185, 196)

U *saŋka/*soŋka 'old' > Mari, Hungarian (ambiguous borrowing: both *a and *o are possible for Mari and Hungarian)
    ← PIr *sanaka 'old' (Holopainen 2019: 235)

U *s/šeɽńa 'gold' > Hungarian, Mansi, Khanty
Possible parallel borrowing U *ć/šeɽńa > Mari
Variant borrowing U *serńa > Mordvin
    ← PIr *dzŕ̥Hnya- ~ *dzəranya- 'gold' (or even AIr/OIr *zaranya?) (Holopainen 2019: 232–234)

The last example does not represent PIr *a, but it also shows variation between different pre-dialects. In Mari, the Malmyž dialect (MariM) has best preserved the original *s, while in the other dialects the *s has coalesced into *š (Metsäranta 2020: 36). However, in *Tscheremissisches Wörterbuch* (Moisio & Saarinen 2008) the closest dialect to Malmyž is marked as Ok (Bolshoy Kiľmez), but it only has s in some of those words (and š in others), in which we find Malmyž s in the *Mari Nyelvjárási Szótár* (Beke 1997–2001). It seems that the Malmyž dialect in the latter dictionary has best preserved the original *s, and the Bolshoy Kiľmez dialect in the first dictionary second best.

In Beke's dictionary, we find Malmyž soŋgo 'old' (page 2449) and Malmyž šörtńö 'gold' (page 2492), the latter pointing towards either the original *ć or *š, being in any case an unexpected reflex. As *s and *š have coalesced in all the East Uralic branches, theoretically we could derive also them from U *šeɽńa. In Subsection 2.5, other examples showing U *š as the substitute for the Indo-Iranian initial affricates were considered.

### 2.6.3. Evidence from sound changes in different branches

The first sound changes appeared at the macro-branch level, although it must be noted that it was no longer a question of a uniform language but instead shared innovations between already regionally separate yet adjacent pre-dialects of the Uralic speech community. Especially diagnostic are the East Uralic sibilant changes shared by Hungarian, Mansi, Khanty, and Samoyedic (drafted in J. Häkkinen 2007: 71–73; reply to recent critique in J. Häkkinen 2023):

1.  U *s and *š merge into *š
2.  U *š > *ʟ (voiceless lateral fricative) or *ϑ (voiceless dental spirant)
3.  U *ć > *s

This chain of three subsequent changes in a certain order makes it practically impossible to assume independent development in different branches or later spreading from one branch to another. These changes must have occurred in a narrow area within a short time span (for wider argumentation, see J. Häkkinen 2023). Based on the previous subsection, the participants were already neighboring pre-dialects when these sound changes occurred, because the loanwords showing different substitutions between branches have also gone through the East Uralic sound changes. Two examples seen already in the previous subsection are the following:

U *saŋka/*soŋka 'old' > EU *ʟaŋka/*ʟoŋka >
    Hungarian *agg* 'old'

U *s/šẹrńa 'gold' > EU *ʟẹrńa >
    Hungarian *arany* 'gold' ~ Mansi *tarəń 'copper' ~ Khanty *ʟarńə 'copper'.

Similarly, the westernmost branches Saami, Finnic, and Mordvin seem to share some sound changes. However, the current evidence points also here to shared innovations between already regionally separate but still adjacent centers or pre-dialects. First, the merger of *δ̆ into *δ intervocalically was possibly shared also by Mari, although it is also possible that the development of the two spirants in Mari differed from each other (Metsäranta 2020: 39). Uralic *ẹ usually yields Mari *ü, but sometimes it behaves like *a,

which changed into Late Proto-Mari *o regularly before velar consonants and occasionally in some other environments (Metsäranta 2020: 81, 314). However, as also Uralic *o has been preserved in Mari before the velar nasal, there the change could have been directly *ẹ > *o.

Even the three westernmost branches do not fully agree on the development of Uralic *ẹ: even though they share the changes *aj–ə > *ij–ə and *ẹ-a > *a–a (in Finnic *ẹ and the *a have totally merged), only Saami and Mordvin seem to share a later merger of *a–ə into *o–a (Aikio 2015: 39). However, it is to be expected that also pairwise changes appeared. Similarly, showing different pairwise distributions, Finnic shares with Mordvin the merger of the word-final *m into *n, and Finnic shares with Saami certain developments concerning round vowels in the second syllable (Kuokkala 2018).

Here is one example from the previous subsection showing that the West Uralic changes were later than the Late Proto-Indo-Iranian loanwords:

U *ćẹta '100' > WU *ćata >
  Saami *ćuotē ~ Finnic *sata ~ Mordvin *śadâ

It is also possible that some loanwords with Proto-Iranian *a were originally substituted by Uralic *ẹ, which then changed into West Uralic *a, but we cannot know for certain, because at that stage there probably were no longer loanwords having regular cognates both in the western and other Uralic branches:

(U *wẹsa 'calf' >) WU *vasa >
  Finnic *vasa ~ Mordvin *vaz
  ← Iranian *wasá 'calf' (Holopainen 2019: 300–301)

There are also later Iranian loanwords borrowed even after some of the branch-specific sound changes, and often also the Iranian original is clearly younger and easy to distinguish from the Proto-Iranian word. Here I give only few examples:

Samoyedic *pu/ilə 'bridge'
  ← Middle Iranian *puhl 'bridge' < LPIIr/PIr *pr̥tu- (Holopainen 2019: 195). If this word was borrowed before the Samoyedic sound changes, it would have been **puj in Late Proto-Samoyedic.

Mansi *\*širγV* 'sword'
  ← Iranian (Alanic) *\*cirγ* 'sword' < PIr *\*tiγra-* (Holopainen 2019: 259). Late Proto-Uralic *\*i* > *\*ä* in Mansi, so this borrowing is clearly later than the Mansi vowel changes. The initial consonant cannot reflect the Proto-Iranian stage.

Permic *\*das* '10'
  ← Iranian *\*das* '10' < PIr *\*daca-* (Holopainen 2019: 379). Late Proto-Uralic *\*a* > *\*u* in Permic, so this borrowing is clearly later than the Permic vowel changes. Also, Permic *\*s* cannot reflect Proto-Iranian *\*c*.

All the evidence presented in the preceding subsections will be considered when I set forth a new model for the disintegration of Late Proto-Uralic in Subsection 5.1. Consequences for the location of Late Proto-Uralic will be considered in Subsection 5.3.

## 3. Paleolinguistic arguments

Due to the tightened criteria, this section contains only few items. For example, the words for 'bee', 'honey', and 'oak' do not have cognates in the easternmost branches, while the words for 'fir' and 'larch' cannot be reconstructed for Late Proto-Uralic due to sound correspondences that are too irregular.

### 3.1. Uralic *\*sęksa* 'Siberian pine'

For a long time, tree names have been used to locate the Proto-Uralic homeland. Many trees are too widespread to be diagnostic (birch, pine, spruce, willow, alder, rowan, bird cherry), but there are two groups of trees spreading in opposite directions and meeting in the Volga-Ural Region: the western deciduous trees (oak, elm, maple, linden, hazel, ash) and the eastern coniferous trees (Siberian pine, fir, larch). Consequently, neither of these tree groups could disqualify the Volga-Ural homeland, but the western trees could disqualify the more distant Siberian homeland, and the eastern trees could disqualify the Upper Volga and the more western homelands, if their names could be reconstructed for Late Proto-Uralic. At the present, there is only one name for an eastern tree that fulfills the required criteria:

U *seksa 'Siberian pine (*Pinus sibirica*, earlier *Pinus cembra sibirica*)' >
    Permic *sus 'Siberian pine; juniper (in Udmurt)' ~ Mansi *tīt 'Siberian
    pine' ~ Khanty *LīɣəL 'Siberian pine' ~ Samoyedic *tititåjŋ 'Siberian pine'
    (UEW: 445; Aikio 2015: 60).

Even though there are no cognates in the westernmost branches, the regu-
lar sound correspondences point to a very old word. The distribution of the
Siberian pine (AgroAtlas: *Pinus sibirica*) excludes Europe except for the
north-easternmost part north from the Upper Kama Region, but it also
excludes the southern half of West Siberia. While the pollen samples were
earlier often dated by stratigraphy, now there are an increasing number of
radiocarbon datings available. According to the new results, the Siberian
pine appeared in the Upper Kama Region only ca. 1300 BCE (Lapteva et al.
2017: 330), so this piece of evidence narrows the Late Proto-Uralic home-
land down to the Central Ural Region or the northern half of West Siberia.

### 3.2. Uralic *ćelə 'elm'

The following is the name for a western tree with the widest Uralic
distribution:

U *ćelə 'elm (*Ulmus*)' >
    Mordvin *śäľəń 'elm' ~ Mari *šolə 'elm' ~ Hungarian *szil* 'elm' (UEW:
    458–459; Aikio 2014: 67).
    New possible cognate: Mansi *sīlt 'linden bast' > TY TCh *sāľť*, KL *sāľt*,
    KM KU Pe VN VS VNZ LM *sēľt*, LL *seľt* 'linden bast' (Kannisto 2013:
    748a), LU So *sālt* 'willow bast?' (in compound words only; Kannisto
    2013: 741b).

Aikio omits the otherwise regular Finnic cognate *salaga 'crack willow
(*Salix fragilis*) etc.' as a Germanic loanword. The Finnic *jalaga 'elm' would
be an otherwise suitable cognate, but here the problem is the lack of fur-
ther examples of the required irregular change *ć > **j. On the other hand,
I have recently proposed a possible Para-Slavic etymology for the Mordvin
word and the Finnic word *halaga 'bay willow (*Salix pentandra*)', speculat-
ing on the possibility of including also the Finnic *jalaga and *salaga there
(J. Häkkinen 2022: 132–133).

The Mordvin *\*ä* is unexpected but explainable, as Aikio (2020: 48–49) presents a parallel case for the change U *\*ẹ* > Mordvin *\*a* > *\*ä* next to *\*ć*: U *\*ẹćə-* 'set' > Mordvin *\*aśəm* > *\*äśəm*. The Hungarian and Mansi words can be regularly derived from U *\*ćẹlə*, but the final *\*-t* in Mansi should be explained as a secondary element. However, such unexplained extra elements are occasionally seen at the end of Uralic nouns and verbs in Mansi (cf. Aikio 2020: 6–8: *\*āľāt*, *\*-ūjt-*; 11: *\*ūləć*). In the Mansi varieties the palatalized *\*ľ* is more common than the expected *\*l*, but this does not appear to be a big problem.

In Mari, *\*ẹ* > *\*ü* is the most frequent outcome, although *\*ẹ* > *\*o* is common in front of velar consonants (Metsäranta 2020: 80–81), and it occurs occasionally also in other environments: U *\*δ̕ẹmə* 'bird cherry' > Mari *\*ľom-bə* | U *\*ẹppə* 'father-in-law' > Mari *\*owə* | U *\*lẹmpə* 'pond, bog' > Mari *\*lo/åp* 'hollow, lowland' (Metsäranta 2020: 314–315). Therefore, the cognates of Mordvin, Mari, Hungarian, and Mansi can all be derived from the common proto-form.

The different meaning of the proposed Mansi cognate requires some attention. First, names for trees do not necessarily follow our modern logic of biological taxonomy, instead they can be motivated by the function of a tree. Bast was taken from both linden and elm, and the Finnish words *pärnä/pernä* and *kynneppää* can refer to both trees; the former word can also mean 'bast' (Vilppula 1984: 196–198; SSA 2: *pärnä*). Furthermore, *lehmus* 'linden' can mean 'soft wood' and *niini* 'bast' can mean 'linden' in some Finnic varieties (SSA 2: *lehmus*, *niini*). These words seem adequate parallel cases to justify the semantic shift 'elm' > 'bast' for the Mansi word.

Second, a semantic shift is to be expected when a language has spread outside the natural habitat of the tree. Mansi is presently spoken in Northwest Siberia, an area to which the elm subspecies have never spread; the easternmost extension ends in the Central Ural Region (AgroAtlas: *Ulmus glabra*, *Ulmus laevis*), and even to the south from the current Mansi region around Chelyabinsk only sporadic pollen finds appear (Lapteva & Korona 2012: 329).

Linden, on the other hand, has spread to Siberia beyond the Central Ural Region, advancing by the present time well past the Irtysh–Tobol confluence (AgroAtlas: *Tilia cordata*), although its presence has remained marginal there (less than 1% of the pollen sum; Volkova et al. 2016: 309). Linden is the primary bast tree in the Southern Mansi region. In the region of the Northern Mansi varieties along the rivers Sosva and Lozva even

linden is not present, so the meaning there has again shifted to 'willow bast'. On the European side of the Urals, the elm arrived in the Middle Kama Region from the southwest already ca. 5000 BCE (Shumilovskikh et al. 2020: 533).

Concerning the names for 'Siberian pine' and 'elm', they both cover only part of the Uralic branches. Naturally, languages tend to lose words when they are no longer needed (semantic shift is, of course, another option). Therefore, we would not even expect to find names for eastern trees in the westernmost languages or names for western trees in the easternmost languages. Consequently, the loss of the word for (or at least the meaning) 'elm' in Mansi, Khanty, and Samoyedic is quite as expected as the loss of the word for 'Siberian pine' in Finnic, Saami, and Mordvin. Such tree names have the best possible excuse for their disappearance from languages, compared to words from any other semantic category.

Against this background, a regular cognate for U *ćęlə 'elm' in Hungarian and Mansi – the latter showing a credible semantic shift – justifies the Finno-Ugric status for the word. Admittedly, Samoyedic seems to be the first Uralic branch to form a regionally separate center, but the associative arguments (see Subsection 5.3) still require the presence of Samoyedic in the immediate vicinity of Finno-Ugric for a long time. Consequently, the name for 'elm' would exclude Siberia as a whole from the possible regions for the Late Proto-Uralic homeland. On the other hand, if we omitted both words due to a lack of cognates in the other end of the language family, we would no longer have any diagnostic tree names left in Late Proto-Uralic.

Interestingly, the eastern trees have different evidentiary value in the cases of the European vs. the Siberian homeland, because these trees originate in Siberia. Considering the Siberian homeland, names for the eastern trees could have appeared in the language already much earlier than at the actual Late Proto-Uralic stage. This in turn would lead to a paradoxical conclusion that it would no longer be necessary to locate Late Proto-Uralic in Siberia – it would be enough to locate the very distant Pre-Proto-Uralic there. Considering the European homeland in the Volga-Ural Region, the names for both the western and the eastern trees could not have been acquired before the trees themselves spread to easternmost Europe from the opposite directions. Therefore, in the Volga-Ural homeland the appearance of the tree names would probably be closer to the actual Late Proto-Uralic stage.

### 3.3. Uralic *wVć(k)V 'metal'

This word is highly diverse in Uralic, and no single reconstruction has convincing cognates with wide enough distribution. Therefore, the dispersal of the word has recently been considered later than Late Proto-Uralic (Aikio 2015: 42–43). Grünthal et al. (2022: supplementary file: 12) have tried to defend a more suitable reconstruction for the Samoyedic cognate, *wäsä, by assuming that the Nganasan cognate is irregular, but even if we accepted LPSy *wäsä instead of *wasa, there would be no certain cognates with *ä elsewhere: Saami *veaškē would regularly come from Uralic *wećka, and even though it might also come from *wäćka, also *waćka could be possible due to sporadic palatalization *a > *e caused by the adjacent *ć. The Finnic *vaski could come from both Uralic *waćkə or *wäćka.

Mordvin and Hungarian rather point to *a in the first syllable, but Mari and Khanty even point to a secondary *a not corresponding to the Uralic *a. In Permic and Mansi the word only appears as the latter part of a compound word and has therefore been badly eroded, although *e seems a possible original vowel there (J. Häkkinen 2023). In conclusion, the word cannot be reliably reconstructed for Late Proto-Uralic. Even if it were a Late Proto-Uralic word, its locational value would be weak, because the reconstructed meaning appears to be 'metal' in general, and copper was used widely in Northern Eurasia already several millennia before Late Proto-Uralic.

### 4. Other arguments

### 4.1. Lack of loanwords from certain languages

Nichols (2021: 355) and Grünthal et al. (2022: 8) write that the lack of Para-Baltic loanwords from the Fatyanovo Culture would testify against a homeland in the Volga-Ural Region. However, there are several reasons why such an argument is not valid.

First, we do not know for certain which language lineages were spoken within the Fatyanovo Culture. Even the Indo-Iranian lineage might have come from the Fatyanovo sphere: at least the Sintashta population was very similar to the Corded Ware populations both at the autosomal level (Saag et al. 2021: 5) and at the Y-chromosomal level (Underhill et al. 2014: 3, 5; Saag et al. 2021: 3).

Second, even though the Northwest Indo-European lineage was probably spoken (also) within the Fatyanovo Culture, it remained phonologically archaic up to the 2nd millennium BCE (see Subsection 2.3), so it is anachronistic to require Para-Baltic loanwords at the 3rd millennium BCE. Instead, there are plenty of proposed Archaic and Northwest Indo-European loanwords which can be connected to that cultural context, although at the moment we cannot stratify them reliably.

Third, the lack of loanwords is not a valid argument, because there is no universal law which would compel adjacent languages to borrow certain words (or any words at all) from each other. Borrowing a word is a complex sociolinguistic situation – it is not an automatic consequence of two languages spoken close to each other. Sometimes words get borrowed very easily, sometimes not.

Fourth, even if there originally were loanwords, there is a possibility that the speakers in the contact zone shifted their language to another, thus losing such loanwords along with their whole language. The Middle Volga was an especially complex area, where at many times several overlapping cultures coexisted (Carpelan & Parpola 2001: 86–89). The situation was not necessarily as simple as the Uralic speakers borrowing loanwords from all of their neighbors and then that language expanding westwards, but there could have occurred language shifts back and forth, following several different cultural influences in different directions. To complicate the situation even more, there are also traces of Paleo-North European languages in the region (J. Häkkinen 2009: 47–49; Aikio 2015: 43–47; Zhivlov 2015), so the closest neighbor of the Uralic speech community in the west was not necessarily an Indo-European language.

Consequently, even though the Fatyanovo Culture belonged to the Corded Ware Cultures, we should not assume that there was only one language present within its whole wide region. Balto-Slavic was only one of possibly several languages spoken in the region, and it becomes recognizable only during the 2nd millennium BCE (Häkkinen 2022: 138–141; Kallio, forthcoming).

Another related argument can be seen in Grünthal et al. (2022: supplementary file, page 13), where the authors write that because the only metal name in Proto-Uralic meant only 'metal' in general, this could be seen as an argument against the homeland close to the rich metallurgical center in the Southern Urals. However, the authors ignore the fact that there were metallurgical centers also near the assumed Siberian homeland candidates

and that there were Indo-Europeans also in South Siberia already since 3300 BCE within the Afanasyevo Culture (which knew bronze metallurgy and even meteoritic iron; see Koryakova & Epimakhov 2006: 188–189). If the Late Proto-Uralic homeland was anywhere near there, should we not expect (Indo-European) metal names borrowed into Uralic also there? Of course, such expectations are in any case unfounded, as argued above.

Moreover, metal names are mainly temporal arguments, and they would have value for locating the Proto-Uralic homeland only if a word could be shown to have been borrowed from a certain locatable language already into Late Proto-Uralic. However, there are no metal names fulfilling these criteria (see Subsection 3.3). I have recently proposed an Indo-Iranian origin for the first part of a Mansi–Permic compound metal name, but it is in any case post-Proto-Uralic (J. Häkkinen 2023).

## 4.2. The taxonomic argument

The taxonomic argument is based on the deepest division within a language family. However, it is easy to find counterexamples like the Turkic, Indo-Iranian, or Celtic homelands, which are not located in the region where the deepest division is nowadays observed – not even within the present distribution of these branches (Kallio 2015b: 84). This argument could only work in cases in which the homeland falls within the present region of the language family – and not necessarily even always when that condition is fulfilled. Consequently, as this argument requires that we already know where the homeland is before we can apply it reliably, it is practically redundant.

Furthermore, the views on the taxonomic structure of the Uralic language family have been notoriously diverse: there are many different results based on many different pieces of evidence. However, sharing of the inherited lexicon, not to speak of only a short list of selected items (like numerals) thereof, cannot be considered a reliable datum, because there are possible distorting processes leading to either increased or decreased lexical sharing between branches (J. Häkkinen 2012). Therefore, the phonological level should always be taken as the starting point, although the lexical level cannot be totally dismissed, as will be shown with the new model for the disintegration of Proto-Uralic (Subsection 5.1).

Related to the taxonomic argument, Saarikivi (2022: 57) writes that the Ugric group is more diverse than other Uralic groups, and therefore the

Uralic homeland could be located close to the homelands of Hungarian, Mansi, and Khanty. However, Ugric is not necessarily a taxonomic branch but instead a unit of areal convergence between the three branches. Even if it were a true branch, its greater diversity could not be interpreted as evidence about its greater temporal depth, because the structure and depth of a branch are purely stochastic. The variables are: (1) How long after the proto-language did the branch-specific sound changes occur? (2) How many consecutive divisions occurred within the branch? (3) How many of these sub-branches survived, without being leveled by closely related dialects or replaced by other languages?

There are several possible results from these variables. It is very well possible that most of the branches within a language family descended from a single recent macro-branch; and it is also possible that a single branch would have greater temporal depth within a language family than a macro-branch with several sub-branches. How it really was, cannot be deduced straightforwardly from the taxonomic structure of the language family – width cannot testify for depth.

Saarikivi (2022: 57) is probably correct when he writes, "If there is such a thing as Proto-Ugric, it is, without doubt, even older than Proto-Samoyedic." This means that the disintegration of Proto-Ugric (if it was a branch) would be older than the disintegration of Late Proto-Samoyedic. Yet this branch-internal disintegration does not necessarily correlate with the external (family-wide) disintegration, for the reasons stated above.

## 4.3. Distances and tendencies

It has been argued that the Volga homeland is improbable, because it would require a movement over a huge distance to the region of Late Proto-Samoyedic (Grünthal et al. 2022: 8). Nevertheless, the known regional distribution of the Uralic branches is what it is, and no matter where the homeland was, some branch has had a greater distance to traverse than some other branch. We could equally well use this argument against the homeland in South Siberia, claiming it implausible that Samoyedic has not moved at all, while the Saami branch has moved over a huge distance. If anything, the distance argument could only be seen to support a somewhat central homeland, where the total distance for any single branch is not extremely long – that would be somewhere around the Volga-Ural Region.

It has also been argued that the general direction of movements in Eurasia has been from the east to the west (Grünthal et al. 2022: 8). However, it is easy to find counterexamples, like the spread of Pre-Proto-Tocharian and Indo-Iranian. A tendency is not a law, and therefore its evidentiary value is weak.

Quite similar is an argument based on language sinks: such regions would pull languages rather than push them, and as a language sink, the Middle Volga Region would be an improbable homeland (Grünthal et al. 2022: 8). Still, a proposed Uralic homeland in the Minusinsk Basin in South Siberia is also a language sink (Nichols & Rhodes 2018: 8). Again, this is merely a tendency, not a law.

Moreover, at least in the archaeological data it is well known that the region of the Upper–Middle Volga has for a long time been an expansion center (Carpelan & Parpola 2001: 79–83). Certainly there were also languages connected to these consecutive cultural expansions, but all those earlier languages later disappeared under the Uralic expansion. Even though the earlier languages have not survived, we should not ignore their earlier existence: by constructing tendencies based only on the very few surviving language families, we cannot reach the complex reality of the past.

## 4.4. Lack of a non-Uralic substrate

This could be a potentially illuminating argument, but at the present state of the art, we know too little about the lost languages, their distribution, and how to even trace them properly. This criterion also works in one direction only: the presence of a non-Uralic substrate in a language can testify that there is no continuity from Late Proto-Uralic in that particular region, but the absence of a visible non-Uralic substrate cannot testify reliably that Late Proto-Uralic was spoken in that region.

Moreover, this criterion could only work for language families within which some extant branch has remained in the original location of the homeland, but in order to be able to fulfill that demand, we should already know where the original homeland was. Consequently, this argument is redundant.

## 5. Late Proto-Uralic – when and where?

After all the relevant arguments and pieces of evidence have been considered, it is a time for a conclusion. As byproducts, a new model of disintegration and taxonomy will be proposed for the Uralic language family (Subsection 5.1), followed by the most resolute possible dating for the Uralic reconstruction stages (Subsection 5.2). Next, the Late Proto-Uralic homeland will be located (Subsection 5.3). Finally, the rules for connecting the linguistic results to the archaeological results will be briefly discussed (Subsection 5.4) and possible counterparts for the Uralic reconstruction stages will be proposed (Subsection 5.5).

### 5.1. The disintegration of Proto-Uralic

The disintegration of Proto-Uralic is connected to both the dating and the locating of the proto-language. Linguists who have touched on the Uralic homeland problem (myself included), have not always been able to properly distinguish between different levels of testimony: evidence from the lexical level, from the level of sound substitution of loanwords, and from the level of sound changes. This has sometimes led to favoring one level of evidence over another in order to solve an apparently contradictory picture. In the present critical examination I aim to remedy this problem, and to develop a model which not only allows us to be aware of and distinguish between all the levels of linguistic evidence, but also to use them together to "triangulate" for the most accurate chronological reconstruction possible.

In Section 2 some seemingly contradicting results from the Indo-Iranian loanword layers were presented:

1. There are no certain Indo-Iranian loanwords in Late Proto-Uralic.
2. Based on the lexical level, Samoyedic was already outside the Indo-Iranian contact zone during the Early Proto-Indo-Iranian loanword layer.
3. Based on the sound substitutions, Samoyedic borrowed loanwords together with the other Uralic branches and even agreed with their sound substitutions during the Late Proto-Indo-Iranian loanword layer.
4. Based on the sound substitutions, different Uralic branches show parallel borrowings from the Proto-Iranian loanword layer onward.
5. Based on the (macro-)branch-specific sound changes, the phonological developments began only after all the previous stages.

A new model of disintegration is needed to take all this evidence into consideration. I propose the following alternative labels for these five reconstruction stages:

U1 = LPU  (Late Proto-Uralic)  = uniform proto-language
U2 = AU   (Ancient Uralic)     = reorientating proto-language
U3 = BU   (Bicentric Uralic)   = bicentric proto-language
U4 = CU   (Common Uralic)      = substitutionally separated pre-dialects
U5 = DU   (Diverging Uralic)   = phonologically separated proto-dialects

It is a matter of personal taste whether one favors numerical or alphabetical labels. It is also debatable whether the first three stages could perhaps be included within the concept of Proto-Uralic, but I find it most clarifying to give every stage its own label. The label Common Uralic was used by Grünthal et al. (2022), but the other labels after the stage U1 are new. Figure 2 illustrates the disintegration of Proto-Uralic.

Centers are part of the same speech community, although regionally separated. Pre-dialects are regionally separated and show independent sound substitutions. Proto-dialects are regionally separated, show independent sound substitutions, and show (macro-)branch-specific sound changes. The next stage would be the branch-specific protolanguages, divided into early, middle, and late proto-stages, if necessary.

Familiarity with family-tree modeling might prevent us from pursuing or comprehending such a model, because in a line-drawn family tree, one trunk abruptly divides into two branches. However, linguistic reality is rarely so simple. A more adaptable illustration for the more complex process of disintegration is a "family funnel", which allows us to stratify features alternating between wide and narrow distribution (Figure 3).

Naturally, later contact phenomena and convergence by chance (like erosion or loss of vowels in unstressed syllables) can occur between branches. Here the focus is only on features so ancient that they have had an impact on views about the taxonomic model of the Uralic language family. An abrupt disintegration means that one branch has immediately moved further from others, and a rigid disintegration means that no shared isoglosses appear between branches after the initial division. Based on the evidence from the Indo-Iranian loanword layers, the disintegration of Late Proto-Uralic was neither abrupt nor rigid.

**U1 = LPU**
**Late Proto-Uralic**

**U2 = AU**
**Ancient Uralic**

**U3 = BU**
**Bicentric Uralic**

Different loanwords in different centers

**U4 = CU**
**Common Uralic**

**U5 = DU**
**Diverging Uralic**

Different sound substitutions
in different pre-dialects

Different sound changes
in different proto-dialects
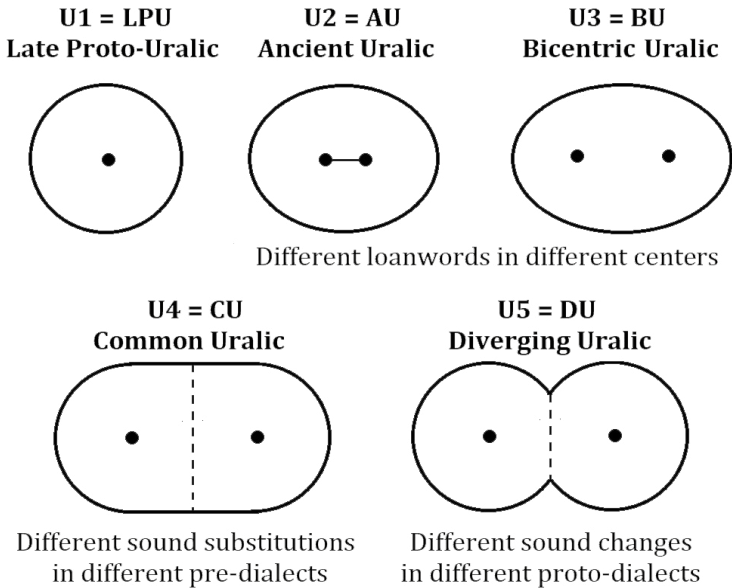
Figure 2: Disintegration of Proto-Uralic: five successive reconstruction stages based on the evidence from the Indo-Iranian loanword layers

**Uralic branches**                  No longer shared changes

U5                                   Last shared changes

U4                                   Further division into several units

U3                                   More partial than total changes

U2                                   First changes which do not spread
                                     to the whole speech community

U1                                   Uniform proto-language

**Late Proto-Uralic**

Figure 3: The family funnel illustrates non-abrupt, non-rigid disintegration of a proto-language. Isoglosses (oval discs) may contain information from different levels of language: lexicon, sound substitutions, and sound changes.

## 5.2. Indo-Iranian evidence for dating Late Proto-Uralic

The absolute chronology of Indo-Iranian is quite firm in the later end, anchored by the following pieces of evidence, from the latest to the earliest:

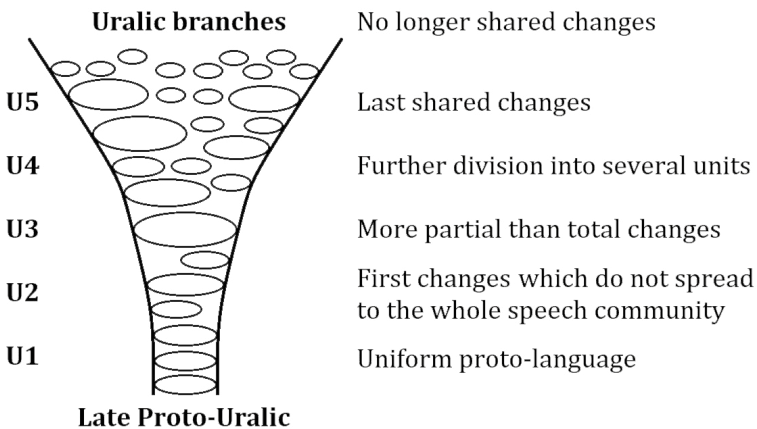1. The first Indic and Iranian writings appear at the mid-first millennium BCE, and they already show many branch-specific sound changes.
2. The Indic language of the *Ṛgveda* is considerably more archaic than Classical Sanskrit, even though the written attestations are not earlier, and thus the oral formation of the *Ṛgveda* has been dated already to the end of the 2nd millennium BCE (Cardona 2017). The same goes with the Iranian Old Avestan language as compared to the Younger Avestan language (Skjærvø 2017). Both of these more archaic varieties were transmitted in liturgical contexts for a long time before they were written down, while at the same time colloquial Indic and Iranian varieties went through more phonological changes.
3. Indic words in the Mitanni and Hittite writings from ca. 1400 BCE are even more archaic than Vedic Sanskrit, close to Proto-Indic (Witzel 2001: 49).
4. The chariot vocabulary shared by Indic and Iranian and certain ritualistic features described in the *Ṛgveda* and *Avesta* are best matched by the archaeological remains of the Sintashta Culture in the Southern Urals ca. 2100–1800 BCE (Anthony 2007: 408–411). Late Proto-Indo-Iranian can therefore be dated and located there.
5. Beyond that, the dating becomes more imprecise. Disintegration of Late Proto-Indo-European is dated to the late 4th millennium BCE (Anthony & Ringe 2015), but the Indo-Iranian sound changes probably began to occur only during the latter half of the 3rd millennium BCE.

The datings of these Indo-Iranian stages can be transferred to the Uralic side through the disintegration model based on Indo-Iranian loanword layers (Table 1).

Table 1: Uralic reconstruction stages and their approximate dating transferred from the Indo-Iranian chronology through the loanword layers

| Indo-Iranian stage | Dating and Uralic stage | Disintegration of Proto-Uralic | | | |
|---|---|---|---|---|---|
| Archaic Indo-European | ~2800 BCE U0 = EPU | Early Proto-Uralic | | | |
| Archaic Indo-European | ~2500 BCE U1 = LPU | Late Proto-Uralic | | | |
| Early Proto-Indo-Iranian | ~2300 BCE U2 = AU | Finno-Ugric | | | Samoyedic |
| Late Proto-Indo-Iranian | ~2000 BCE U3 = BU | Finno-Ugric | | | Samoyedic |
| Proto-Iranian/Indic | ~1800 BCE U4 = CU | P r e - d i a l e c t s | | | |
| Archaic Iranian/Indic | ~1500 BCE U5 = DU | West Uralic | | East Uralic | |
| Archaic Iranian/Indic | ~1200 BCE | Saami, Finnic, Mordvin | Mari, Permic | Hungarian, Mansi, Khanty | Samoyedic |

## 5.3. Evidence for locating Late Proto-Uralic

For locating purposes, the earlier stages of the Indo-Iranian lineage carry the most weight, being closer in time to Late Proto-Uralic. Based on the connection between the chariot-related vocabulary and the ceremonial practices described in the *Ṛgveda* and *Avesta* on the one hand, and the archaeological remains of chariots and graves on the other hand, Late Proto-Indo-Iranian is connected to the Sintashta Culture (ca. 2100–1800 BCE), and the language only spread beyond the river Tobol around 2000 BCE, when the Sintashta-rooted Andronovo Complex spread to Southwest Siberia and Northern Central Asia (Anthony 2007: 389–390, 397; E. Kuz'mina 2007: 451).

All the stages leading to Late Proto-Indo-Iranian developed on the European side of the Urals (E. Kuz'mina 2007: 305). There are no serious challenging views for the original European homeland of the Indo-Iranians, which is significant considering the location of the Uralic stages. Traditionally the Indo-Iranian lineage has been connected to archaeological

cultures of the European steppe, like the Poltavka Culture (Mallory & Adams 1997: 439–440), but also influence from the Abashevo Culture into the Sintashta Culture has been acknowledged (Anthony 2007: 382–387).

Even though the Poltavka Culture began already ca. 2600 BCE (Morgunova & Khokhlova 2013), there is no reason to believe that the Indo-Iranian phonological developments began so early. Most of the Indo-Iranian loanwords were borrowed into Uralic only after the vowel merger in Late Proto-Indo-Iranian, and the rest of them need not be much earlier, reflecting already most of the Middle Proto-Indo-Iranian sound changes (see Subsection 2.5). Therefore, even though already the Poltavka Culture spread beyond the Southern Urals, it is highly uncertain to try to explain the Early Proto-Indo-Iranian loanwords in Uralic by its extension beyond the Urals, as implied by Grünthal et al. (2022: 10). More likely Early Proto-Indo-Iranian developed only slightly before the appearance of the Sintashta Culture in the Southern Ural Region ca. 2100 BCE.

Moreover, it is not certain that the Indo-Iranian lineage can be connected to the Poltavka Culture at all. The Sintashta Culture has roots both in the steppe cultures and in the Abashevo Culture, and the latter has roots also in the Fatyanovo Culture (belonging to the Corded Ware Cultures; Anthony 2007: 383). Based on the recent genetic results, the Sintashta population was most similar to the populations of the Fatyanovo Culture and the other Corded Ware Cultures, both at the autosomal level (Saag et al. 2021: 5) and at the Y-chromosomal level (Underhill et al. 2014: 3, 5; Saag et al. 2021: 3). Consequently, we cannot exclude the Fatyanovo Culture as the possible origin of the Indo-Iranian lineage.

For Uralic, the exact cultural counterpart for Early Proto-Indo-Iranian is irrelevant, because all the candidates (the Poltavka, the Abashevo, and the Fatyanovo-Balanovo Culture) coexisted in just about the same Volga-Ural Region at the late 3rd millennium BCE. The partial overlapping of the Fatyanovo-Balanovo and the Abashevo Cultures could explain the regionally spread features shared by Balto-Slavic and Indo-Iranian, like satemization and the *ruki*-rule (see Section 2.3), as suggested by Parpola (2022: 264).

Although the Early Proto-Indo-Iranian loanword layer directly requires locating only the Finno-Ugric center in the vicinity of the Southern Urals (likely to the north from the Indo-Iranians, in the Central Ural Region), there are associated arguments requiring also the presence of the Samoyedic center right next to it for a long time:

1) Samoyedic still shares some Late Proto-Indo-Iranian loanwords with other Uralic branches during the stage U3, showing in different words the same arbitrary substitute for the LPIIr *a* as the Finno-Ugric branches. These shared loanwords would be impossible to explain if Samoyedic were already located in South Siberia (see also Kallio 2015b: 82, footnote 5).

2) Samoyedic appears to have participated in the chain of three subsequent sibilant changes shared with Hungarian, Mansi, and Khanty, which requires the presence of Samoyedic next to the three other branches (which descend from the Finno-Ugric center) still during the stage U5.

3) Hungarian, Mansi, and Khanty share Early and Late Proto-Indo-Iranian loanwords (with the same sound substitutions) with the western branches, so they must still be located around the Central Ural Region through the stages U2 and U3. After that, they could have moved to Siberia together with Samoyedic, in which case the East Uralic sibilant changes at the stage U5 could have occurred in Siberia. On the other hand, it is also possible that all these branches were still in the Central Ural Region during that stage.

We may note a strikingly compatible pattern between the distribution of the Indo-Iranian loanwords in Uralic and the reconstructed dispersal of Indo-Iranian: first, during the stage U2, when Early Proto-Indo-Iranian was spoken in the Volga-Ural Region, the Samoyedic center was outside the contact zone. Later, during the stage U3, when Late Proto-Indo-Iranian spread to the east and was spoken in the Southern Trans-Urals between the headwaters of the rivers Ural and Tobol within the Sintashta Culture, the Samoyedic center got involved in the Indo-Iranian contacts. Therefore, we can locate the Finno-Ugric center in the western part of the Central Ural Region and the Samoyedic center in the eastern part of the Central Ural Region (see Figure 4 below).

The crucial question is whether the uniform Late Proto-Uralic (the stage U1) was spoken in that very same area or somewhere else. It seems impossible that the disintegration of Late Proto-Uralic could have occurred in some distant location, like around the Middle Irtysh Region or even beyond. If that was the case, Samoyedic would have remained there, while only Finno-Ugric would have moved to the Central Ural Region, and we could not explain how Samoyedic could share with Finno-Ugric some Late Proto-Indo-Iranian loanwords with the very same arbitrary sound substitutions at the stage U3. Neither could we explain how Samoyedic could have participated in the chain of three successive sibilant changes

together with the ancestral stages of Hungarian, Mansi, and Khanty at the stage U5.

Consequently, there is no avoiding the inevitable conclusion: the disintegration of Late Proto-Uralic must have begun in the Central Ural Region. Of course, it is possible that its immediate ancestor arrived from South Siberia only slightly before the beginning of the disintegration, but that possible stage in South Siberia could not be labeled Late Proto-Uralic. Earlier stages of the Uralic lineage fall beyond the scope of this scrutiny, but hopefully future research produces more results about that topic. At the moment scholars might locate the distant Pre-Proto-Uralic both in Siberia (Aikio 2022: 26–27) and in the Volga-Ural Region (Parpola 2022).

This conclusion has an important consequence: through the above-presented associative arguments, every piece of evidence which is absent in Samoyedic yet shows a Finno-Ugric distribution, phonologically regular enough, and semantically credible cognates, now has an impact on the location of Late Proto-Uralic itself. By anchoring the Finno-Ugric center, any such piece of evidence anchors also the Samoyedic center in the immediate vicinity of the Finno-Ugric center until after the stage U3, and in the vicinity of the other East Uralic branches until after the stage U5. I shall label this factor the "Uralic bundle effect".

While the Indo-Iranian loanword layers pull Late Proto-Uralic to the west, the area of the Siberian pine holds the reins for that pull. Even today, the natural habitat of the Siberian pine on the European side of the Urals does not reach south from the Upper Kama Region, although random occurrences might appear in a wider area in the north-eastern part of European Russia (AgroAtlas: *Pinus sibirica*). However, random occurrences can hardly explain the preservation of the tree name in the daughter languages for over four millennia – clearly the languages must have been spoken very close to the natural habitat of the Siberian pine.

Pollen of this tree appears in the Upper Kama Region only ca. 1300 BCE (Lapteva et al. 2017), which requires the presence of the speakers of Late Proto-Uralic firmly in the Central Urals, excluding the Middle Volga homeland and every homeland candidate further to the west. However, the South Siberian homeland encounters problems, too. Even though the Siberian pine has for a long time been present in the Sayan Region (Blyakharchuk & Chernova 2013), it is not present in the southern part of West Siberia (AgroAtlas: *Pinus sibirica*), through which the Uralic language could probably be supposed to have extended were the homeland

in South Siberia. To secure the presence of the Siberian pine along the expansion route, the language should have moved first to the north along the Yenisei, and only then to the west.

The Finno-Ugric name for 'elm' excludes the West Siberian homeland, but the Central Ural Region and the European homelands are acceptable. Perhaps the word was never even adopted into Samoyedic, but in any case, the "Uralic bundle effect" requires Samoyedic in the immediate vicinity, i.e. in the eastern slopes of the Central Urals. During the 3rd millennium BCE, the Central Ural Region was the only region where the Siberian pine and elm met each other (Figure 4).

However, names for these trees did not necessarily appear in the language at the same moment. It is possible that one of them appeared earlier in Pre-Proto-Uralic either in the more western or in the more eastern region, and the other one was adopted later. Nevertheless, no matter which scenario we favor, the evidence always pulls Late Proto-Uralic back to the Central Ural Region.

Even if one rejects the name for the elm as uncertain due to the semantic shift in Mansi, the final result would not change. In that case, the name for the Siberian pine could have been borrowed already earlier in Siberia, but the Indo-Iranian loanword layers still require Late Proto-Uralic and the subsequent stages in the Central Ural Region (Figure 4). The Koptyaki Culture appears to have been in the right place at the right time (see Subsection 5.5).

The boundaries of the Siberian pine, elm, and forest-steppe on the map in Figure 4 are based on the present distribution, but the latter two have not changed for many millennia, although the boundary of Siberian pine was somewhat further to the east during the Late Proto-Uralic stage, as described above.

In the Central Trans-Urals, the forest-steppe reaches up to the river Iset (the boundary as drawn in O. Kuz'mina 2021: 1209) and has remained rather stable for several millennia, since long before the Late Proto-Uralic stage (Lapteva & Korona 2012: 329–330). Distinguishably Mansi place-names reach from the present Mansi region to the south, between the headwaters of the Neiva and Iset (Matveev 2011: 445) – this region is exactly on the Central Ural Passage. Consequently, there is no need to locate the ancestral stages of Hungarian, Mansi, and Khanty in any more southern environment in order to explain the horse-related vocabulary.
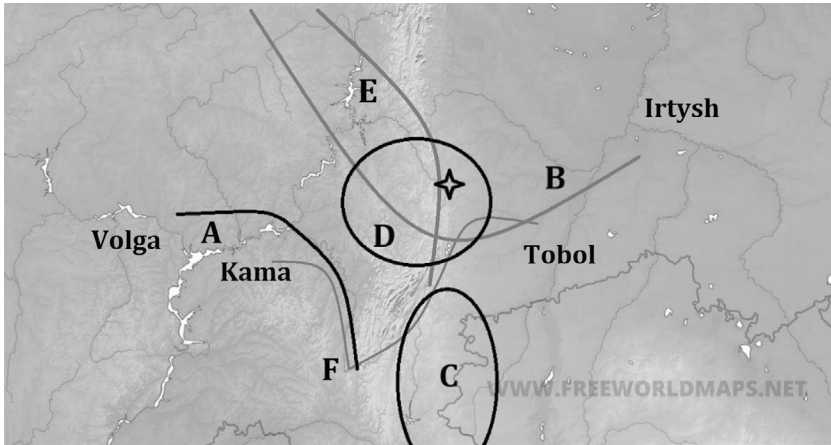
A = The north-eastern boundary of the Abashevo Culture (early stage)
B = The southern boundary of natural habitat of Siberian pine
C = The approximate area of the Sintashta Culture
D = The approximate area of the Koptyaki Culture
E = The north-eastern boundary of natural habitat of elm
F = The northern boundary of forest-steppe
✥ = The Central Ural Passage

Figure 4: The most probable homeland for Late Proto-Uralic based on evidence from Indo-Iranian loanwords (A, C) and the names 'Siberian pine' (B) and 'elm' (E). The Koptyaki Culture (D) matches the required spatial and temporal coordinates. Forest-steppe (F) also reached to the area, explaining the Ugric horse-related vocabulary.

## 5.4. Connecting linguistic and archaeological results

In the following subsection I consider possible extra-linguistic counterparts for the Uralic dispersal. It must be emphasized that what is suggested here, is not a result as much as a starting point for further research. Nevertheless, even the starting point requires certain conditions: (1) that we accept the linguistic results and (2) that we can find an extralinguistic counterpart in the right place at the right time, its later stages spreading in the right direction(s).

Methodological pitfalls of multidisciplinary surveys have been recognized for a long time in Indo-European and Uralic studies (e.g., Mallory 1989; 2001; Aikio & Aikio 2001; Heggarty 2007; J. Häkkinen 2010; Saarikivi & Lavento 2012), and I shall not delve deeper into the methodology here. I only mention a few important points.

First, an archaeological culture could contain several language communities, and a language community could correspond to several cultures. Therefore, if we consider an ancient language as a phenomenon tightly following cultural boundaries or distribution of cultural features, we are not in reality even dealing with language but with some imaginary pseudo-linguistic level: an abstract projection of cultural boundaries, erroneously labeled as "language".

Second, cultures are usually polythetic, which means that a distribution may vary from item to item. How then could we ever guess to which an item a language would best correspond? How could we know whether an ancient speech area matched better with a distribution of ceramic pots, bronze axes, or certain type of graves?

Third, a correlation between a language and an archaeological or a genetic phenomenon is always only momentary. In a different place or at a different time the same phenomenon could be connected to a different language. This is an inevitable conclusion from the fact that language is not inherited dependent on any cultural or genetic phenomena. Assuming otherwise is again dealing with some irreal pseudo-linguistic level. For every step of a language expansion, the counterpart should be looked for independently.

Fourth, there are always several possible counterparts for language. A language always has only one genealogical root (except for real mixed languages), while cultures and populations usually have several roots, and so do their ancestral cultures and populations, etc. Therefore, when trying to follow a language back in time, the probability to choose the right counterpart grows cumulatively lower by each step beyond the starting point (the initial spatial-temporal correlation).

Fifth, archaeological continuity usually corresponds to linguistic discontinuity: archaeological continuity is to some extent evident everywhere, yet the linguistic landscape is mostly a result of quite recent language expansions. The wider the area of the language family, the lower the probability that any random region was the original homeland.

One cannot discern language from culture or DNA, and archaeology or genetics do not have methods for studying language. If language is included in a multidisciplinary survey, then the most reliable linguistic results must be taken as the starting point. If there appears a discrepancy in time or space between the results of different disciplines, one cannot ignore the linguistic results and keep on claiming that a certain language must be dated earlier or located in another place in order to save that particular correlation. The only scientific way is to acknowledge the discrepancy and comprehend that clearly this extra-linguistic phenomenon is a poor match for this particular language. The only way forward is to find a better-matching candidate without discrepancies in time, space, or direction of expansion.

Consequently, when in the following subsections I will present an archaeological phenomenon as a possible counterpart for a language, it only means that there appears an apparent spatial-temporal match. I do not intend that this language is confined within the limits of such a phenomenon, nor that this language is the only possible language within those limits, nor that this language is transmitted to the following generations along with certain archaeological phenomena, nor that this language can be assumed to descend from a certain local or non-local ancestor of that appointed counterpart. Nevertheless, even a connection as thin as assumed here is still a connection to more concrete prehistorical events, which makes it easier to comprehend the context where the speakers of this language lived.

On the Indo-European side, there are some fortunate anchors between the linguistic and archaeological realities, which connect a certain language in a certain place and time: the wagon vocabulary in Late Proto-Indo-European, finding its counterpart in the remains of early wagons within the Yamnaya Culture in the late 4th millennium BCE (Anthony & Ringe 2015), and the chariot vocabulary in Indo-Iranian, finding its counterpart in the remains of the first chariots within the Sintashta Culture ca. 2000 BCE (Anthony 2007: 408–411). On the Uralic side we are not so fortunate, but we can anchor our reconstruction stages through the Indo-European loanword layers, as was demonstrated in Subsection 5.2.

## 5.5. The Koptyaki Culture and the Seima-Turbino Network

The Koptyaki Culture in the Central Ural Region occupied the natural trade route over the Central Urals: the plateau between the adjacent headwaters of the rivers Iset in Siberia and Chusovaya in Europe. Sites of the culture show bronze items of both the Samus-Kizhirovo type and the Seima-Turbino type, and its chronology and origin remain so far rather unclear. It probably derives partly from the local Ayat Culture, partly from eastern and southern influences (Korochkova et al. 2010; Korochkova 2019; Grigoriev 2021: 22). The Koptyaki Culture had contacts with the cultures to the south, unlike the contemporaneous cultures in the West Siberian southern taiga zone (Korochkova 2012: 146). This agrees with the Indo-Iranian loanword layers in Uralic.

Within the Koptyaki Culture, considerable variation is visible from the Middle Kama Region in the west to the Tobol Region in the east, but the ceramics are considered the common denominator (Korochkova 2019: 734). Chronologically probably only the later stages U3–U5 could be connected to the Koptyaki Culture itself, and Late Proto-Uralic might be connected to its poorly known local predecessor (local for the reasons explained in Subsection 5.3). The Koptyaki Culture was succeeded by the Cherkaskul and the Mezhovskaya Cultures (see below; Grigoriev 2021: 22).

Interestingly, the Central Ural Plateau was exactly the passage through which the main trade route of the Seima-Turbino Network ran. According to the distribution maps of the Seima-Turbino items, the main river route from the Sayan Region to Europe was the Irtysh–Tobol–Iset–Chusovaya–Kama–Volga (Carpelan & Parpola 2001: 99–111). Another route from Tobol was slightly more northern but ended very close in the watershed area, where the recently found site of Shaytanskoye Ozero is located: along the Tobol–Tura–Neiva–Revda–Chusovaya (Chernykh et al. 2017: 48). One might suspect that the people of the Koptyaki Culture profited greatly from the use of these routes by the traders of the Seima-Turbino Network.

It was Carpelan who first proposed that the Seima-Turbino Network was connected to the spread of Samoyedic from the Volga Region to South Siberia (Carpelan & Parpola 2001: 109). Kallio (2006) considered its connection to the spread of Proto-Uralic, followed by J. Häkkinen (2009) and Parpola (2012b; 2017). Recently Grünthal et al. (2022) connected it to the spread of Finno-Ugric westwards from West Siberia.

While the Abashevo metallurgy derived from the Circum-Pontic metallurgical tradition and was based on arsenical bronze, the Seima-Turbino metallurgy was based on tin bronze. Tin came from and was first utilized in the Altay-Sayan Region, but copper deposits were found in a wider area in the Ural Region and Kazakhstan (Koryakova & Epimakhov 2006: 28–29).

The initial stage of the network in Southwest Siberia ca. 2200–2000 BCE cannot be associated with the Uralic speakers, for the reasons presented in Subsection 5.3, but perhaps the later stages in Europe could. After securing the passage over the Central Urals (datings from the Shaytanskoye Ozero site right on the passage are ca. 2000–1650 BCE; Korochkova 2019: 733), the Seima-Turbino Network established new centers in the Kama Region and the Mid-Upper Volga Region ca. 1900–1600 BCE (Chernykh et al. 2017: 51–52; Marchenko et al. 2017).

Interestingly, the Pepkino Kurgan on the southern side of the Middle Volga in Mari El contained the remains of 28 Abashevo warriors who were killed probably around 2000 BCE (Chernykh et al. 2017: 53). As the Seima-Turbino sites appear further to the west soon after that, it seems possible that the Seima-Turbino Network managed to take control over the Volga route. However, there are traces of hostilities also between the people of the Abashevo and the Balanovo Cultures, so it is only speculation that the Abashevo mass grave and the expansion of the Seima-Turbino Network would be causally connected.

At the same time with the Seima-Turbino expansion, widespread east–west contacts can be seen also in the shared features of ceramics between the Krotovo Culture (in the Middle Irtysh Region), the Garino Culture (in the Middle Kama Region), and the Chirkovo Culture (in the Mid-Upper Volga Region; Vybornov et al. 2019: 19). These locations match with the Seima-Turbino centers in Rostovka, Turbino, and Yurino, respectively.

As ceramics in prehistoric Eurasia are widely considered to have been the realm of women, perhaps this spread of certain features in ceramics reflects the exchange of brides between the groups participating in the Seima-Turbino Network, or perhaps the bronze traders brought their families with them. Together the shared extension of both bronze items and ceramic wares seems to testify to the movement of both men and women within the wide network, which offers an adequate background also for a possibility of a language expansion.

Could the Uralic speakers, as the gatekeepers of the Central Ural Passage, have demanded their share of the network on the European side of the Urals? They already had established contacts with the neighbors to the south and southwest, as testified by the Indo-Iranian loanwords. Joining in the network and expanding it could be the ultimate reason behind the Uralic expansion to the west. The Middle–Upper Volga centers could correspond to the West Uralic branches (Saami, Finnic, Mordvin, West Chudic, and Meryanic, and perhaps Mari), and the Middle Kama center could correspond to the Central Uralic branches (Permic and perhaps Mari). The original region of Mari is uncertain: this branch seems to share a surprisingly low number of pairwise words and common innovations with both Permic (Metsäranta 2020: 285–286, 290–291) and Mordvin (Itkonen 1997: 259), so perhaps Mari was for a long time separated from those branches by unknown Uralic or non-Uralic neighboring languages.

The East Uralic branches (Hungarian, Mansi, Khanty, and Samoyedic) could be connected to the Cherkaskul Culture, which existed in the very same Central Ural Region between Middle Kama and Tobol after the Koptyaki Culture, ca. 1800–1500 BCE, as suggested by Parpola (2012a). The influence of this culture is visible also in the Upper Ob and Irtysh Region, which is considered as the homeland of Late Proto-Samoyedic. The Cherkaskul Culture is included among the Andronoid Cultures, the label reflecting a strong influence from the steppe Andronovo Complex; especially in ceramics the connection to the Fyodorovka Culture is clear (Grigoriev 2021: 24). There also appeared an expansion from the Cherkaskul Culture to the southern directions, to both sides of the Urals (Korochkova 2011: 28–29).

The Ugric branches possibly continued together in the Central Ural Region within the following Mezhovskaya Culture ca. 1500–1000 BCE. This archaeological framework presented by Parpola (2012a) matches nicely with the datings achieved from the Indo-Iranian loanword layers: the East Uralic sibilant changes were probably fully developed by 1500 BCE at the latest. The leap of Samoyedic to South Siberia separated it from the other Uralic branches for millennia to come, until much later the northward-advancing Samoyeds met the eastward-advancing Khanty between the Ob and Yenisei.

For Hungarian, Mansi, and Khanty the eastern side of the Urals is far enough: from there Hungarian moved first to the south and later to the west, while Mansi and Khanty moved to the north and northeast. The

forest-steppe zone extends to the river Iset, so the Ugric horse-related vo-
cabulary from some unknown language could have been adopted right
there in the Central Ural Region.

There are already recent models of dispersal agreeing with the rejuve-
nated Proto-Uralic (Parpola 2012a; 2012b; 2017; 2022; Lang 2020), and it is
not possible to go through the whole dispersal process here. Suffice it to say
that connecting the spread of Saami and Finnic westwards from the Upper
Volga Region to the later stages of the Netted Ware tradition at the end of
the 2nd millennium BCE seems possible, although the linguistic results
could also agree with a somewhat later dispersal.

To conclude, the Seima-Turbino Network is only a partial match for the
early Uralic expansion: its later western extension might be connected to
the early dispersal of Uralic westwards from the Central Ural Region, but
its earlier eastern core region cannot be related to the expansion of Late
Proto-Uralic or the subsequent stages. Even during the stage U5 closing to
the mid-second millennium BCE, the Uralic proto-dialects appear to have
spread only within a narrow strand, reaching from the Volga–Oka Region
through the Lower/Middle Kama Region to the Central Ural Region. At
the same time, the Seima-Turbino Network had already reached its ulti-
mate width from Mongolia to Finland.

Nevertheless, as people and items spread quicker than languages, it
is possible that the Uralic speakers were somehow involved in the whole
width of the Seima-Turbino Network, but their number and proportional
density was sufficient to expand their language only within a few centers
close to their core area. Parpola (2012b: 159–160) has earlier proposed that
only the European side of the network was Uralic-speaking, while the Si-
berian side spoke Indo-Iranian. However, the new datings for the Seima-
Turbino Network in Siberia (Chernykh et al. 2017; Marchenko et al. 2017)
are too early for Indo-Iranian, so the language in the Siberian part of the
network must have been originally something else.

Mallory (2001) has proposed that the Indo-Iranian influence and a new
societal structure could be behind the Uralic expansion. There are indeed
several Late Proto-Indo-Iranian loanwords which could reflect a bronze-
trade context: metal tools *ora 'awl' and *waćara 'hammer, ax'; *ćišta 'wax',
which could be connected to bronze casting; numerals *ćęta '100' and
*ćasra '1000', as well as *a/ęrwa 'value', possibly connected to high-volume
bronze trade; and *asora 'lord, prince' connected to the new social organi-
zation (all these loanwords are from Holopainen 2019).

Additionally, I have recently suggested an Indo-Iranian etymology for the first component of a compound metal name U *ä(j)sVn-weć(k)V 'tin and/or lead' found only in Permic and Mansi but still preceding the East Uralic sound changes. U *ša/okara 'armor', found only in Khanty, could be already a Proto-Iranian loanword, yet still earlier than the East Uralic sound changes (J. Häkkinen 2023). The oldest body armor in the relevant region is a lamellar armor made of horn blades, found in the Seima-Turbino burial ground in Rostovka, in the Middle Irtysh Region (Koryakova & Epimakhov 2006: 107). Recent datings from Rostovka range mostly between 2200–2000 BCE (Marchenko et al. 2017; Chernykh et al. 2017).

However, there is nothing in these words pointing specifically towards the Seima-Turbino Network: the words could be related to other metallurgical traditions or cultures as well, for example to the Abashevo and the Sintashta Cultures. It seems probable that there were several factors behind the Uralic expansion: (1) contacts with the Indo-Iranians to the south and southwest, (2) contacts with the Seima-Turbino Network to the east, and (3) a critical location controlling the Central Ural Passage, which was of paramount importance for the trade routes running in the Eurasian forest zone.

Based on the known later development (prevailing of the Uralic languages both to the west and to the east from the Urals instead of Indo-Iranian, Paleo-West Siberian, or Paleo-North European languages), it seems that the Uralic speakers managed to capitalize on their strategic position to the maximum. The first wave of expansion was directed to the west from the Central Ural Region around the second quarter of the 2nd millennium BCE, and the second wave was advancing westwards from the Upper Volga Region (Finnic, Saami, Mordvin, Meryanic, and West Chudic; see Rahkonen 2013) from the late 2nd millennium BCE onward. In the east, only Samoyedic leaped far away to the Sayan Region probably at some point during the latter half of the 2nd millennium BCE, while the Ugric branches remained in the vicinity of the Central Ural Region for a long time. We cannot exclude the possibility of Para-Samoyedic languages existing earlier in Southwest Siberia, but the traces probably would have been wiped away by later successive expansions of Iranian, Yeniseian, and Turkic languages.

## 6. Conclusion

In this critical examination I have considered every relevant piece of evidence fulfilling the criteria for Late Proto-Uralic word and carrying locational evidentiary value, and I have mapped out the most accurate possible regions concerning individual pieces of evidence. The evidence leaves room for only one possible homeland for Late Proto-Uralic: the Central Ural Region.

However, the sphere of this homeland reaches towards the Middle Kama in the west and Middle Tobol in the east, partially overlapping with some recent homeland propositions (Parpola 2022: 270, 264; Saarikivi 2022: 56). The evidence dismisses homeland candidates further to the west (e.g. the Middle Volga Region and the Upper Volga Region) or to the east (e.g. the Middle Irtysh Region and the Sayan Region). Nevertheless, some of these rejected regions could be the homeland of some earlier stage preceding Late Proto-Uralic.

Another result of this scrutiny is a more resolute stratification of the Indo-Iranian loanword layers. Based on these loanword layers, a new model for the disintegration of Late Proto-Uralic is presented. This kind of flexible model is necessary to account for the non-abrupt, non-rigid disintegration process of Late Proto-Uralic. Moreover, through these loanword layers the Uralic reconstruction stages are anchored to the Indo-Iranian chronology. It is argued that even though Late Proto-Uralic was divided into two centers (Samoyedic and Finno-Ugric) already soon after ca. 2500 BCE, both centers must have remained close to each other until ca. 2000 BCE.

Only in the first quarter of the 2nd millennium BCE did the Uralic speech area disperse into a narrow strand reaching from the Upper Volga Region (> Saami, Finnic, West Chudic, Meryanic, Mordvin, and possibly Mari) through the Volga-Kama confluence (> Permic, possibly Mari and possible extinct branches between them) to the Central Ural Region (> Hungarian, Mansi, Khanty, and Samoyedic). This dispersal led to several Uralic pre-dialects, probably corresponding to the branch ancestors. During the second quarter of the 2nd millennium BCE, the first macro-branch-specific sound changes occurred, giving rise to the Uralic proto-dialects (probably four regional units: West Uralic, East Uralic, Mari, and Permic). Only ca. 1500 BCE could the individual branches have begun to advance farther from each other, and this concerns also Samoyedic, which must be located in the vicinity of the Central Ural Region until that time.

## Abbreviations

AIE — Archaic Indo-European: any dialect following Late Proto-Indo-European but not yet showing branch-specific sound changes

AIr — Archaic Iranian: between Proto-Iranian and Old Iranian

AU = U2 — Ancient Uralic: reorientating proto-language

BU = U3 — Bicentric Uralic: bicentric proto-language

CU = U4 — Common Uralic: several pre-dialects, showing independent sound substitutions

DU = U5 — Diverging Uralic: several proto-dialects, showing shared sound changes

EPIE — Early Proto-Indo-European: the common ancestor to all the Indo-European languages

EPIIr — Early Proto-Indo-Iranian

EPU — Early Proto-Uralic: a recent ancestor, preceding Late Proto-Uralic by several centuries; a stage during which the earliest Archaic Indo-European loanwords were possibly borrowed

EU — East Uralic (comprising Hungarian, Mansi, Khanty, and Samoyedic)

IPA — International Phonetic Alphabet

KL — East Mansi dialect of Lower Konda

KM — East Mansi dialect of Middle Konda

KU — East Mansi dialect of Upper Konda

LL — West Mansi dialect of Lower Lozva

LM — West Mansi dialect of Middle Lozva

LPIE — Late Proto-Indo-European: the common ancestor to all the Indo-European branches after Anatolian and Tocharian split off

LPIIr — Late Proto-Indo-Iranian

LPU = U1 — Late Proto-Uralic: immediately preceding the disintegration

LU — North Mansi dialect of Upper Lozva

MPIIr — Middle Proto-Indo-Iranian

NwIE — Northwest Indo-European: archaic dialect continuum of the predecessors of at least Balto-Slavic, Germanic, Celtic, and Italic

OIr — Old Iranian, during the 1st millennium BCE

Pe — West Mansi dialect of Pelymka

PIr — Proto-Iranian

PrePU — Pre-Proto-Uralic: a distant ancestor or long continuum preceding Late Proto-Uralic by several millennia

So — North Mansi dialect of Sosva

TCh — South Mansi dialect of Great Chandyri on Tavda

TY — South Mansi dialect of Yanichkova on Tavda

UPA — Uralic Phonetic Alphabet

VN — West Mansi dialect of Sotnikova on North Vagilsk

VNZ — West Mansi dialect of Zaozernaya on North Vagilsk

VS — West Mansi dialect of South Vagilsk

WU — West Uralic (comprising Saami, Finnic, Mordvin, and probably the extinct West Chudic and Meryanic)

# References

AgroAtlas = Afonin, A. N. & Greene, S. L. & Dzyubenko, N. I. & Frolov, A. N. (eds.). 2008. *Interactive agricultural ecological atlas of Russia and neighboring countries: Economic plants and their diseases, pests and weeds.* http://www.agroatlas.ru

Aikio, Ánte. 2006. On Germanic-Saami contacts and Saami prehistory. *Journal de la Société Finno-Ougrienne* 91. 9–55. https://doi.org/10.33340/susa.81944

Aikio, Ánte (Luobbal Sámmol Sámmol Ánte). 2014. The Uralic-Yukaghir lexical correspondences: Genetic inheritance, language contact or chance resemblance? *Finnisch-Ugrische Forschungen* 62. 7–76. https://doi.org/10.33339/fuf.86078

Aikio, Ánte (Luobbal Sámmol Sámmol Ánte). 2015. The Finnic 'secondary e-stems' and Proto-Uralic vocalism. *Journal de la Société Finno-Ougrienne* 95. https://doi.org/10.33340/susa.82642

Aikio, Ante (Luobbal Sámmol Sámmol Ánte). 2020. Uralic etymological dictionary: Draft version of entries A–Ć published online on Academia.edu.

Aikio, Ánte (Luobbal Sámmol Sámmol Ánte). 2022. Proto-Uralic. In Bakró-Nagy, Marianne & Laakso, Johanna & Skribnik, Elena (eds.), *The Oxford guide to the Uralic languages*, 3–27. Oxford: Oxford University Press. https://doi.org/10.1093/oso/9780198767664.003.0001

Aikio, Ante & Aikio, Aslak. 2001. Heimovaelluksista jatkuvuuteen: Suomalaisen väestöhistorian tutkimuksen pirstoutuminen. *Muinaistutkija* 4/2001. 2–21. http://www.sarks.fi/mt/pdf/2001_4.pdf

Anthony, David W. 2007. *The horse, the wheel, and language: How Bronze-Age riders from the Eurasian steppes shaped the modern world*. Princeton: Princeton University Press.

Anthony, David W. & Ringe, Don. 2015. The Indo-European homeland from linguistic and archaeological perspectives. *Annual Review of Linguistics* 2015: 1. 199–219. https://doi.org/10.1146/annurev-linguist-030514-124812

Beekes, Robert S. P. 2011. *Comparative Indo-European linguistics: An introduction.* Revised and corrected by Michiel de Vaan. Amsterdam/Philadelphia: John Benjamins Publishing Company.

Beke, Ödön. 1997–2001. *Mari nyelvjárási szótár* 1–9. Unter Mitarbeit von Zsófia M. Velenyák und József Erdődi, neu redigiert von Gábor Bereczki, bearbeitet von Margarita Kuznecova, herausgegeben von János Pusztay (Bibliotheca Ceremissica 4). Szombathely: Berzsenyi Dániel Tanárképző Főiskola.

Blyakharchuk, Tatiana A. & Chernova, Natalia A. 2013. Vegetation and climate in the Western Sayan Mts according to pollen data from Lugovoe Mire as a background for prehistoric cultural change in southern Middle Siberia. *Quaternary Science Reviews* 75. 22–42. http://dx.doi.org/10.1016/j.quascirev.2013.05.017

Byrd, Andrew Miles. 2018. The phonology of Proto-Indo-European. In Klein, Jared & Joseph, Brian & Fritz, Matthias (eds.), *Handbook of comparative and historical Indo-European linguistics*, Volume 3, 2056–2079. Berlin & Boston: De Gruyter Mouton. https://doi.org/10.1515/9783110542431-042

Cantera, Alberto. 2017. The phonology of Iranian. In Klein, Jared & Joseph, Brian & Fritz, Matthias (eds.), *Handbook of comparative and historical Indo-European linguistics*, Volume 1, 481–503. Berlin &Boston: De Gruyter Mouton. https://doi.org/10.1515/9783110261288-033

Cardona, George. 2017. The documentation of Indic. In Klein, Jared & Joseph, Brian & Fritz, Matthias (eds.), *Handbook of comparative and historical Indo-European Linguistics*, Volume 1. 309–325. Berlin &Boston: De Gruyter Mouton. https://doi.org/10.1515/9783110261288-025

Carpelan, Christian & Parpola, Asko. 2001. Emergence, contacts and dispersal of Proto-Indo-European, Proto-Uralic and Proto-Aryan in archaeological perspective. In Carpelan, Christian & Parpola, Asko & Koskikallio, Petteri (eds.), *Early contacts between Uralic and Indo-European: Linguistic and archaeological considerations* (Mémoires de la Société Finno-Ougrienne 242), 55–150. Helsinki: Société Finno-Ougrienne.

Chernykh, E. N. & Korochkova, O. N. & Orlovskaya, L. B. 2017. Issues in the calendar chronology of the Seima-Turbino Transcultural Phenomenon. *Archaeology, Ethnology & Anthropology of Eurasia*. 45/2. 45–55.

Derksen, Rick. 2015. *Etymological dictionary of the Baltic inherited lexicon* (Leiden Indo-European Etymological Series 13). Leiden & Boston: Brill.

Grigoriev, Stanislav. 2021. Andronovo problem: Studies of cultural genesis in the Eurasian Bronze Age. *Open Archaeology* 2021: 7. 3–36. https://doi.org/10.1515/opar-2020-0123

Grünthal, Riho & Heyd, Volker & Holopainen, Sampsa & Janhunen, Juha A. & Khanina, Olesya & Miestamo, Matti & Nichols, Johanna & Saarikivi, Janne & Sinnemäki, Kaius. 2022. Drastic demographic events triggered the Uralic spread. *Diachronica*, Volume 39, Issue 4. 490–524. https://doi.org/10.1075/dia.20038.gru

Häkkinen, Jaakko. 2007. *Kantauralin murteutuminen vokaalivastaavuuksien valossa*. University of Helsinki. (Master's thesis.) https://helda.helsinki.fi/handle/10138/19684

Häkkinen, Jaakko. 2009. Kantauralin ajoitus ja paikannus: perustelut puntarissa. *Journal de la Société Finno-Ougrienne* 92. 9–56. https://doi.org/10.33340/susa.82020

Häkkinen, Jaakko. 2010. Jatkuvuusperustelut ja saamelaisen kielen leviäminen. *Muinaistutkija* 1/2010. 19–36. http://www.sarks.fi/mt/pdf/2010_1.pdf

Häkkinen, Jaakko. 2012. After the protolanguage: Invisible convergence, false divergence and boundary shift. *Finnisch-Ugrische Forschungen* 61. 7–28. https://doi.org/10.33339/fuf.85674

Häkkinen, Jaakko. 2022. Recurring irregularities in West Uralic 1: Para-Slavic loanwords. In Kaheinen, Kaisla & Leisiö, Larisa & Erkkilä, Riku & Qiu, Toivo E. H. (eds), *Hämeenmaalta Jamalille: Kirja Tapani Salmiselle 07.04.2022*, 127–144. Helsinki: HELDA Open Books. https://doi.org/10.31885/9789515180858.11

Häkkinen, Jaakko. 2023. An old Uralic compound metal name. In Holopainen, Sampsa & Kim, Jeongdo & Metsäranta, Niklas (eds.), *Elämä ja etymologia: Janne Saarikiven 50-vuotisjuhlakirja*. 102–121. Helsinki: HELDA Open Books. https://doi.org/10.31885/9789515182838

Häkkinen, Kaisa. 1984. Wäre es schon an der Zeit, den Stammbaum zu fällen? *Ural-Altaische Jahrbücher, Neue Folge*, Band 4. 1–24.

Häkkinen, Kaisa. 1996. *Suomalaisten esihistoria kielitieteen valossa* (Tietolipas 147). Helsinki: Suomalaisen Kirjallisuuden Seura.

Hegedűs, Irén. 2012. The RUKI-rule in Nuristani. In Whitehead, Benedicte Nielsen & Olander, Thomas & Olsen, Birgit Anette & Rasmussen, Jens Elmegård (eds.), *The sound of Indo-European: Phonetics, phonemics, and morphophonemics* (Copenhagen Studies in Indo-European, Vol. 4). 145–167. Copenhagen: Museum Tusculanum Press.

Heggarty, Paul. 2007. Linguistics for archaeologists: Principles, methods and the case of the Incas. *Cambridge Archaeological Journal* 17:3. 311–340.

Holopainen, Sampsa. 2018. Indo-Iranian loans confined to Saami? In Holopainen, Sampsa & Saarikivi, Janne (eds.), *Perì orthótētos etýmōn* (*Περὶ ὀρθότητος ἐτύμων*): *Uusiutuva uralilainen etymologia* (Uralica Helsingiensia 11), 135–179. Helsinki: Suomalais-Ugrilainen Seura.

Holopainen, Sampsa. 2019. *Indo-Iranian borrowings in Uralic: Critical overview of the sound substitutions and distribution criterion*. University of Helsinki. http://urn.fi/URN:ISBN:978-951-51-5729-4 (Doctoral dissertation.)

Holopainen, Sampsa. 2021. On the question of substitution of palatovelars in Indo-European loanwords into Uralic. *Journal de la Société Finno-Ougrienne* 98. 197–233.

Hyllested, Adam. 2014. *Word exchange at the gates of Europe: Five millennia of language contact*. University of Copenhagen. (Doctoral dissertation.)

Iivonen, Antti. 2012. Kielten vokaalit kuuloanalogisessa vokaalikartassa. *Puhe ja kieli* 32:1. 17–43. https://journal.fi/pk/article/view/6332

Itkonen, Terho. 1997. Reflections on Pre-Uralic and the "Saami-Finnic protolanguage". *Finnisch-Ugrische Forschungen* 54,3. 229–266.

Janhunen, Juha. 2009. Proto-Uralic – what, where, and when? In Ylikoski, Jussi (ed.), *The quasquicentennial of the Finno-Ugrian Society* (Mémoires de la Société Finno-Ougrienne 258). 57–78. Helsinki: Société Finno-Ougrienne. https://www.sgr.fi/sust/sust258/sust258_janhunen.pdf

Jasanoff, Jay H. 2017. The impact of Hittite and Tocharian: Rethinking Indo-European in the 20th century and beyond. In Klein, Jared & Joseph, Brian & Fritz, Matthias (eds.), *Handbook of comparative and historical Indo-European linguistics*, Volume 1, 220–238. Berlin &Boston: De Gruyter Mouton. https://doi.org/10.1515/9783110261288-018

KALLIO, PETRI. 2006. Suomen kantakielten absoluuttista kronologiaa. *Virittäjä* 1/ 2006. 2–25. http://www.kotikielenseura.fi/virittaja/hakemistot/jutut/2006_2.pdf

KALLIO, PETRI. 2008. On the 'Early Baltic' loanwords in Common Finnic. In Lubotsky, Alexander & Schaeken, Jos & Wiedenhof, Jeroen (eds.), *Evidence and counter-evidence: Essays in honour of Frederik Kortlandt 1, Balto-Slavic and Indo-European linguistics* (Studies in Slavic and General Linguistics 32), 265–277. Leiden: Brill. https://doi.org/10.1163/9789401206358_020

KALLIO, PETRI. 2009. Stratigraphy of Indo-European loanwords in Saami. In Äikäs, Tiina (ed.), *Máttut – máddagat: The roots of Saami ethnicities, societies and spaces/places* (Publications of the Giellagas Institute 12). 30–45. Oulu: Giellagas Institute, University of Oulu.

KALLIO, PETRI. 2012. The non-initial-syllable vowel reductions from Proto-Uralic to Proto-Finnic. In Hyytiäinen, Tiina & Jalava, Lotta & Saarikivi, Janne & Sandman, Erika (eds.), *Per Urales ad Orientem: Iter polyphonicum multilingue, Festskrift tillägnad Juha Janhunen på hans sextioårsdag den 12 februari 2012* (Mémoires de la Société Finno-Ougrienne 264), 163–175. Helsinki: Société Finno-Ougrienne. https://www.sgr.fi/sust/sust264/sust264_kallio.pdf

KALLIO, PETRI. 2015a. The stratigraphy of the Germanic loanwords in Finnic. In Askedal, John Ole & Nielsen, Hans Frede (eds.), *Early Germanic languages in contact*, 23–38. Amsterdam: John Benjamins Publishing Company. https://doi.org/10.1075/nss.27.02kal

KALLIO, PETRI. 2015b. The language contact situation in prehistoric Northeastern Europe. In Mailhammer, Robert & Vennemann, Theo & Olsen, Birgit Anette (eds.), *The linguistic roots of Europe: Origin and development of European languages* (Copenhagen Studies in Indo-European, Vol. 6). 77–102. Copenhagen: Museum Tusculanum Press.

KALLIO, PETRI. Forthcoming. *The Proto-Balto-Slavic loanword stratum in the Uralic languages*. Manuscript.

KANNISTO, ARTTURI. 2013. *Wogulisches Wörterbuch*. Bearbeitet von Vuokko Eiras, herausgegen von Arto Moisio (Lexica Societatis Fenno-Ugricae XXXV). Helsinki: Société Finno-Ougrienne.

KIM, RONALD I. 2018. The phonology of Balto-Slavic. In Klein, Jared & Joseph, Brian & Fritz, Matthias (eds.), *Handbook of comparative and historical Indo-European linguistics*, Volume 3. 1974–1985. Berlin &Boston: De Gruyter Mouton. https://doi.org/10.1515/9783110542431-036

KOIVULEHTO, JORMA. 1991. *Uralische Evidenz für die Laryngaltheorie* (Österreichische Akademie der Wissenschaften, Philosophisch-Historische Klasse, Sitzungsberichte, 566. Band). Wien: Verlag der österreichischen Akademie der Wissenschaften.

KOIVULEHTO, JORMA. 2001. The Earliest contacts between Indo-European and Uralic speakers in the light of lexical loans. In Carpelan, Christian & Parpola, Asko & Koskikallio, Petteri (eds.), *Early contacts between Uralic and Indo-European: Linguistic and archaeological considerations* (Mémoires de la Société Finno-Ougrienne 242), 235–264. Helsinki: Société Finno-Ougrienne.

Koivulehto, Jorma. 2002. Contact with non-Germanic languages II: Relations to the east. In Bandle, Oscar & Braunmüller, Kurt & Jahr, Ernst Hakon & Karker, Allan & Naumann, Hans-Peter & Telemann, Ulf & Elmevik, Lennart & Widmark, Gun (eds.), *The Nordic languages: An international handbook of the history of the North Germanic languages*, 583–594. Berlin & New York: Walter de Gruyter.

Korochkova, Olga Nikolaevna. 2011. *Взаимодействие культур в эпоху бронзы в Среднем Зауралье и подтаежном Тоболо-Иртышье: Факторы, механизмы, динамика* [Interaction of cultures during the Bronze Age in the Central Trans-Urals and the southern taiga Tobol-Irtysh Region: Factors, mechanisms, dynamics]. Russian Academy of Sciences. (Summary of doctoral dissertation.)

Korochkova, Olga Nikolaevna. 2012. Первый металл в культуре таежных аборигенов Западной Сибири и Среднего Урала [The first metal in the culture of the taiga natives of Western Siberia and the Central Urals]. *Вестник НГУ, Серия: История, филология*, Т. 11, Вып. 3. 137–147.

Korochkova, O. N. (Корочкова, О. Н.) 2019. Среднее Зауралье и таежное Приобье в системе связей Западноазиатской металлургической провинции [Middle Trans-Urals and Taiga Ob' in the System of Relations of the West Asian Metallurgical province]. *Вестник Санкт-Петербургского университета: История*, Т. 64, Вып. 2 [*Vestnik of Saint Petersburg University: History*, Vol. 64, iss. 2]. 718–739. https://doi.org/10.21638/11701/spbu02.2019.216

Korochkova, Olga Nikolaevna & Kuzminykh, Sergei Vladimirovich & Serikov, Yuri Borisovich & Stefanov, Vladimir Ivanovich. 2010. Metals from the ritual site of Shaitanskoye Ozero II (Sverdlovsk Oblast, Russia). *Trabajos de prehistoria* 67, N.º 2. 489–499. http://dx.doi.org/10.3989/tp.2010.10052

Koryakova, Ludmila & Epimakhov, Andrej. 2006. *The Urals and Western Siberia in the Bronze and Iron Ages* (Cambridge World Archaeology). Cambridge University Press.

Kuokkala, Juha. 2018. Finnic-Saamic labial vowels in non-initial syllables: An etymological evaluation. In Holopainen, Sampsa & Saarikivi, Janne (eds.), *Perì orthótētos etýmōn* (Περὶ ὀρθότητος ἐτύμων): *Uusiutuva uralilainen etymologia* (Uralica Helsingiensia 11), 11–74. Helsinki: Société Finno-Ougrienne.

Kuz'mina, Elena. 2007. *The origin of the Indo-Iranians*. Mallory, J. P. (ed.) (Leiden Indo-European Etymological Dictionaries series 3). Leiden: Brill.

Kuz'mina, O. V. 2021. The Abashevo Culture. *Вестник Санкт-Петербургского университета: История*, Т. 66, Вып. 4 [*Vestnik of Saint Petersburg University: History*, Vol. 66, iss. 4]. 1206–1229. https://doi.org/10.21638/11701/spbu02.2021.411

Lang, Valter. 2020. *Homo Fennicus: Itämerensuomalaisten etnohistoria*. Helsinki: Suomalaisen Kirjallisuuden Seura.

Lapteva, Elena G. & Korona, Olga M. 2012. Holocene vegetation changes and anthropogenic influence in the forest-steppe zone of the Southern Trans-Urals based on pollen and plant macrofossil records from the Sukharysh cave. *Vegetation History and Archaeobotany* (2012) 21. 321–336. https://doi.org/10.1007/s00334-011-0333-z

LAPTEVA, E. G. & ZARETSKAYA, N. E. & KOSINTSEV, P. A. & LYCHAGINA, E. L. & CHERNOV, A. V. 2017. First data on the Middle to Late Holocene dynamics of vegetation in the Upper Kama Region. *Russian Journal of Ecology* 2017, Vol. 48, No. 4. 326–334.

LUBOTSKY, ALEXANDER. 1990. La loi de Brugmann et *$H_3$e*-. *La reconstruction des laryngales* (Bibliothèque de la Faculté de Philosophie et Lettres de l'Université de Liège, fascicule CCLIII). 129–136. Liège-Paris. https://hdl.handle.net/1887/2666

LUBOTSKY, ALEXANDER. 2018. The phonology of Proto-Indo-Iranian. In Klein, Jared & Joseph, Brian & Fritz, Matthias (eds.), *Handbook of comparative and historical Indo-European linguistics*, Volume 3, 1875–1888. Berlin & Boston: De Gruyter Mouton. https://doi.org/10.1515/9783110542431-031

MALLORY, J. P. 1989. *In search of the Indo-Europeans: Language, archaeology and myth*. London: Thames and Hudson.

MALLORY, J. P. 2001. Uralics and Indo-Europeans: Problems of time and space. In Carpelan, Christian & Parpola, Asko & Koskikallio, Petteri (eds.), *Early contacts between Uralic and Indo-European: Linguistic and archaeological considerations* (Mémoires de la Société Finno-Ougrienne 242). 345–366. Helsinki: Société Finno-Ougrienne.

MALLORY, J. P. & ADAMS, D. Q. (eds.). 1997. *Encyclopedia of Indo-European culture*. London & Chicago: Fitzroy Dearborn Publishers.

MALLORY, J. P. & ADAMS, D. Q. 2006. *The Oxford introduction to Proto-Indo-European and the Proto-Indo-European world*. Oxford & New York: Oxford University Press.

MARCHENKO, Z. V. & SVYATKO S. V. & MOLODIN V. I. & GRISHIN A. E. & RYKUN M. P. 2017. Radiocarbon chronology of complexes with Seima-Turbino type objects (Bronze Age) in Southwestern Siberia. *Radiocarbon* 2017. 1–17. https://doi.org/10.1017/RDC.2017.24

MATASOVIĆ, RANKO. 2005. Toward a relative chronology of the earliest Baltic and Slavic sound changes. *Baltistica* XL(2). 147–157.

MATVEEV, A. V. 2011. *Материалы по мансийской топонимии горной части Северного Урала* [Materials on Mansi toponymy in the mountainous part of the Northern Urals]. Министерство образования и науки Российской Федерации Государственное образовательное учреждение высшего профессионального образования.

METSÄRANTA, NIKLAS. 2020. *Periytyminen ja lainautuminen: Marin ja permiläisten kielten sanastontutkimusta*. University of Helsinki. (Doctoral dissertation.) http://urn.fi/URN:ISBN:978-951-51-6870-2

MOISIO, ARTO & SAARINEN, SIRKKA. 2008. *Tscheremissisches Wörterbuch*. Aufgezeichnet von Volmari Porkka, Arvid Genetz, Yrjö Wichmann, Martti Räsänen, T. E. Uotila und Erkki Itkonen (Lexica Societatis Fenno-Ugricae XXXII). Helsinki: Suomalais-Ugrilainen Seura.

MORGUNOVA, N. L. & KHOKHLOVA, O. S. 2013. Chronology and periodization of the Pit-Grave Culture in the region between the Volga and Ural rivers based on radiocarbon dating and paleopedological research. In Jull, A. J. T. & Hatté, C. (eds.), *Proceedings of the 21st International Radiocarbon Conference*, Part 2 (*Radiocarbon* Vol 55, Nr 3), 1286–1296. https://doi.org/10.1017/S0033822200048190

NICHOLS, JOHANNA. 2021. The origin and dispersal of Uralic: Distributional typological view. *Annual Review of Linguistics* 2021: 7. 351–369.

NICHOLS, JOHANNA & RHODES, RICHARD A. 2018. Vectors of language spread at the central steppe periphery: Finno-Ugric as catalyst language. In Kroonen, Guus & Iversen, Rune (eds.), *Digging for words*, 58–68. Oxford: BAR Publishing.

OLLETT, ANDREW. 2014. Evidence of laryngeal coloring in Proto-Indo-Iranian. *Historische Sprachforschung = Historical Linguistics* 127. 150–165.

PARPOLA, ASKO. 2012a. The problem of Samoyed origins in the light of archaeology: On the formation and dispersal of East Uralic (Proto-Ugro-Samoyed). In Hyytiäinen, Tiina & Jalava, Lotta & Saarikivi, Janne & Sandman, Erika (eds.), *Per Urales ad Orientem: Iter polyphonicum multilingue, Festskrift tillägnad Juha Janhunen på hans sextioårsdag den 12 februari 2012* (Mémoires de la Société Finno-Ougrienne 264), 227–250. Helsinki: Société Finno-Ougrienne. https://www.sgr.fi/sust/sust264/sust264_parpola.pdf

PARPOLA, ASKO. 2012b. Formation of the Indo-European and Uralic (Finno-Ugric) language families in the light of archaeology: Revised and integrated "total" correlations. In Grünthal, Riho & Kallio, Petri (eds.), *A linguistic map of prehistoric northern Europe* (Mémoires de la Société Finno-Ougrienne 266), 119–184. Helsinki: Société Finno-Ougrienne.
https://www.sgr.fi/sust/sust266/sust266_parpola.pdf

PARPOLA, ASKO. 2017. Finnish *vatsa* ~ Sanskrit *vatsá* and the formation of Indo-Iranian and Uralic languages. *Journal de la Société Finno-Ougrienne* 96. 245–286.

PARPOLA, ASKO. 2022. Location of the Uralic proto-language in the Kama River valley and the Uralic speakers' expansion east and west with the 'Sejma-Turbino Transcultural Phenomenon' 2200–1900 BC. *Археология евразийских степей* 2022(2). https://doi.org/10.24852/2587-6112.2022.2.258.277

POKORNY, JULIUS. 2007. *Proto-Indo-European etymological dictionary: A revised edition of Julius Pokorny's Indogermanisches Etymologisches Wörterbuch*. Dnghu Association.

RAHKONEN, PAULI. 2013. *Southeastern contact area of Finnic languages in the light of onomastics*. University of Helsinki.
https://helda.helsinki.fi/handle/10138/38908 (Doctoral dissertation.)

RINGE, DON. 2017. Indo-European dialectology. In Klein, Jared & Joseph, Brian & Fritz, Matthias (eds.), *Handbook of comparative and historical Indo-European linguistics.*, Volume 1, 62–75. Berlin & Boston: De Gruyter Mouton.
https://doi.org/10.1515/9783110261288-006

SAAG, LEHTI & VASILYEV, SERGEY V. & VARUL, LIIVI & KOSORUKOVA, NATA-
    LIA V. & GERASIMOV, DMITRI V. & OSHIBKINA, SVETLANA V. & GRIFFITH,
    SAMUEL J. & SOLNIK, ANU & SAAG, LAURI & D'ATANASIO, EUGENIA & METS-
    PALU, ENE & REIDLA, MAERE & ROOTSI, SIIRI & KIVISILD, TOOMAS & SCHEIB,
    CHRISTIANA LYN & TAMBETS, KRISTIINA & KRIISKA, AIVAR & METSPALU,
    MAIT. 2021. Genetic ancestry changes in Stone to Bronze Age transition in the
    East European plain. *Science Advances* Vol 7, Issue 4.
    https://doi.org/10.1126/sciadv.abd6535
SAARIKIVI, JANNE. 2022. The divergence of Proto-Uralic and its offspring: A
    descendent reconstruction. In Bakró-Nagy, Marianne & Laakso, Johanna &
    Skribnik, Elena (eds.), *The Oxford guide to the Uralic languages*, 28–58. Oxford
    University Press.
SAARIKIVI, JANNE & LAVENTO, MIKA. 2012. Linguistics and archaeology: A crit-
    ical view of an interdisciplinary approach with reference to the prehistory of
    Northern Scandinavia. In Damm, Charlotte & Saarikivi, Janne (eds.). *Net-
    works, interaction and emerging identities in Fennoscandia and beyond: Papers
    from the conference held in Tromsø, Norway, October 13–16 2009* (Mémoires de
    la Société Finno-Ougrienne 265), 177–216. Helsinki: Société Finno-Ougrienne.
    https://www.sgr.fi/sust/sust265/sust265_saarikivi_and_lavento.pdf
SALMINEN, TAPANI. 1989. Classification of the Uralic languages. In Grünthal,
    Riho & Penttinen, Sirpa & Salminen, Tapani (eds.), *IFUSCO 1988: Proceedings
    of the Fifth International Finno-Ugrist Students' Conference, Helsinki, 22–26
    May 1988* (Castrenianumin toimitteita 35), 15–24. Helsinki: Suomalais-Ugri-
    lainen Seura & Helsingin yliopisto.
SALMINEN, TAPANI. 2002. Problems in the taxonomy of the Uralic languages in
    the light of modern comparative studies. *Лингвистический беспредел: Сбор-
    ник статей к 70-летию А. И. Кузнецовой.* 44–55. Москва: Издательство
    Московского университета.
SAMMALLAHTI, PEKKA. 1988. Historical phonology of the Uralic Languages.
    Sinor, Denis (ed.), *The Uralic languages*, 478–554. Leiden: Brill.
SHUMILOVSKIKH, LYUDMILA S. & SCHMIDT, MONIKA & PERESKOKOV, MIKHAIL &
    SANNIKOV, PAVEL. 2020. Postglacial history of East European boreal forests in
    the mid-Kama region, pre-Urals, Russia. *Boreas* 49(3). 526–543.
    https://doi.org/10.1111/bor.12436
SIMON, ZSOLT. 2020. Urindogermanische Lehnwörter in den uralischen und
    finno-ugrischen Grundsprachen: Eine Fata Morgana? *Indogermanische For-
    schungen* 125, issue 1. 239–265.
SKJÆRVØ, PRODS OKTOR. 2017. The documentation of Iranian. In Klein, Jared &
    Joseph, Brian & Fritz, Matthias (eds.), *Handbook of comparative and historical
    Indo-European linguistics*, Volume 1, 471–481. Berlin & Boston: De Gruyter
    Mouton. https://doi.org/10.1515/9783110261288-032

SSA = Itkonen, Erkki & Kulonen, Ulla-Maija (eds.). 1992–2000. *Suomen sanojen alkuperä: Etymologinen sanakirja* 1–3 (Suomalaisen Kirjallisuuden Seuran Toimituksia 556 & Kotimaisten kielten tutkimuskeskuksen julkaisuja 62). Helsinki: Suomalaisen Kirjallisuuden Seura & Kotimaisten kielten tutkimuskeskus.

Tirkkonen, Jani-Matti. 2012. ”*Lopullinen totuus pohjoiseurooppalaisten alkuperästä*”? *Kalevi Wiik vallankumouksellisen paradigman edustajana juuretkiistassa.* University of Eastern Finland. (Master's thesis.)

UEW = Rédei, Károly. 1986–1991. *Uralisches Etymologisches Wörterbuch.* Band I–II: Unter Mitarbeit von Marianne Bakró-Nagy, Sándor Csúcs, István Erdélyi †, László Honti, Éva Korenchy †, Éva K. Sal und Edit Vértes, Band III: Register Zusammengestellt von Attila Dobó und Éva Fancsaly. Wiesbaden: Harrassowitz.

Underhill, Peter A. & Poznik, G. David & Rootsi, Siiri & Järve, Mari & Lin, Alice A. & Wang, Jianbin & Passarelli, Ben & Kanbar, Jad & Myres, Natalie M. & King, Roy J. & Di Cristofaro, Julie & Sahakyan, Hovhannes & Behar, Doron M. & Kushniarevich, Alena & Šarac, Jelena & Šaric, Tena & Rudan, Pavao & Pathak, Ajai Kumar & Chaubey, Gyaneshwer & Grugni, Viola & Semino, Ornella & Yepiskoposyan, Levon & Bahmanimehr, Ardeshir & Farjadian, Shirin & Balanovsky, Oleg & Khusnutdinova, Elza K. & Herrera, Rene J. & Chiaroni, Jacques & Bustamante, Carlos D. & Quake, Stephen R. & Kivisild, Toomas & Villems, Richard. 2014. The phylogenetic and geographic structure of Y-chromosome haplogroup R1a. *European Journal of Human Genetics* (2014). 1–8. https://doi.org/10.1038/ejhg.2014.50

Vilppula, Matti. 1984. Lisiä lehmuksen, jalavan ja lehmuksen kuoriaineksen nimityksiin. *Virittäjä* 88(2). 193–200.

Volkova, V. S. & Kuz'mina, O. B. & Gnibidenko, Z. N. & Golovina, A. G. 2016. The Paleogene/Neogene boundary in continental deposits of the West Siberian Plain. *Russian Geology and Geophysics* 57. 303–315.

Vybornov, Aleksandr A. & Lychagina, Eugeniya L. & Kulkova, Marianna A. & Melnichuk, Andrey F. 2021. Periodization and Chronology of the Eneolithic in the Upper and Middle Kama Regions. *Materiality and objects: Multi-disciplinary approaches to archaeological material and contexts, Proceedings of the Finnish-Russian Archaeological Symposium, Hämeenlinna, 10-12 October 2019* (*Iskos* 24). 7–20.

Witzel, Michael. 2001. Autochthonous Aryans? The evidence from Old Indian and Iranian texts. *Electric Journal of Vedic Studies* 7:3. 1–93.

Zhivlov, Mikhail. 2015. *Неиндоевропейский субстрат в финно-волжских языках.* (Presentation manuscript.)

*Sampsa Holopainen*
University of Helsinki & University of Vienna

# Notes on an old problem of Hungarian historical vocalism: the sporadic (?) change of Uralic *u > Hungarian a, á

This article discusses the alleged sound change Proto-Uralic *u > Hungarian a, á. The etymologies manifesting this change that have been presented in earlier etymological literature are critically examined, and it is shown that a significant portion of them are wrong or based on outdated reconstructions. New explanations for many etymologies are presented, and possible causes for the few convincing examples of *u > a are discussed.

## 1. Introduction

The purpose of this article is to investigate certain problems of Hungarian historical vocalism and discuss the methodological problems involved in postulating "sporadic" sound changes and tendencies instead of regular sound laws.[1] The article consists of a presentation of methodology and an overview of recent studies on Hungarian historical phonology, including

---

the discussion of Proto-Ugric vocalism and Old Hungarian evidence. After that, the suggested etymologies manifesting the lowering *u > a or *u > á are discussed, and in the end, conclusions are presented.

The assumption of sporadic sound change is contrary to the Neogrammarian principle of regular sound change or sound law; the latter assumes that change is always regular (under the same conditions, the same phoneme changes in certain ways), whereas the former view assumes that sounds change in unexpected ways, and no regularity can be assumed. A term "tendency" is used by some researchers (such as Csúcs 2005; Róna-Tas 2017) to denote such sporadic change.

The development of Uralic *u in Hungarian is a good example of a situation where numerous different modern Hungarian vowels (at least *a, á, o, u, ú, ë*; Csúcs et al. 1991: 22–37, 65–66) have been assumed to reflect the same Proto-Uralic phoneme, without clear rules or conditions. In this article, the examples of an alleged sound change *u > a, á are analyzed, and it is shown that the development of *u in Hungarian is much more regular than has been hitherto assumed. The results show that resorting to evidence for "key languages", notably Finnic, has resulted in a misleading picture of Proto-Uralic vowel reconstruction and thus has also led to erroneous views on vowel developments in the prehistory of Hungarian (see Kallio 2012 on the problems with Finno-centric vowel reconstructions). Although the change *u > a, á might seem like too marginal an issue for one article, there are actually surprisingly many etymologies showing this alleged change. I intend to return in future articles to other irregular changes in the prehistory of Hungarian.

In studies on Uralic historical phonology, the idea that sound change is sporadic has played a significant role, and this line of thought has been especially pursued in Uralic research done in Hungary (see e.g. Csúcs 2005: 8–9; WOT: 1036–1037; Honti 2013: 6; Róna-Tas 2017: 79; also Gerstner 2018 speaks of "tendencies" in his article on historical phonology in the most recent handbook of the history of Hungarian). Honti (2013: 6) sums up this attitude, stating that "sound changes are often less than 'sound laws';

usually they are mere 'tendencies of sound change'". On the other hand, since the late 1970s a more rigorous line of research and historical phonology has emerged, represented notably by Pekka Sammallahti (1988) and Juha Janhunen (1981). The rigorous "Neogrammarian" line of research has gained recognition in more recent publications, and the idea of sporadic sound change has been criticized, with convincing arguments being presented in favor of the regularity of sound change (see especially Zhivlov 2010; 2014 and Aikio 2012; 2013b: 161; 2015; see also Abondolo 1996: 3–4 and Reshetnikov & Zhivlov 2011); Zhivlov (2014: 113) firmly explains the stark difference in the two approaches, arguing that as the regularity of sound laws is a basic principle, no sporadic developments can be accepted in a serious study of historical phonology:

> The basic tenet of this methodology is the principle of regularity of sound laws (…). Taking this principle seriously means that we cannot invoke 'sporadic developments' as an explanation in historical phonology.

The most recent comprehensive studies on Uralic historical phonology and etymology follow the Neogrammarian approach; in addition to the ones mentioned above, Häkkinen (2007), Pystynen (2018), Metsäranta (2020), and Aikio (in preparation) should especially be mentioned as good examples.

In this paper, the methodological aim is to explain the developments with as little irregularity as possible, and consequently tendencies are *a priori* considered implausible explanations. It is naturally possible that some developments rejected in this paper will be revisited and rehabilitated by later research, if conditions for the apparent tendencies can be found.

Although the meaning of "tendency" is not often defined in the research literature, WOT (1036) and Róna-Tas (2017: 79) speak of strong and weak tendencies, assuming that strong tendencies are the ones that follow clear rules and show only very few exceptions or no exceptions at all. Weak tendencies, on the other hand, are the ones that most of the examples obey, but which show a significant number of exceptions. The strong tendencies, as described by Róna-Tas, can be compared to sound laws in that they usually have no exceptions. However, the idea of weak tendencies is, in my view, methodologically much more problematic, as it is difficult to determine how much irregularity is allowed in such cases.

It should be noted here that regular sound change is the mainstream view of historical linguistics (see handbooks like Anttila 1989: 57–65,

85–86; Campbell 1999: 17–18; Salmons 2020: 30). One can state that it would be impossible to work on historical phonology without expecting that phonemes change according to rules and not arbitrarily. The following quotation from Ringe (2004: 237) serves well to describe the situation and the importance of regular sound change:

> Modern work in sociolinguistics has shown that the scenario just summarized is slightly oversimplified; most importantly, sound changes pass through a variable phase before "going to completion," and occasionally the progress of a sound change is arrested in the variable phase, giving rise to irregularities (see, e.g., Labov 1994 for discussion). But the statistical preponderance of regular sound changes remains impressively massive, and it is almost always methodologically advisable to treat explanations involving irregular sound changes with suspicion.

A quite similar view is represented by Fox (1995: 136–137, 304), who acknowledges the sociolinguistic arguments against regularity in practice represented by Labov, but notes that in order for the comparative method to work, it is a necessary assumption that sound change is regular. Some other important handbooks of historical linguistics also accept the fact that sound change is not necessarily *absolutely* regular but it is still a useful or even necessary tool in historical linguistics, especially in research on sound change (see Campbell 1996; Kiparsky 2015: 70–72). Moreover, the exceptions to regular sound change often have some reasons behind them, such as taboo or the effects of word associations, sometimes also the role of spelling conventions, as listed by Kiparsky (2015: 70, endnote 14). Moreover, Campbell (1996) discusses such cases in detail and concludes that they do not present a serious obstacle to the idea that sound change is mostly regular. Kiparsky (2015: 72) also mentions lexical diffusion as one of the obstacles, but Labov (2020) offers a detailed discussion of this and concludes that even if sound change spreads gradually, it operates in a regular way.

It can be thus stated that all the major handbooks of historical linguistics stress the importance of regular sound change as the core of the comparative method, even if some irregularity is allowed. A clear statement against the regularity of sound change is presented by Clackson (2007: 31–33), who argues that it is not necessary to assume that *all* sound changes are regular. He notes that "most" sound changes are regular, and this gives enough proof for the historical-comparative method to work. In a way this can be understood in a similar vein as Ringe's quotation above, meaning

that some degree of regularity has to be assumed in order for the comparative method to work. Clackson's remark has been criticized by De Vaan (2008: 1230).

In the history of Hungarian, the idea of irregular and sporadic change has been especially influential and has persisted for an especially long time (cf. the references above). This situation is partly due to the complicated phonological developments that have taken place in the history of Hungarian, as well as in the Ob-Ugric languages (Khanty and Mansi) which are usually grouped together with Hungarian under the Ugric branch. As it has been more challenging to explain the Hungarian/Ugric developments than those in Finnic or Saami, for example, it is rather understandable that such researchers have assumed that sound change simply is not regular. However, this assumption is mistaken and cannot be substantiated. In more recent research, such as Aikio (2002; 2006; 2015), it is shown that the Hungarian developments fit the reconstructions resulting from a rigorous approach to regularity.

One must state here that Hungarian historical phonology, and especially the development of vocalism, forms a special case in Uralic studies. The main problem is that the broad lines of developments leading to Hungarian are known but many details are uncertain; this has also been stated by Aikio (2022: 5). In contrast to Proto-Permic vocalism, which includes similar problems, Hungarian vowel history has been approached more through a *laissez-faire* approach. In this respect, it resembles the study of Mari historical phonology (see Aikio 2014a: 142; 2022: 5), and also the study of Ob-Ugric vowel history (Zhivlov 2006), which is obviously closely related to the problems of Hungarian vocalism through the close relationship of Ob-Ugric and Hungarian. The only truly Neogrammarian approach to Hungarian vocalism is Sammallahti (1988), which is by now outdated in certain points. An additional problem in Sammallahti's presentation is that it is heavily based on the traditional, binary classification of the Uralic family. Because of this, Sammallahti frequently projects some changes to Proto-Ugric without a real need, clinging onto the binary classification of Hungarian and Ob-Ugric. The problems in the reconstruction of Proto-Ugric vocalism and their relevance for the present article will be explained below.

An additional source of problems in the research into Hungarian and other eastern Uralic languages is the role of Finnic as a key language. The problems of Finnic as a key language in traditional Uralic reconstructions have been discussed by Kallio (2012); see also Abondolo (1996: 3–4).

## 2. Recent studies in the historical phonology of Hungarian

The most recent overview of Hungarian historical phonology is found in WOT (1011–1069). Unfortunately, this overview is based mostly on the etymologies of the UEW, which makes the picture skewed, and moreover the presentation of WOT follows the outdated method of "tendencies" instead of sound laws. This means that WOT does not bring much new information about the historical phonology of Hungarian. Some phonological issues are also discussed in the brief monograph of Honti (2017), written as a criticism of WOT. Also, Róna-Tas (2017), in a reply to Honti, deals with some issues of historical phonology. On the whole, these recent works offer good explanations for some individual etymologies, but their scrutiny of phonological developments does not give satisfying explanations for problems of Hungarian historical phonology.

An important work on the Uralic background of Hungarian is the recent article of Aikio (2018) which deals with some specific issues of the development of Proto-Uralic consonants in Hungarian. Aikio presents two new sound laws for Hungarian: PU \**jŋ* > Hu *gy* (for example, PU \**ajŋi* > Hu *agy* 'brain', PU \**wajŋi* > Hu *vágy* 'lust') and PU \**nč* > *r* (for example, PU \**ponči* > Hu *far* 'backside', (derivative) *farok* 'tail'; PU \**künči* > Hu (derivative) *kör-öm* 'nail').

Relevant here is also the work of Tálos, who has in several works argued for a reconstruction of Proto-Uralic with two tongue-heights only, and has also written specifically on problems of Ugric and Hungarian historical phonology (see Tálos 1975; 1984). His views were followed in Abondolo's (1996) brief monograph that approaches the Uralic vowel history from an Ob-Ugric point of view, and many of Abondolo's remarks are also relevant for Hungarian. Even though the reconstructions and ideas of Proto-Uralic vowel rotation pursued by Tálos and Abondolo have not been widely accepted (but see Kümmel 2019; 2020, who frequently cites the Uralic reconstructions of Tálos), it is appropriate to mention them here as followers of a strict methodology. Their ideas have, in any case, been influential, and their remarks on many individual etymologies have proven useful. Even though the reconstruction of a Proto-Uralic high labial vowel \**u* is not influenced by the idea of two tongue heights, Abondolo's (1996) comments on the phonology of some Ob-Ugric etymologies will be relevant in this paper and they will be addressed below in the discussion of etymologies.

## 3. Problems in the reconstruction of Proto-Ugric vocalism

On the whole, the historical phonology of the Ugric languages (Hungarian, Khanty, and Mansi) is less well known than that of many other branches of Uralic, although some recent studies (Zhivlov 2006; 2014; Aikio 2015; 2018) have improved the situation. Especially problematic is the question of possible common sound changes shared by these three languages. Although Proto-Ugric changes are suggested by some sources like WOT and Sammallahti (1988), it remains unclear whether any common changes can be reconstructed when up-to-date etymological material is used. Some of these problems are discussed below.

The most problematic aspect of the question of an Ugric proto-language is that no commonly accepted reconstruction of Proto-Ugric phonology exists (see Bakró-Nagy 2013: 173–175 for a recent overview of certain problems of Ugric reconstruction). The only comprehensive, yet not widely accepted, presentations of Proto-Ugric are Sammallahti (1988) and WOT (1011–1069). It is actually methodologically rather surprising that most of the proponents of Proto-Ugric do not work with proper reconstructions of the Ugric proto-language, with even Honti (2017: 171), one of the main supporters of Ugric unity, noting that no commonly accepted reconstruction exits. The presentations of Sammallahti and WOT both include various problems, even though they are useful in many details. The biggest problems concern vocalism, and only those will be presented here in detail.

Sammallahti's (1988) reconstruction of vocalism can today be regarded as outdated in many points. A notable problem is that he assumes that Proto-Finno-Ugric, the stage preceding Proto-Ugric, had long vowels, which were then shortened in Proto-Ugric. This cannot be correct: Sammallahti's idea of Proto-Finno-Ugric long vowels is based on Janhunen's (1981) reconstruction of a Proto-Uralic phoneme *x that was vocalized in Proto-Finno-Ugric, and it has been convincingly demonstrated by Aikio (2012) that this idea is not correct. Sammallahti's Proto-Ugric reconstructions are thus for the most part identical with modern reconstructions of Proto-Uralic: for example, Proto-Ugric *ńïlï is identical with Aikio's Proto-Uralic *ńïli. The other problematic points in Sammallahti's reconstruction include the reconstruction of full and reduced vowels; it is unclear what this opposition is really based on.

There are also more general problems of a methodological nature. While Sammallahti's Proto-Ugric vowel history must be applauded for its attempt to follow the Neogrammarian principle rigorously, his attempt is complicated by the assumed binary structure of the Ugric and Uralic family tree. Because he assumes that intermediary forms have to be reconstructed for the vowel systems of Proto-Hungarian, Proto-Khanty, and Proto-Mansi, he resorts to complicated and flip-flopping changes. For example, Sammallahti (1988: 500, 504) assumes that Proto-Uralic *e* changes to Proto-Ugric *i* in *-i*-stems, but this change is hard to substantiate, as at least Hungarian shows no special development of *e* in such a context.

The same can be said of Sammallahti's Proto-Ob-Ugric reconstruction. It is necessary to keep in mind here that also no commonly accepted reconstruction of Proto-Ob-Ugric vocalism exists – also, Honti's (1982) Ob-Ugric vowel reconstruction has been criticized for postulating changes that are required only by strictly following a binary division of the Uralic family tree (Tálos 1984: 90, 97), but for which there is no actual evidence in Khanty and Mansi.

The problems with the Proto-Ugric phonology reconstructed by WOT are different in nature. It is also outdated, but mostly because it is based on the outdated material of the UEW (WOT also mentions Sammallahti's 1988 article as one source, but it remains unclear what parts of the presentation are based on that). WOT has to be given credit for its criticism of some of the UEW's more problematic etymologies and for commenting on some sound changes in greater detail, but as a whole the Proto-Ugric reconstruction is not very useful. The binary model problem is naturally true also in this case. An even bigger problem is the use of tendencies instead of sound changes (see WOT: 1036 for discussions of the methodological premises) to explain phonological developments. There are also problems in the use of Ob-Ugric evidence: for example, WOT assumes that in Proto-Ugric the vowels in the unstressed syllables became reduced, and only one vowel is reconstructed for this position, but this clearly cannot be correct, as the Uralic *-a-* and *-i*-stems have different reflexes in Khanty and Mansi, showing that they could not have merged in Proto-Ugric. Also the retention of vowels in the second syllable in the earliest Mansi written sources (from the 18th century) makes the idea of reduced vowels in Proto-Ugric quite unlikely (for example, "Old" Mansi *амба* [amba], modern North Mansi *āmp* 'dog', see Honti 1982: 126).

Although many details of Proto-Ugric are unclear, it is quite clear that there were few if any changes in vocalism common to the predecessor of Hungarian, Khanty, and Mansi. The vowel system that can be reconstructed on the basis of these three languages is very close to Proto-Uralic.

## 4. Notes on the Old Hungarian evidence

When Hungarian historical phonology is discussed, the situation is different from many other Uralic languages in that Hungarian has a long written history, going back to the eleventh century. However, the interpretation of Old Hungarian material often presents challenges, especially when the vowels are concerned.

The problems of Old Hungarian orthography are connected to sound changes that took place during the Old Hungarian period. Because lowering of vowels (*u* > *o*, *o* > *a*) indeed took place during this period, it is often difficult to determine what exactly the phonetic values of graphemes like *u* or *o* are, and often the material is open to various interpretations (see Bárczi 1958; Benkő 1980; E. Abaffy 2003).

Benkő (1980: 89–121) describes in detail the problems involved in the interpretation of Old Hungarian vowels. There are differing views among researchers on how the vowel graphemes in the Old Hungarian texts should be interpreted, and this is complicated by changes that took place over the Old Hungarian period. A notable problem concerning Old Hungarian *u* is that as both labial *å* and illabial *a* existed in the language at this period, there was a "chain shift" in the graphemes: if *a* was [a] and *o* was [å], then *u* was used sometimes to mark [o]. On the other hand, *u* was also used to mark [u] (Benkő 1980: 89–94; Korompay 2018: 87). Benkő (1980: 94–95) notes that it is far from certain what kind of linguistic situation this practice actually reflects.

Naturally, this does not mean that the Old Hungarian evidence would play no role at all in research into the history of Hungarian vocalism. But it means that much of the evidence is controversial, and all the etymologies should be investigated separately.

## 5. The case of Proto-Uralic *u and its reflexes in Hungarian

### 5.1. Overview of the problem

It is often argued that there are cases of PU *u being reflected by *a* and *á* in Hungarian: see for example Barczi (1958), Aikio (2002: 45, 48), and Maticsák (2020: 388). The UEW lists 12 cases of Hungarian *a*, two cases with Hungarian variants with *a* and *o*, and four cases of Hungarian *á* (Csúcs et al. 1991: 37); the etymologies with uncertain vowel reconstruction are not included in this calculation. Sammallahti (1988) assumes such a change in only five words, including reconstructions where *u is given as one possibility – the difference is partly due to different etymological material but also to different reconstructions of some Proto-Uralic words. Some additional examples have been suggested by Aikio (2002). There has been little discussion on the conditions of this change, however. The problem is that these cases are in the minority, as usually the reflex of PU *u is Hungarian *o* or *ú (such as PU *wuďi 'new' > Hu *új*, PU *kuńci 'urine' > Hu *húgy* 'piss'); some convincing examples of *u > Hu *u* are also known (such as PU *kuńa- 'close the eyes' > Hu *huny*). The conditions for the different developments are not quite clear, but some cases of long *ú* can be explained through contraction caused by glides preceding or following the vowel, such as PU *uji- 'swim' > Hu (derivative) *úsz-ik* (see also the example of Hu *új* above).

The development of Proto-Uralic *u in Hungarian is thus far from settled, and it would require more than one paper to solve this question. However, the words showing *u > *a* or *á* are a good place to start investigating the problem, as this group of words seems to include several unclear etymologies with competing explanations. In sieving out the problematic etymologies displaying this alleged sound change, the way is opened to investigating the problems of PU *u in Hungarian on the basis of more reliable etymological material.

The history of Proto-Uralic *u is also complicated by the fact that some words allegedly showing Proto-Uralic *u with aberrant reflexes in Hungarian should be reconstructed with PU *ẹ instead, such as PU *jẹxi- (UEW: 103 *juγe- (*juke-)) 'drink' > Fi *juo-*, Hu *i-*, *iv-* (see Zhivlov 2014: 115–117).

It should be added here that even though the reflexes of other Uralic back vowels in Hungarian are better known and regular developments have been suggested, there are also problems with the reflexes of Proto-Uralic *a in Hungarian, as both long *á* and short *a* are found as reflexes of this vowel

(compare PU *ćara- 'dry' > Hu száraz but PU *pata 'cauldron' > Hu fazék).
A possible solution is presented by Zhivlov (2014: 117–124), but many details still remain unclear.

In the following, the etymologies possible showing Hungarian long vowels a or á as the reflexes of Proto-Uralic *u are critically analyzed. It will be determined whether the etymologies themselves are valid, and in the cases where the Uralic or Ugric etymology turns out to correct, it will be discussed whether the vowel *u can indeed be reconstructed as the predecessor of Hungarian a or á. After analyzing the etymologies, possible causes for the different reflexes are briefly discussed.

The etymological material has been collected from the main Uralic etymological sources: the UEW, Sammallahti (1988), and the recent articles of Aikio (2002; 2006; 2015), as well as Aikio's (2013a) handout which lists the Uralic words with back vocalism.

## 5.2. Suggested Uralic and Ugric etymologies
###  allegedly manifesting *u > Hu a or á

PU *jupta- 'tell' > (?) Hu  játszik, (?) OHu 1198 ioatec, Fi juttele- 'speak, tell', juttu 'tale', Md jovtams, joftams 'tell, say', Ngan đebtaðasa (< PSam *jəptə- 'speak, tell', cognate also in Selkup, Janhunen 1977: 35) (UEW: 104; Helimski 1999; Aikio 2002: 48; 2013a)

The Uralic (or Finno-Ugric) background of the Hungarian word is an old idea (see the references in the UEW). However, it is considered uncertain by EWUng (640) and UEW (s.v. jukta-) due to semantics, but the possible connection to Proto-Uralic *juktV- 'tell' is mentioned. EWUng (640) notes that u > a in Hungarian is unusual. EWUng notes that the oldest meaning of the Hungarian word játszik is 'tell (erzählen)' rather than 'play' as is prevalent in modern Hungarian. This is close to the meanings attested in related languages, but it is unclear whether we are really dealing with the same word in Old and Modern Hungarian. The word does not appear in Sammallahti's (1988) list of Proto-Uralic words, and SSA does not mention Hungarian játszik among the cognates of Finnish juttu, juttele-.

Regarding the Uralic reconstruction, in earlier sources such as the UEW *juktV- was preferred, but we now know that *juptV- has to be the correct reconstruction, thanks to Helimski (1999) who added the Samoyedic cognate – the earlier reconstruction would have accounted for the Hungarian, Finnic, and Mordvin forms, but Samoyedic requires *pt as *kt would

have developed into *\*t* in Proto-Samoyedic (Sammallahti 1979: 46–47). As there are no other Hungarian words that derive from a Uralic word with a cluster *\*pt*, it is difficult to assume whether the vocalization of the stop *\*p* could have played a role in the development of the vocalism. However, the etymology remains uncertain because of the semantic problems mentioned above, and an uncertain etymology can have only little value in discussions of Hungarian historical vocalism.

PU *\*juri-* 'spin' > (?) Hu *jár* 'go', Kh Trj *jŏrəγʟə-* 'forget' (< PKh *\*jurəγʟə-*), Ms P *jōrl-* id. (< PMs *\*jɔryʟ*), SaN *jorrat* 'go around', Ud *jiromi̮-* 'go astray', TN *yurə°-* 'forget' (< PSam *\*jürə-*) (UEW: 102; Aikio 2002: 46–48)

Reconstructing Proto-Uralic *\*juri-* is convincing based on Aikio's (2002: 46–48) argumentation, but the relationship of Hungarian *jár* 'go' to this word family remains problematic. The UEW reconstructs Proto-Uralic *\*jori-*, but Aikio (2002) argues that the Proto-Uralic word probably had *\*u* instead. It should be noted that the UEW considers Hungarian *jár* an uncertain reflex of this Uralic stem. There are some problems in the reconstruction of the Uralic word's vocalism, as the Proto-Samoyedic vowel *\*ü* is irregular, but the Saami, Permic, and Ob-Ugric cognates point clearly to *\*u* (Aikio 2002: 47–48). Aikio notes that *\*j-* might have caused the secondary fronting in Samoyedic. Aikio also points to the irregularity of *u > á* but notes that there are parallels for this irregular lowering, though he does not discuss the issue in detail. Although Hungarian *jár* is discussed by Aikio (2006) as a cognate to PU *\*juri-*, the etymology is not mentioned in Aikio's (2013a) list of words. EWUng also mentions PFU *\*jorkV-* as a possible pre-form for the Hungarian, while the UEW (102) also reconstructs a proto-form *\*jorkV-* and lists Hungarian *jár* as an uncertain cognate.

However, it is possible that Hungarian *jár* is not a Uralic word at all. A Turkic etymology for *jár* has been suggested by Palló (1982: 123–125), who assumes a loan from Turkic *\*yor(ï)-* 'nomadize, wander' (> East Old Turkic *yorï-* id.). WOT (1200–1203) is critical towards the etymology, but the criticism stems more from the problematic connection of Hungarian *jár* to the verb *nyargal* 'gallop', both of which have been derived from the same Turkic source. As is noted by WOT, it is obvious that *jár* and *nyargal*[2] are

---

2. The etymology of *nyargal* 'gallop' is not clear, but it is interesting that many other Hungarian horse terms are of unknown origin, such as *nyerëg* 'saddle' (see Holopainen 2022: 108–109) that also features word-initial *ny-*. Hungarian *nyargal* might be a loan from a substrate language, like *nyerëg* probably is.

not regularly related (a change *j > *ny would be completely irregular), but it can be argued that *jár* could still be a Turkic loan. There seem to be no phonological or semantic problems in the Turkic etymology of *jár*: Palló notes that there are few examples of Turkic *o being reflected as Hungarian *á* in loanwords, but phonetically this substitution is not implausible. Furthermore, WOT (1120) lists some examples of this substitution, such as Hu *áporodik* 'decay' ← West Old Turkic *op(u)ra- 'grow old', Hu *váj* 'hollow out' ← West Old Turkic *vay- id., Hu *vályú* 'trough, tray' ← West Old Turkic *valuɣ 'trough'. As the Uralic background of *jár* is uncertain, the idea that the Hungarian word is borrowed from *yor(ï)- 'nomadize, wander' is a plausible etymology that can be rehabilitated.

There is one problem with the Turkic etymology, however, namely that there are no good parallels for the Turkic glide *y corresponding to *j in Hungarian.[3] While there are no phonetic problems in deriving *jár* from *yori-, the Turkic loans in Hungarian reflect the sound change *y > Oghuric *ǰ > Chuvash *ś* (for example, Hungarian *gyűrű* 'ring' ~ Chuvash *śerĕk* id.; Hungarian *szél* 'wind' ~ Chuvash *śil* id.); see WOT (1092–1093) for a discussion of the different reflexes. However, we must keep in mind that *y was retained in the "Common Turkic" branch (cf. East Old Turkic *yüzük* 'finger ring'). Although the majority of the early Turkic loanwords in Hungarian point to an Oghuric (Chuvash-type) donor language, Róna-Tas and Berta (WOT: 1071) admit that they cannot exclude the presence of non-Oghuric languages among the group they lump under the umbrella term "West Old Turkic", so a borrowing from an Common Turkic type language would probably be possible. Further research on this problem is clearly needed, but it does not seem to be an impossible idea to derive *jár* from a Turkic source that has *y-.

**PU/PUg** *kaďma* **(UEW:** *kuďmV***)** > Hu *hamu* 'ashes', Kh Vj *kajem* < PKh *kājm*, Ms TJ *kōľə̑m* < PMs *kūľm* id. (UEW: 194; Abondolo 1996: 93; Zhivlov 2014: 120)

Abondolo (1996: 93) has argued that Proto-Ugric *kaďma* is a derivative of *kaďa- 'leave' (the same explanation is presented also by Aikio *apud* Zhivlov 2014: 120). This is a plausible idea semantically, and *-ma is a known deverbal nominalizer in Proto-Uralic, so this etymology can be accepted. The *a vocalism presumed by this explanation is reflected regularly

---

3. I am grateful to Christopher Culver for pointing this out to me.

by Hungarian *a*, Proto-Khanty *\*ā*, and Proto-Mansi *\*ū* (cf. Zhivlov 2014: 124). The UEW's (194) idea of reconstructing *\*kudˊmV* is impossible, as none of the Ugric languages regularly point to *\*u*, and it is unclear why *\*u* was reconstructed by the UEW in the first place.

**PUg** *\*kaja-* **or** *\*koja-* > Hu *hajt* 'treiben, jagen', Ms So *χujt-* 'tempt' (< PMs *\*kujt-*) (UEW: 854)

The UEW gives two alternative reconstructions for this Proto-Ugric word, but neither is completely clear. Uralic or Ugric *\*a–a* stems are usually reflected as long *\*ū* in Mansi (PU *\*kala* 'fish' > Ms *\*kūl* id.). However, several examples of Proto-Mansi short *\*u* reflecting Uralic *\*o–a* stems can be found in the material of Aikio (2015: 60–62), such as PU *\*śona* 'sledge' > PMs *\*śun* id., PU *\*kodˊka* 'spirit' > PMs *\*kuľ* id., and PU *\*kompa* 'wave' > PMs *\*kump* id.

However, it is also possible that the words in Hungarian and Mansi are not related at all. The meanings of the two verbs are rather different: even though 'tempt' and 'drive, pursue' can probably be derived from a common source, the connection is not that obvious.

Furthermore, Aikio (2014b: 1–2) has recently connected the Mansi word to Proto-Khanty *\*kūć-* 'tempt' (> North Khanty *χuś-*; this was earlier, e.g. in SSA s.v. *kutsua*, connected cautiously to Finnish *kutsu-* (< Proto-Finnic *\*kuccu-*) and North Saami *gohccu-* (< Proto-Saami *\*koććō-*) but Aikio shows that the etymology is impossible due to irregular phonological correspondences; the Finno-Saami word is probably a loanword from Baltic *\*ku̯aiti̯a-* 'call; sue', as also noted by SSA as one possibility[4]). The correspondence between Mansi *\*kujt-* and Khanty *\*kūć-* is regular, and Proto-Uralic (Proto-Ob-Ugric?) *\*kujtV-* can be reconstructed as their common predecessor. It is probable that Hungarian *hajt* is unrelated, as it is also semantically quite far from the meaning 'call' or 'tempt' that can be reconstructed for the (Proto-Ob-Ugric?) predecessor of the Ob-Ugric words.

---

4. Also an earlier, Proto-Indo-European etymology for Fi *kutsu-* has been suggested (Koivulehto 1986: 272–274 assumes a loan from Proto-Indo-European *\*gu̯oti̯-*, reflected in Armenian *koč̣em* 'name, call somewhere'), but *Suomen vanhimman sanaston etymologinen verkkosanakirja* (s.v. *kutsua*) deems this less likely, as the Baltic etymology is phonologically plausible and more convincing in the case of a loan limited to Finnic and Saami (https://sanat.csc.fi/wiki/EVE:kutsua).

PU *kujV*, *kowja* or (?) *koja* > Hu *háj* 'fat', Fi *kuu*, MdE *kaja*, Ma *kaja*, *koja*, Ud *kęj*, *kwaj* (UEW; Sammallahti 1988: 544; Aikio 2013a: 15; Zhivlov 2014: 137; YSuS s.v. *kuu²*)

The Uralic word is reconstructed with *u* in the UEW, but later research has shown that this reconstruction has to be erroneous, even though various problems in the reconstruction of this word remain. The reconstruction of *ow* (YSuS) rather than *u* explains at least some of the reflexes more regularly (see Kallio 2018: 253). The word is not mentioned in Aikio's (2015) list of Uralic *o–a* stems, however.

For Hungarian, *o* is clearly better than *u*, as parallel examples of *o > á* abound (PU *kota* 'hut' > Hu *ház* 'house', PU *ola* 'jaw' > Hu *áll* id.). In Mari one would rather expect *u* (*ońća > užaš*, *ćoďa- > šudala-* 'course'). It is unclear, however, how *ow* would regularly develop in Mari. A possible parallel example would be Proto-Mari *åmaš* 'mosquito curtain' (Mari *omaš*, *amaš*), from Proto-Uralic *owdimi* or *awdimi* (unclear reconstruction, see Aikio 2015: 65). Also the Finnic cognate (Fi) *uudin : uutime-* shows similar vocalism as *kuu < *kowja*. On the other hand, Proto-Mari *å* often reflects Proto-Uralic *a*. This would be a possible reconstruction for Hungarian, too. The Mordvin cognate rather points to an *a–a* stem. The Permic vocalism is difficult: Udmurt *kwaj* could reflect Proto-Uralic *kowja*, cf. Udmurt *kwa-la* < Proto-Uralic *kota*, but the Komi cognate could not be derived from such a form. The vowel correspondence between the Komi and Udmurt cognates is in any case unexpected and fits any Proto-Uralic vowel combination poorly.

Both the UEW and EWUng also mention Turkic *qoyï* 'thick (flowing)' in the context of *háj*, but it is not clearly stated what kind of relationship the Turkic word should have with the Uralic etyma. On purely phonological grounds, *háj* could probably be explained as a loan from Old Turkic *qoyï* (cf. the discussion of vocalism in the context of Hungarian *jár* above), but semantically the Uralic *comparanda* denoting 'fat' are closer.

To sum up, there are various problems in the reconstruction of the Proto-Uralic word, probably because we do not know enough about the development of *Vw* sequences in the Uralic languages. But none of the languages here, with the possible exception of Finnic, point to Proto-Uralic *u*, and as we have seen, also the Finnic vowel can be explained otherwise.

PU *\*kulki-* > Hu (der.) *halad* 'proceed', Fi *kulke-* 'go, wander', SaN *golga-* 'run, float', Md E *koľge-* 'drip, flow', Ko *ķilal-* 'drift downstream', Kh V *kɔɣəl-* 'stride' (< PKh *\*kɔ̄ɣəl-*), TN *xæsy°* 'go; become' (< PSam *\*kəj-*; cognates also in Forest Nenets, Yurats, and Mator; Janhunen 1977: 51) (UEW: 197; Sammallahti 1988: 544; Aikio 2013a: 13)

This is a convincing Proto-Uralic etymology, and the reconstruction *\*kulki-* is universally accepted. *\*u* is clearly the only possibility based on comparative evidence (Finnic, Saami, Mordvin, Permic, and Samoyedic unambiguously point to a reconstruction *\*kulki-*). This is therefore a plausible example of Proto-Uralic *\*u* being reflected by Hungarian *a*, unlike many of the etymologies discussed here.

The reasons for this might be due to a conditional sound change: the possible factors could be the word-initial *\*k* or the word-internal consonant cluster. Here *\*k* has been lost through spirantization, which could have influenced the development of the vowel and caused the lowering.

PU *\*kumpa* or PU *\*kompa* 'wave' > Hu *hab* 'foam', Kh V *kŏmp* 'wave', Kh V, Vj *kump* (< PKh *\*kūmp*), Ms So *χump* id., P *kup* < PMs *\*kump*, (?) Fi *kumpu* 'hill', (?) SaL *kåbbå* 'a small hill', SaS *gabpe* 'small mountain' (Hasselbrink 1981–1985: 537) MdE *kumboldoms* 'wave, rise in waves' (cognate according to Aikio 2013a), MaE *wüt-kowǝ*, MaW *koe, ko*, Ko (?) *gibad*, TN *χampa*, Ngan *koŋhu*, Slk (Taz) *qōmpi̮* (< PSam *\*kåmpå*, cognate also in Forest Nenets and Enets; Janhunen 1977: 59) (UEW: 203–204; Sammallahti 1988: 537; Aikio 2013a: 12; 2014c: 83; 2015: 60)

The Proto-Uralic vocalism of this word is somewhat uncertain: many branches point to *\*o* rather than *u*, and it is not even clear that all the words mentioned in earlier sources as cognates really belong into the same word family. Sammallahti (1988: 537) reconstructs Proto-Uralic *\*kompa*. Such a reconstruction would account well for the Hungarian word, as *\*o > a* is a regular sound change. Proto-Samoyedic *\*kåmpå* certainly cannot reflect *\*kumpa* regularly, and also the Ob-Ugric words point to *\*o* rather than *\*u*. It is not completely certain that Finnic *kumpu* is a real cognate here, as no meaning 'wave' is attested in Finnic. Lule Saami *kåbbå* that is listed as a cognate by the UEW (203–204) is not mentioned by Aikio (2015: 60), nor is the assumed Komi cognate *gibad* that shows an aberrant *g*.

In addition to the mismatches in vocalism, there are also rather large semantic differences among the cognates: some languages denote 'wave' (Hungarian 'foam' can be derived from an earlier meaning 'wave'), others

'hill'. It is probable that the words denoting 'hill' and 'wave' are originally different stems that have been mixed up in some Uralic languages.[5]

In conclusion, it seems probable that Hungarian *hab* reflects Proto-Uralic *\*kompa* that is reflected also at least by Samoyedic and the Ob-Ugric languages. This word is not an example of Proto-Uralic *\*u* > Hungarian *a*.

**PU** *\*kunta* or *\*konti* > Hu *had* 'army', Kh V *kăntəγ* 'Khanty' (< PKh *\*kįntəγ*), Ms TJ *kānt* 'army' (< PMs *\*kānt*), (?) Fi *kunta* 'community', (?) Est *kond* id. (UEW: 206–207, 208; Sammallahti 1988: 544; Aikio 2013a: 15)

This etymology involves similar problems as *\*kompa* > *hab* mentioned above. It is possible that the traditional comparison includes more than one PU stem. The reconstruction *\*kunta* in the UEW is based on Finnic evidence, but even the Finnic cognates (Finnish *kunta* ~ Estonian *kond*) are not regular. Zhivlov (2014: 140) reconstructs *\*konti-* 'hunt, murder' and assumes that Hungarian *had* is a reflex of this Proto-Uralic stem.

Mansi *\*ā* can regularly reflect PU *\*o* in an *-i*-stem, and also the Khanty form with *\*į* can be derived from this, if it is an ablaut variant of *\*a* (Zhivlov 2014: 124). Based on Hungarian and Mansi, the Proto-Uralic form had *\*o*, and even though it seems that various details require further research, Hungarian *had* does not reflect a Proto-Uralic form that had *\*u*. All the Ugric cognates can be derived from *\*konti*.

**PU** *\*kuńci* > Hu *hangya* 'ant', Fi *kusiainen*, Ko *koʒul*, Ud *kuʒiľi*, Ms TJ *künš* id. (< (?) PMs *\*kunš-*) (UEW: 209; *kuńće*, *kuće*)

This Proto-Ugric etymology involves various problems, and the entire etymology should probably be rejected. The etymology is listed by SSA, but it is missing from Sammallahti's (1988) list of words. Hungarian *ngy* as the reflection of PU *\*ńć* is irregular, as is noted already by the UEW, and the Permic vocalism does not point to *\*u* (Proto-Permic *\*į* would be the regular outcome). It remains unclear what the exact connection between these words is, but they are certainly no regular cognates. The Finnish form *kusiainen* has probably been influenced by *kusi* 'piss, urine' due to folk etymology (SSA s.v. *kusiainen*).

---

5. Recently Zhivlov (2023: 162, 164) has also reconstructed two separate stems: Proto-Uralic *\*kumpi* 'hillock, tussock' and *\*kompa* 'wave'.

It remains a possibility that the words in various Uralic languages are loanwords from somewhere, but it is difficult to say anything more certain, as no source form is known. If Finnish *kusiainen* is unrelated, the words in Mansi, Hungarian, and Permic could reflect a loan from a substrate language in Central Eurasia, but more research would be needed before this can be proven. See Holopainen (2022: 105–107) for a discussion of other potential *Wanderwörter* with a similar distribution. Further research on the etymology of Hungarian *hangya* is certainly needed, but as the Uralic etymology has to be rejected as irregular, this is, again, not an example of a change *u > a in Hungarian.

To sum up, the Uralic etymology shows too many irregularities to be accepted as such. It is a matter of methodology whether such irregular etymologies can be accepted.

PU *kuŋi- > Hu *hó* : *hava-* 'moon', Fi *kuu*, Md E *koŋ*, Kh Kaz *χŭw*, Kam *ki* id. (UEW: 211–212; Sammallahti 1988: 537; Aikio 2013a: 13)

Hungarian *hó* 'moon' shows the oblique stem *hava-*, meaning that the word originally had *a* and the *ó* in the nominative is due to later contraction. In the various sources different Proto-Uralic reconstructions have been given, concerning both the word-internal consonant and the vocalism. Erzya dialectal *koŋ* can regularly reflect only *kuŋi-, but the rest of the forms are ambiguous. In Hungarian, *ŋ is usually reflected by *g* (the change *ŋ > *ŋk is shared with Ob-Ugric), but also many cases of *ŋ* disappearing and leaving only a hiatus filler are known. Probably there is a conditioned change that we do not understand completely. Bakró-Nagy (2003) presents a detailed account of the reflexes of *ŋ in the Ugric languages, but the exact conditions of the different reflexes remain unclear; a possible solution has been suggested by Zhivlov (2015), who assumes different developments of *ŋ in vocalic and consonantal stems in Ugric, with later analogical leveling, but the matter requires further research.

Sammallahti (1988) reconstructs Proto-Uralic *kiχi-. However, most languages seem to point to *u. Together with *kulki- > *halad* this is one of the few cases where Hungarian probably really does show *a* as a reflex of Proto-Uralic *u.

PU *kura (? *kurV) 'crooked' > ? Hu (der.) *harántos* 'slanted, skewed', Ko *ḳiṛiṣẹn*, Ud *ḳiṛiž̌*, Kh V *kŏr* 'curve in a river', TN *xəra* 'bend, curve; reason' (< PSam *kərå, cognates also in Forest Nenets, Selkup, and Kamas, Janhunen 1977: 55) (UEW: 220; Aikio 2013a)

This etymology is a complicated one. Aikio (2013a) considers Hungarian *horog* 'hook' a reflex of PU *kura* 'crooked', assuming Finnic *kura* 'left' (> Est *kura*) as cognate in addition to the Permic and Samoyedic forms listed above. Hu *horog* and *harántos* could not regularly reflect the same Proto-Uralic word due to the different vocalism. Semantically *horog* 'hook' would be an unproblematic reflex of PU *kurV* 'crooked'.

Also, *harántos* 'slanted, skew' could semantically be connected with the Uralic words denoting 'curve' or 'curved', but as *horog* shows the regular development *u > o, it is more probable that *horog* is the real reflex of Proto-Uralic *kura*. The UEW also mentions verbal forms with *hár-* occurring in Hungarian dialects, such as *hárít-* 'ablenken, abwenden'; these are semantically close to *harántos* and probably belong to the same Uralic word family.

PU *kura- > Hu (der.) *harmat* 'dew', Fi *kuura* 'hoarfrost', (?) SaL *kārrō-* 'hoarfrost forms in the trees', (?) Ko *gi̮ẹr*, (?) Ud *gẹr*, Slk *kurə* 'fine snow, hoar frost', Kam *kuro* 'frost, hoarfrost' (UEW: 215; Sammallahti 1988: 544)

Despite being included in Sammallahti's (1988) list of words, the Uralic etymology is quite problematic. Aikio (2013a) does not mention the etymology. The suggested Permic cognates are not regular: the relationship between the Komi and the Udmurt words is irregular (the Komi sequence *i̮ẹ* does not regularly correspond to Udmurt *ẹ*), making even the reconstruction of a Proto-Permic word impossible.[6] The suggested Samoyedic cognates could formally reflect a Proto-Uralic word with *u (cf. Sammallahti 1988: 495), but it is unclear whether this Selkup word really exists, as it is not found in dictionaries (such as Alatalo 2004).

The Finnic and Hungarian words could technically be derived from *kura, but the similarity might also be accidental. However, the exact composition of Hungarian *harmat* is unclear; if the word is a reflex of a stem *kura, it is uncertain what the part *-mat* represents, as the word does not look like any regular derivative. SSA (s.v. *kuura*) considers the Uralic etymology unlikely. Also, a competing Germanic etymology for the Finnic

---

6. Aikio (personal communication) notes that the Komi word has probably emerged through contraction.

word also exists: the word has been derived from Proto-Germanic *skūra* (> Middle High German *schur* 'hail(storm)'), see SSA (s.v. *kuura*). LÄGLOS (s.v. *kuura*) considers the Germanic etymology possible but uncertain; also the Uralic etymology is mentioned in LÄGLOS, but it is noted that the vocalism is irregular. In my view, the Germanic etymology is clearly a better explanation for the origin of the Finnic word. It remains unclear whether the suggested Lule Saami cognate could also be borrowed from the same Germanic word. The reconstruction of a Proto-Uralic *kura* is in any case doubtful, and it is most unlikely that Hungarian *harmat* derives from any Proto-Uralic word that had *u*.

UEW also lists similar forms in the Turkic, Tungusic, and Mongolic languages (such as Turkic *qïraɣu* 'hoarfrost', Tungusic *kiraha-* 'fall (of fine snow)', Mongolic *kīraɣu* 'hoarfrost'). It cannot be ruled out that some of the Uralic forms could be explained as loans from Turkic, but it is notable that a possible Turkic origin of *harmat* is not mentioned by WOT. As Hungarian *harmat* means 'dew', it is not semantically very close to these Altaic words.

PU *kuri-* > Hu *harag* 'anger', Kh Vj *korəm-* (< PKh *karəm-*), Ms K *χor-* (< PMs *kɔr*) (Zhivlov 2006: 117), MdE *kor* 'anger' (UEW: 220–221)

Here Mordvin *kor* could regularly reflect *u*, although also other pre-forms for Mordvin *o* are possible. The Ob-Ugric cognates show the same vocalism as the reflexes of PU *puna-* 'braid' (> Mansi *pɔn*, Khanty *panəl-*, Zhivlov 2006: 117), so it seems possible that the Ob-Ugric cognates reflect a Proto-Uralic *kuri-* or *kura-*. There are eight etymologies displaying this Ob-Ugric vowel correspondence in Zhivlov's material, which is a notable number of etymologies considering the generally small number of Proto-Uralic stems that can be reconstructed. On the other hand, most of the Uralic *u*-words in Aikio's (2013a) account of Uralic vocalism do not display this vowel correspondence in Ob-Ugric, and Zhivlov (2014: 121) has noted that the development of Proto-Uralic *u* in Khanty is not fully understood and requires further research. Because of this, it seems uncertain whether a Proto-Uralic form *kuri-* can indeed be reconstructed. A more convincing option is presented by Aikio (in preparation) who considers Hungarian *harag*, Mordvin *kor*, and the Ob-Ugric words reflexes of Proto-Uralic *kira-*; this cognate set also includes Finnish *kiro* 'curse' and North Saami *garru* id. Other examples of disharmonic *i–a* stems reflected by Hungarian *a* have been suggested, such as Proto-Uralic *wiča-* > Hungarian *vásik* 'wears away' already by Sammallahti (1988: 551), so the change can be considered regular.

PU *kuttV* > Hu *hát* 'back', Kh J *juw kutsa* 'under the tree, in the shelter of the tree', *kutəʌ* : *juw kutəʌnə* 'in the shelter of the tree', Ms KU *χūtəj* 'in the shade' (< PMs (?) *kūtəγ*), Slk (Taz) *qottä*, *qott* 'backwards' (UEW: 225; Alatalo 2004, No. 1934)

This etymology limited to Ugric and Samoyedic is mentioned as a plausible Proto-Uralic etymology, but it is missing from more recent word lists of Sammallahti (1988) and Aikio (2013a) and the scarce attestation in Ob-Ugric and Samoyedic raises suspicion. This is also one of the very few suggested examples of Hungarian *á* reflecting Proto-Uralic *u*.

Abondolo (1996) mentions the Ugric cognates but omits the Selkup cognate without comment. Abondolo reconstructs *ïï*, following his own ideas of Proto-Uralic tongue-height and length. A central vowel *ẹ* is unlikely, as it would not yield *ū* in Ob-Ugric. In his unfinished *Marginalia ad UEW*, Helimski (manuscript) mentions the Proto-Uralic etymology but reconstructs PU *kottȝ*, arguing that North Selkup *o* cannot reflect *u* but the Ugric allows either *o* or *u*. However, due to semantics, word-class differences (only an adverb in Samoyedic), and the limited distribution of the word within Uralic, Helimski does not consider the etymology completely certain.

Although Helimski rightly refutes the reconstruction with Proto-Uralic *u*, his arguments about the vocalism of this word are not entirely convincing: Hungarian *á* can reflect either *a* or *o*, and also Mansi *ū* can point to both an *a–a* and *o–a* stem. *u* is out of the question here. East Khanty *u* can reflect Proto-Khanty *ū*, which would not fit any of the possible vowels mentioned here: in *a–a* stems Proto-Khanty *ū* appears regularly after a labial or word-initially (Zhivlov 2014: 117). However, Aikio (personal communication) points out that East Khanty *u* could also reflect Proto-Khanty *ō*, which is also the middle ablaut grade of Proto-Khanty *ā*. The vowel correspondence Proto-Mansi *ū* ~ Proto-Khanty *ā* could reflect an older *-a*-stem. It seems possible, then, that the Hungarian, Mansi, and Khanty cognates could be explained from Proto-Ugric *katta*.

The connection with the Selkup word remains uncertain, and the similarity might also be accidental. The Uralic reconstruction *kuttV* should in any case be abandoned.

PU (?) *kuttV* > Hu *hat* 'six', Fi *kuusi*, SaN *gutta*, MdE *koto*, Ma *kut*, Ko *kvať*, Ud *kwať*, Kh V *kut*, Ms TJ *kat* id. (UEW: 225; Sammallahti 1988: 544)

The exact reconstruction of this Proto-Uralic numeral is disputed. The Hungarian can point to *\*u* or *\*o*, the Saami cognate is irregular from *\*kuttV*. Similar problems are involved in the reconstruction of many Uralic numerals (such as *\*kulmi* ~ *\*kolmi* ~ *\*kormi*, see Abondolo 1996: 94), meaning that the word has only limited value in the discussion of Hungarian historical vocalism. However, if we assume that *\*u* > *a* is regular in this environment (after *\*k*), there are no problems in deriving Hu *hat* from *\*kuttV*.

The Ob-Ugric vocalism is likewise problematic: Zhivlov (2006: 140) reconstructs Proto-Ob-Ugric *\*kātu*, PMs *\*kāt*, and PKh *\*kōt*. This is not a regular correspondence of any PU back vowel in Ob-Ugric, and also Abondolo (1996: 95) notes that the correspondence is unusual.

It is dubious whether the problems with the vocalism of this Proto-Uralic numeral can be solved, but as several branches of Uralic show contradictory vocalism, this etymology cannot be used as evidence of a change Proto-Uralic *\*u* > Hungarian *a*.

PU *\*kuwli-* or *\*kowli-* > Hu *hall* 'hear', Fi *kuule-*, SaN *gulla-*, MdE, M *kuľe-*, Ma *kola-*, Ud, Ko *k̮l̥-*, Kh V *kɔl-*;
PU *\*kunta-li-* > Hu *hall*, OHu *hadl*, Ms So *χūntl-* (< PMs *\*kʷāntəl-*), Kh V *kunγəl-* (< PKh *\*kuntəγl-*), Fi *kuuntele-* 'listen'
(UEW: 196–197; Sammallahti 1988: 544; Aikio 2006: 17; 2013a: 15; YSuS)

The UEW assumes that Hungarian *hall* represents contamination between two Uralic verbs, *\*kunti-li* and *\*kuli-*. This makes the analysis of this etymology challenging. Both verbs can clearly be reconstructed for Proto-Uralic, but the reconstruction *\*kuli-* is now outdated. Finnic long *uu* points to an earlier *Vw* sequence that could be reconstructed as either *\*uw* or *\*ow* (cf. PU *\*kowsi* > Fi *kuusi*, Aikio 2012: 242), but the Permic vocalism more clearly points to *\*u(w)*. The Saami vocalism (PSa *\*u*) also points to *\*u*, cf. PU *\*suxi-* > *suhka-* 'row'.

Abondolo (1996: 95) reconstructs the pre-form of Hungarian *hall/hadl* as *\*kanta-li-*, and he assumes that the Finnic high vowel *uu* is due to an ablaut variant. This explanation cannot be correct, but the pre-form *\*kanta-li-* would indeed be more probable for the Hungarian word. The Finnic vocalism might be explained through contamination with the unrelated but semantically close word family *kuule-* (cf. *kuulella*, SSA).

Due to the cluster *dl*, Old Hungarian *hadl* is clearly a reflex of Proto-Uralic *\*kVntili-*, and it is quite difficult to say for certain whether the reflexes of the two Uralic verbs have merged in the history of Hungarian.

**PU** *\*mu-* (?) 'this, that; another (?)' > Hu *más* 'other', *másik* 'another', *ma* 'today', *majd* 'soon', *most* 'now', Mari *molə̑* 'other', Fi *muu* id., SaS *mubpie* 'other; second', Ud *mịd*, Ko *męd* 'another', Ms TJ *mēt* 'another, second' (UEW: 281–282)

Kulonen (1993: 197–199) assumes that two pronominal stems, *\*mo-* and *\*mu-*, can be reconstructed for Proto-Uralic, as the vocalism in several branches of the family points to two distinct stems (for example, the Saami forms like *mubpie* as well as Udmurt *mịd* and Komi *męd* point to a stem *\*mu-*). Also, in Hungarian, there are forms like *most* 'now' pointing to *\*u* in addition to *más* which points instead to *\*o*. Hungarian *ma* probably also points to Proto-Uralic *\*mo-*. Finnic *muu* can reflect different Pre-Finnic forms (cf. the discussion of *fa* and *puu* under PU *\*pawi* below).

Helimski (1997: 301) suggests that Hungarian *ma* is possibly a cognate to Proto-Samoyedic *\*mä* 'today', retained only in Mator *mā* 'today' and these words would reflect the same Proto-Uralic lexeme. This is an interesting point that warrants further research, but it is not immediately clear how the Uralic word should be reconstructed (*\*mawi* would probably yield both Hungarian *ma* and Proto-Samoyedic *\*mä* regularly). The limited distribution of the word is also suspicious.

**PU** *\*muča-* 'illness' > Hu *hagy-máz* 'typhoid fever', Ko *mịž*, Ud *mịž* 'illness', Kh V *mɔč* 'Schaden', Ms *maš* 'hole' (UEW: 283, Aikio 2002: 13–15; 2013a)

It is assumed in the UEW that Hungarian *hagy-máz* reflects an opaque compound consisting of two words of Uralic origin. The idea of a compound as such is plausible, and the part *hagy-* has a convincing Uralic etymology (see Aikio 2002: 13–15; 2015: 60), but the issue with *\*muča-* is more problematic. A Hungarian sibilant *z* from *\*č* is completely irregular, and there are no convincing parallels for *á* as the reflex of PU *\*u* (cf. the discussion of *hát* above). This makes the etymology very dubious.

Mari *mə̂ž*, *muž* is mentioned as a cognate by the UEW, but the Mari word is not listed by Aikio (2013a). Problems with the Mari etymology have been noted also by Bereczki (2013: 153–154), who writes that Mari *ž* from *\*č* is irregular, but he argues that parallels exist. It remains unclear whether the Mari word could be connected here somehow, for example as a loan

from Permic. Proto-Saami *mocēs* (> South Saami *muhtsies* 'slovenly, untidy, messy') is mentioned as a new cognate by Aikio (2013a).

PU *muŋki* > Hu *mag* 'seed', *maga* 'self (reflexive pronoun)', Ma *moŋγ̂r* 'trunk, body', Ud *mugor*, *mi̯gor* 'body, form, build, shape, bodily appearance, Gestalt', Ko *miger* 'туловище, стан', TN *maŋk°* (< PSam *måŋkut* 'bosom', also in Forest Nenets and Enets, Janhunen 1977: 88–89) (UEW: 286–287; Aikio 2013a: 12)

The Proto-Uralic word is reconstructed as *moŋki* by Aikio (2013a: 12), who adds Proto-Samoyedic *måŋkut* to this cognate set. There is no evidence for a reconstruction with *u, so this word does not serve as an example of the alleged sound change in Hungarian.

As a side note it can be mentioned that Helimski (2002: 108) separates Hungarian *mag* 'seed' from *maga*, arguing that the latter is borrowed from an Alanic word that yielded Ossetic (Iron) *myg*, (Digor) *mugæ* 'sperm' (this idea was suggested already by Abaev 1965: 531). This etymology remains possible, especially in the light of semantic differences of *mag* and *maga*, although the vowel substitution in this Alanic etymology is not quite clear and involves similar problems as the Uralic etymology. (The Ossetic word possibly reflects Proto-Indo-Iranian *muka-, cognate to Latin *mūcus* 'snivel', Greek μύξα, Abaev 1958–1989 II: 137. However, this Indo-European etymology is far from clear, see Beekes 2010: 977–978; De Vaan 2010: 392 so the background of Ossetic *y/u* is not quite clear here.)

PU *muri-* > Hu *mar* 'bite, gnaw, etch; mill', Kh *mɔri̯-* 'break', Ms So *mur-* 'break', TN *mərda-* 'break through' (< PSam *mərə-*, cognates in all Samoyedic languages except Mator, Janhunen 1977: 87–88), Fi *murta-* 'break' (UEW: 288, Sammallahti 1988, Aikio 2013a: 13)

Aikio lists Finnish *murta-*, Khanty *mɔ̄rəj-*, and Proto-Samoyedic *mərə-* as reflexes of Proto-Uralic *muri-*. It is not completely clear that Hungarian *mar* indeed belongs here, especially as the semantic connection is not obvious, inasmuch as the other cognates denote breaking, whereas the primary meaning of Hungarian *mar* is 'bite'. The semantics are not an obstacle as such, but together with the phonological problem they can be considered to speak against the etymology.

An alternative etymology for Hungarian *mar* has been suggested: Katz (2003: 283–284) assumes that the Uralic words were borrowed from Indo-Iranian *marH-*, attested in Old Indo-Aryan *mari̯-*, *mr̥ṇāti* 'crushes'

(EWAia II: 321–322). *u might be a substitution for the Indo-Iranian zero-grade *r̥ attested in forms like the present mr̥ṇāti. Formally Hungarian mar could be a later Iranian loanword, cf. Ossetic (prefixed) læmaryn/læmarun 'press out, squeeze out', even though in this case, too, semantic problems remain. In any case, the possible Uralic origin of Hungarian mar is so uncertain that this etymology cannot be used to prove that Proto-Uralic *u can yield a in Hungarian.

PU *pawi > Hu fa (< fá with secondary shortening) 'tree', Fi puu, Ma pu, Ko pu, Ud pu (< PP *pŭ), TN pya id. (< PSam *pä, cognates in all Samoyedic languages, Janhunen 1977: 117) (UEW: 410–411; Sammallahti 1988: 539; Aikio 2013a: 9; Holopainen et al. 2017: 115, footnote 5; YSuS s.v. *puu)

The *u found in earlier sources like the UEW is probably reconstructed mostly based on Finnic evidence. However, the other languages do not clearly point to *u, instead *ow or *aw would probably be possible, as Finnic long uu can probably result from various Vw sequences (Aikio 2012: 241–243; see also Kallio 2018: 253). Aikio (2013a) and YSuS reconstruct *aw here, and this has been supported by Holopainen et al. (2017: 115, footnote 5). Hungarian and Samoyedic quite clearly rather point to *a, whereas Mari and Permic are ambiguous.

PU *pućirta- > Hu facsar 'squeeze', Fi puserta-, Ko piẓ̌irt-, Ud piẓ̌irt-, Kh V posər- (< PKh *pasər-) id. (UEW: 397; Aikio 2013a: 14)

Here the evidence for *u is quite overwhelming: Finnic and Permic both point clearly to *u, and also the Khanty reflex can be derived from that. The Uralic etymology is probably correct, and this is one possible case of the change *u > a indeed taking place in Hungarian. However, Hungarian cs is not the regular reflex of Proto-Uralic *ć (Sammallahti 1988: 517 mentions Hungarian csomó as the only example showing such a reflex, assuming a secondary affricate), and the inclusion of Hungarian facsar into this cognate set cannot be regarded as completely certain.

The issue is also complicated by the UEW's idea that the verb *puć3-r3- (as reconstructed by the UEW) includes the same verbal root as *puńća-, *puća- 'press, wring out', reflected by Khanty (V, N) pos-, (DN) pus-, Mansi (TJ, P, So) pos-, (KU) pas-, Komi pićkị-, Mari (W) pə̂nze-, (E) puńće-, puńćala-, Lule Saami påhtjē-, and related forms in other Saami languages (UEW: 404). The idea clearly cannot be correct as such, as the cognates allegedly reflecting *puńća-, *puća- are irregular and it is clear that they

cannot reflect the same Proto-Uralic verb (for example, the Komi and Saami forms cannot reflect a cluster *ńć, but this is required by Mari, and the Saami vocalism is also irregular). It remains dubious whether any of the forms listed under this verb in the UEW can be connected with *pućirta-.

It has been noted (EWUng: 348) that also a variant *csafar* has risen in Hungarian through metathesis, and the meaning and phonological shape of *csafar* have been influenced by the unrelated verb *csavar* 'turn (something), waggle'. The unexpected vowel and affricate in *facsar* might also have been influenced by a contamination with *csavar*, although it is admittedly difficult to prove this.

PU *puna* 'hair' > Hu *fan*, *fon* 'pubic hair', Kh V *pun* 'hair, wool, feather', Ms TJ *pon* 'feather, hair', Fi *puna* 'red', Md *pona* 'hair, wool', Ma *pân* 'hair, feather' (UEW: 407 s.v. *puna*; Sammallahti 1988; Aikio 2013a: 14)

This is a convincing Uralic etymology accepted by all the relevant sources, and it is clear that *u* has to be reconstructed for Proto-Uralic. It is unclear why both variants *fan* and *fon* are attested in Hungarian, but it can be assumed that *fon* is the regular form here while *a* reflects a dialectal development: both forms are attested already in Old Hungarian (EWUng: 354). As pointed out above in Section 3, the lack of clarity in the interpretation of Old Hungarian vocalism has been noted by Benkő (1980: 89–94), but as forms with *a* and *o* can be found in both modern Hungarian and already in Old Hungarian sources, it is probable that *fan* indeed existed in Old Hungarian. Further research on this dichotomy is needed, but it should be noted that as phonemes do not split spontaneously, it would be good to find some reason for the dichotomy *fan ~ fon*. The variant *fon* in any case shows the expected development *o < *u*.

PUg *purɜ* > Hu *far*, *farok* 'tail', Kh V *pįr* 'back part' (UEW: 407, 880)

Aikio (2018) argues that Hungarian *far*, *farok* reflect Proto-Uralic *ponči* 'tail'. This is a convincing explanation in the light of Aikio's new sound law *nč* > Hu *r*. Aikio also notes that the vowel correspondences between the Hungarian and Khanty cognates suggested in the UEW are irregular, so the UEW's reconstruction of a Proto-Ugric noun *purV* has to be rejected in any case.

PU *ruŋkV > Hu *rág* 'chew', Ms L *rågn-* id., TN *luŋkiba-* 'gnaw, nibble', Ngan *ľüŋgüśa* 'gnaw' (UEW 426)

This is an irregular and unconvincing etymology, as noted already by the UEW. None of the cognates suggested here can reflect a pre-form *ruŋkV-regularly. Mansi *g* from Proto-Uralic *ŋk* would be an irregular development that has no parallels in other etymologies. Tundra Nenets *l* cannot regularly reflect PU *r*, and it is quite probable that Proto-Uralic phonotactics did not allow words beginning with *r-* (see Hahmo 2003/2004). In Sammallahti's (1988) list of words, no Uralic cognates with word-initial *r-* are listed.

PU *śukkV ~ *śakkV 'piece, bit, part' > Hu (dial.) *szak* 'small piece' (also in compounds *észak* 'north', *éjszaka* 'night', and in the derivatives *szakad*, *szakít* 'tear'), Kh Vj *săk* 'crumbled', (?) Fi *sukku* 'state of being crushed' (UEW: 457)

This etymology offers again no real evidence for Proto-Uralic *u*, as the Uralic etymology is considered uncertain even by the UEW, and none of the languages point really to the reconstruction of Proto-Uralic *u*. The suggested Finnic cognate *sukku* is poorly attested and a semantically dubious cognate. East Khanty *ă* is not the regular reflex of Proto-Uralic *u*. The UEW is right in assuming that Hungarian *szak* and Khanty *săk* can indeed belong together, and it can be noted that they can reflect Uralic *śakki* or *śokki* (cf. Zhivlov 2014: 124).

PU *śuwi or *śawi > Hu *száj* 'mouth', Fi *suu*, SaS *tjovve* (< PSa *ćuvę* or *ćovę*) Ma *šu* (< PMa *šu*), Ud *śu* 'mouth', (?) Ko *śu-* in compounds *śu-kęś* 'kvass', *śu-kęśaś-* 'drink (verb)' (< PP *śu*) id., Ms K *sunt* 'mouth of a river', So *sūp* 'mouth' (< ? PMs *su-*) (UEW: 492–493; Aikio 2013a: 14; YSuS s.v. *suu*)

This etymology is a similar case as *pawi* 'tree' above: although the Uralic etymology as such is plausible, various details of the reconstruction are unclear. Among the Uralic cognates, only Finnic forms like Finnish *suu* point to Proto-Uralic *u*, but as it was discussed above, various Proto-Uralic sequences of *Vw* can result in Finnic *uu*. Proto-Permic *u* does not point regularly to Proto-Uralic *u*, and Mari *u* can also reflect various pre-forms, meaning that this is not a certain case of *u > a* in Hungarian. The Permic cognates are considered uncertain already by the UEW. Mansi short *u* points to Proto-Uralic *u* rather than *a* (see also the discussion of Hungarian *szád* below). Proto-Saami *ćuvę* or *ćovę* could not regularly reflect Proto-Uralic *a*.

PU *śuwinti* > (?) *szád* 'mouth of a river', Ms K *sunt*, SaKld *čont* (< PSa *čuntẹ*) id. (Aikio 2013a: 14)

It is unclear whether this Uralic word reflects the same stem as Hungarian *száj* 'mouth' and its cognates discussed above. The vocalism of the Saami cognates does not point regularly to Proto-Uralic *u*, but Aikio assumes *uw* that could have probably developed into *u* in Saami. Due to the limited attestation and unclear vowel developments in both Saami and Hungarian, this word does not give reliable proof of the sound change *u* > *a* in Hungarian.

PU *šurV* 'cut' > (?) Hu *irt-* 'destroy', (?) *arat* 'harvest', Ko *šẹr-*, Ud *šịr-*, Kh V *lŏrt-*, O *lărt-*, TN *tyarocy* 'be divided; share' (< PSam *tär-*, cognates in all Samoyedic languages, Janhunen 1977: 154–155) (UEW: 503–504; Sammallahti 1988: 550; WOT: 1232)

This is a rather problematic etymology, as both *irt* and *arat* are considered possible reflexes of the same stem *šurV* in the UEW; this cannot be correct, as it is impossible to connect these forms etymologically due to the different vocalism. The UEW (492) considers *arat* uncertain. The whole existence of a Proto-Uralic verbal stem *šurV* is based on very uncertain evidence, as at least the suggested Samoyedic cognates clearly do not regularly point to *u*. The Permic and Khanty cognates can reflect Proto-Uralic *šurV*, but neither Hungarian *arat* or *irt* reflects *šurV* regularly. Nevertheless, the UEW's explanation of the origin of Hungarian *irt* is accepted by WOT. However, *arat* is considered a Turkic loan (see below). Sammallahti (1988) also mentions the etymology, although with a question mark, reconstructing *šị/ura-* 'remove' and mentioning only Hungarian *irt* as a cognate. Sammallahti does not mention the Samoyedic forms listed in the UEW.

WOT (70–71) considers Hungarian *arat* a possible loanword from the Old Turkic verb *or-* 'mow' (> East Old Turkic *or-* 'mow (grass), reap (cereal crops)'), perhaps from its unattested causative form *or(a)t-*. This Turkic etymology is phonologically and semantically plausible. The etymology of *irt* remains open and requires further research, but due to the probable Turkic origin of *arat*, it is improbable that *irt* is related to it through a lexical split.

PU *tukti* > Hu *tat*, Kh V *tŏγət*, MaW *tâktâ*, Ko *tị̆k*, TN *tade* 'crossbar' (UEW: 534, Sammallahti 1988: 550; Aikio 2013a: 14)

The Mari, Permic, and Samoyedic cognates point to *u. Also, the Khanty form can be derived from *u regularly. The cognate set is regular, so this appears to be one of the few genuine cases of PU *u > Hu *a. It can be assumed here that the loss of *k in the cluster *kt has impacted the development of *u. Aikio noted this word as a probable example of an "irregular lowering" *u > a in Hungarian. However, Aikio did not deal with any of the examples in detail, and as loss of velar consonants (stop *k or spirant *x) in word-internal position is involved in many other etymologies showing this lowering (see the discussion of *halad*, *tó* : *tava-*), it can be argued that this change is not irregular but occurred under certain conditions.

PUg *tul-* > Hu *táltos* 'sorcerer', Kh Kaz *tɔʌt* 'help, relief (in sickness or poverty)', *tɔʌta* 'without (bigger) difficulty, without noise; suddenly', N *tolt* 'Riese (eigtl. Zauberer)', *toltn* ~ *tolten* 'with magic', Vj *tolt* 'fever' (< PKh ? *tolt-), Ms N *tūlt(en)* 'easily' (< PMs *tūlt-) (Honti 1982: 188; UEW: 895; Abondolo 1996: 44; WOT: 841–843; Honti 2017: 62–67)

The Proto-Ugric etymology in the UEW involves numerous problems and it has been doubted by Abondolo (1996: 44) and WOT (841–843). Abondolo notes that the Ugric etymology involves various problems and it is not even certain that the Khanty words grouped together in the UEW are related to each other, while Honti's Proto-Khanty reconstruction is problematic. Also, semantic problems can be added to this etymology.

WOT lists a possible Turkic etymology for *táltos*, assuming that Hungarian *táltos* < ? *taltučV is a loan from West Old Turkic *taltutči 'the one who exercises loss of consciousness' (derived from a Turkic verb *tal- 'faint'). This explanation is plausible, as Hungarian *á* ← Turkic *a is a well-attested substitution, and *s* can also be derived from earlier *č without problems. Honti (2017: 62–67) discusses both the Ugric and the Turkic etymology in detail, analyzing especially the semantic developments, and he supports the Ugric etymology. Honti does not offer any specific arguments against the etymology presented in WOT, however.

PU *tuli- > Hu *talál* 'find', Fi *tule-* 'come', Ma *tola-*, TN *tosy°* id. (UEW: 535; Aikio 2002: 29–30)

The Uralic etymology of Hungarian *talál* is considered uncertain by the UEW due to semantics, as all the other cognates point to the meaning 'come'. There is no compelling reason to assume that *talál* is from *\*tuli-*. A more convincing alternative has been suggested by Aikio (2002: 29–30), who assumes a possible connection with PU *\*tolwa-* or *\*talwa-* 'bring' (> SaN *doalvut*, Nganasan *tojbu-* 'take, transport, deliver', Kam *tu-* 'arrive; reach' < PSam *\*tajwa-*). Hu *a* from PU *\*a* (or *\*o*) would be a regular development (Zhivlov 2014: 124). However, also in this case the semantic connection of the Hungarian word with the rest of the cognates is not quite transparent, and the etymology remains somewhat uncertain. In later works, Aikio (2013a; 2015) does not list the Hungarian word among the reflexes of PU *\*talwa-* or *\*tolwa-*.

PU *tuna- > Hu *tan-*, MdE *tonado-* < PMd *\*tonadə-*, Ma *tunema-* < PMa *\*tŭnemä-* 'get used to, learn', Ko *tunal-*, Ud *tunal-* < PP *\*tŭn-* 'seer, soothsayer', *\*tŭnal-* 'foretell'; TN *tənarasy* 'train, teach' < PSam *\*tə̂nå-* 'teach, train', cognates also in Enets, Forest Nenets, and Selkup, Janhunen 1977: 147 (< Pre-Samoyedic *\*tun-ta-*) (UEW: 537; Sammallahti 1988: 550; Aikio 2013a: 14; in preparation)

The Proto-Uralic verbal stem *\*tuna-* is attested only in derivatives in Hungarian, Mordvin, Mari, and Samoyedic. In modern Hungarian, a word *tan* is attested, but this is a modern back-formation from the verbs *tanul*, *tanit* (EWUng: 1477–1478). It is unclear whether Komi *tun* reflects the underived stem, but most Permic reflexes are clearly derivatives. Sammallahti gives the Uralic reconstruction as *\*toni-*, but Aikio (2002: 44–45) reconstructs *\*u*, noting that the Hungarian cognate is "apparently irregular". Concerning the stem vocalism, the UEW reconstructed *\*tuna-*, and although Aikio (2013a) reconstructed *\*tuni-*, he has more recently (in preparation) convincingly argued that the word was an *-a*-stem *\*tuna-* (both Mordvin and Samoyedic point to an *\*-a*-stem).

Proto-Uralic *\*o* would be a more regular predecessor of Hu *a*, and the Permic cognates can point to *\*o* as well, so they are ambiguous in this sense. However, the Mordvin cognate points to earlier *\*u*, and also the Mari word can be regularly derived from *\*u*. Aikio's arguments are convincing, and the reconstruction *\*tuni-* can regularly explain most of the cognates.

The Hungarian word is the most problematic one in this cognate set, as the vowel *a* does not fit any of the rules described above. However, the vowel *a* might be explained through contamination with the unrelated word *tanú* 'witness' which is a loan from West Old Turkic *\*tanug*, cf. East Old Turkic *tanug* 'witness', a derivative of the verb *tanï-* (WOT: 848–852). In earlier etymological literature, it was occasionally assumed that *tanú* and *tan-* are etymologically related, and even though this is not the case, it is easy to assume that a connection has been made between the two similar verbs by Hungarian speakers through folk etymology. It is also possible that instead of the noun *tanú*, the speakers of Proto-Hungarian borrowed the Turkic verb *\*tanï-*, and the native *\*tuni-* has merged with the borrowed, phonologically similar verb. This kind of situation is difficult to prove, but a parallel case is Finnish *ahta-* (< Proto-Finnic *\*akta-*), which is a Uralic verb semantically influenced by a Baltic loan.

It has been already suggested by Ikeda (2000: 66) that the Hungarian verb was semantically influenced by Turkic *\*tanï-*. Ikeda does not comment on the phonological developments, however.

To sum up, Hungarian *tan* can be included among the cognates of Proto-Uralic *\*tuni-*, so Aikio's statement that this is an irregular reflex of *\*tuni-* is plausible as such. However, it is probable that the *a* vocalism is the result of influence from an unrelated Turkic word.

PU *\*turV* > Hu *tar* 'withers', *tarja* 'cow's spine with flesh', Kh V *tur* 'neck', Ms TJ *tor* id., ? Fi *turja* 'back of the neck' (UEW: 538)

Hu *torok* 'throat' is probably the real, regular cognate here (Aikio 2013a: 15). The relationship to *tar, tarja* is uncertain. The two Uralic stems *\*turV* and *\*tura* listed in the UEW probably belong together somehow. The issue is not quite clear, but *torok* in any case shows the expected reflex of Proto-Uralic *\*u*. Further research into the etymology of *tar* and *tarja* would be needed, and it is possible that these forms are unrelated to Proto-Uralic *\*turV* or that they show later dialectal developments.

PU *tuxi- 'lake' > Hu *tó : tava-* 'lake', Kh V *tŏγ*, Ms TJ *tō*, Ko *ti̯*, Ud *ti̯*, TN *to* id. (UEW: 532, Sammallahti 1988: 540, Aikio 2013a: 14)

This word displays a regular morphophonemic alternation in modern Hungarian, where Hu *ó* is due to contraction of the sequence *aw* (retained in the oblique stem *tava-*). Most languages (including Permic) point clearly to PU *u rather than *o as reconstructed by the UEW. This is thus probably another case of *u > a in Hungarian.

PU *u- ~ *o- 'that' > Hu *az* 'that', ? MdE *ombo* 'another', Ud *oti̯n* 'there' (UEW: 332)

The UEW gives parallel reconstructions, with *u- as one option. This pronominal stem is usually reconstructed as *o-, see recently Janhunen (2020: 132), who assumes that the pronominal *o- found in Hungarian *az* is connected to the Proto-Uralic copula *o- (> Proto-Finnic *oma 'is', *omat 'they are' > Veps *om*, *omad* id.). This does not fit the Mordvin evidence very well, as *o would not yield Mordvin o, but Hungarian a- can be regularly derived from Proto-Uralic *o. The vocalism of Proto-Uralic pronominal stems is in general very complicated to reconstruct (see also the discussion on Hungarian *más* above), but there is no compelling reason to reconstruct *u here.

PUg *urV ~ *arV > Hu *aránt* 'against', *iránt*, *ëránt* 'into direction', Kh V *ur*, Ms TJ *or* 'mountain ridge' (UEW: 833–834; EWUng: 622)

The UEW presents a Ugric reconstruction with alternative vocalism. In EWUng, it is stated that *ur₃ is the likely reconstruction, and no reconstruction with *a- is mentioned. The assumed Khanty and Mansi cognates are grouped under a different PU stem, namely *wara 'mountain, hill' (> SaN *várri*, Nganasan *béru* 'mountain, cliff', Kam *bōr* 'mountain, ridge') by Aikio (2012: 233) and Zhivlov (2014: 120): this explanation is phonologically regular, and there is no reason to reconstruct a separate Ugric stem to account for the Ob-Ugric forms. Whatever the etymology of Hu *aránt* is, it cannot reflect PU *wara, as the word-initial *w- should have been retained. The relationship between *aránt* and *iránt* requires further research, but neither of these words can be derived from a reconstruction *urV. A full account of the etymology of this word family would require a careful philological discussion of the Old Hungarian data, but as the alleged cognates listed in the UEW have been shown to be unrelated to these Hungarian words, this word does not belong in discussions of Proto-Uralic *u.

## 6. Conclusions

Based on the analysis above, a significant part of the etymologies manifesting the alleged change of *u to a or á in Hungarian turned out to be wrong on some level: many etymologies were shown to be implausible, whereas some cases of this sound change were based on reconstructions that turned out to be wrong, even if the etymologies themselves are correct. The results show that there is little reason to assume a sporadic change *u > a or *u > á in the history of Hungarian.

However, some plausible examples displaying this change remain, and it can be argued that *u > a (but not *u > á) indeed took place in the history of Hungarian under some conditions. The convincing Uralic etymologies that clearly show this change are the following: *kulki- > halad, (?) *kuwli- > hall, *kuŋi/kuwi > hó : hava-, *tukti > tat, *tuxi > tava-. In addition to these, the etymologies of facsar and fan have a Uralic background that was considered as plausible or probable in the discussion of etymologies. It is possible that some words that show disputed vocalism also reflect *u, but at the present state of research this cannot be shown and further research is needed before the issue can be settled. The change *u > a is reflected in a very small group of etymologies, and it is dubious whether far-reaching conclusions on historical vocalism can be drawn based on them.

However, most of these words involve the loss of the velar stop *k or the velar spirant *x in word-internal position: *k is lost in *kulki- and *tukti-, and *x in *tuxi-. A possible explanation to account for this change is that the loss of *k and *x caused the lowering of the preceding vowel *u that then merged with o that regularly developed into *a in *-i-stems. For hall (< Old Hungarian hadl) and *kuŋi/kuwi- a similar explanation does not hold as such, but as hall can be assumed to reflect contamination of Proto-Uralic *kuwli- and *kantili-, the vowel a can be explained as a regular reflex of the *a of the latter Uralic verb. If hó : hava- goes back to *kuŋi, it can be assumed that *ŋ first became *x and was lost after that, causing the lowering as happened in tó : tava- < *tuxi-.

The rules presented above do not explain all the possible examples. However, the discussion has shown that a significant majority of the examples can be explained otherwise, and it can be claimed that the fact that most of the etymologies allegedly manifesting this change can be rejected shows that the methodological principle of regular sound change can lead to a clearer picture of Uralic and Hungarian historical phonology.

Ante Aikio (personal communication) notes that in the case of *turja* >
Hu *torok*, *tarja* and *puna* > Hu *fan*, *fon*, the Ob-Ugric reflexes show sim-
ilar correspondences, and it is possible that a different vowel combination
should be reconstructed in such cases.

It is certainly possible that further research will find additional exam-
ples of words that fit the cautious conclusions presented above. It is also
possible that some further convincing examples of Proto-Uralic *u* > Hun-
garian *a* will be presented, and the conditions for this development will
become more apparent. It is in any case clear that there is much to do
concerning the reflexes of *u* in Hungarian and the Ob-Ugric languages.

## Abbreviations

| | | | | | |
|---|---|---|---|---|---|
| Est | Estonian | | Ms | Mansi | |
| Fi | Finnish | | | K | East Mansi (Konda) |
| Hu | Hungarian | | | KU | East Mansi (Lower Konda) |
| Kam | Kam | | | | |
| Kh | Khanty | | | L | West Mansi (Lozva) |
| | DN | South Khanty (Demjanka, Narygin) | | N | North Mansi |
| | | | | P | West Mansi (Pelym) |
| | J | East Khanty (Jugan) | | So | North Mansi (Sosva) |
| | Kaz | North Khanty (Kazym) | | TJ | South Mansi (Janychkova) |
| | N | North Khanty | | | |
| | O | North Khanty (Obdorsk) | Ngan | Nganasan | |
| | | | OHu | Old Hungarian | |
| | Trj | East Khanty (Tremjugan) | PFU | Proto-Finno-Ugric | |
| | | | PKh | Proto-Khanty | |
| | V | East Khanty (Vakh) | PMa | Proto-Mari | |
| | Vj | East Khanty (Vasjugan) | PMd | Proto-Mordvin | |
| Ko | Komi | | PMs | Proto-Mansi | |
| Ma | Mari | | PP | Proto-Permic | |
| | E | East | PSa | Proto-Saami | |
| | W | West | PSam | Proto-Samoyedic | |
| Md | Mordvin | | PU | Proto-Uralic | |
| | E | Erzya | PUg | Proto-Ugric | |
| | M | Moksha | Sa | Saami | |
| | | | | Kld | Kildin Saami |
| | | | | L | Lule Saami |
| | | | | N | North Saami |
| | | | | S | South Saami |
| | | | Slk | Selkup | |
| | | | TN | Tundra Nenets | |
| | | | Ud | Udmurt | |

## References

Abaev 1958–1989 = Абаев, Василий И. 1958–1989. *Историко-этимологический словарь осетинского языка*, I–V. Москва & Ленинград: Академия Наук СССР.

Abaev 1965 = Абаев, Василий И. 1965. К алано-венгерским лексическим связям. In Ortutay, Gyula & Bodrogi, Tibor (eds.), *Europa et Hungaria: Congressus Ethnographicus in Hungaria*, 517–537. Budapest: Akadémiai Kiadó.

Abondolo, Daniel. 1996. *Vowel rotation in Uralic: Obug[r]ocentric evidence* (SSEES Occasional Papers 31). London: School of Slavonic and East European Studies, University of London.

Aikio, Ante. 2002. New and old Samoyed etymologies. *Finnisch-Ugrische Forschungen* 57. 9–57.

Aikio, Ante. 2006. New and old Samoyed etymologies (Part 2). *Finnisch-Ugrische Forschungen* 59. 9–34.

Aikio, Ante (Luobbal Sámmol Sámmol Ánte). 2012. On Finnic long vowels, Samoyed vowel sequences, and Proto-Uralic *x. In Hyytiäinen, Tiina & Jalava, Lotta & Saarikivi, Janne & Sandman, Erika (eds.), *Per Urales ad Orientem: Iter polyphonicum multilingue, Festskrift tillägnad Juha Janhunen på hans sextioårsdag den 12 februari 2012* (Mémoires de la Société Finno-Ougrienne 264), 227–250. Helsinki: Société Finno-Ougrienne.

Aikio, Ante (Luobbal Sámmol Sámmol Ánte). 2013a. Uralilaisen kantakielen vokaalistosta. (Handout of a paper presented at Etymologia ja kielihistoria: Erkki Itkosen ja Aulis J. Joen 100-vuotisjuhlaseminaari 19.4.2013, Helsinki.)

Aikio, Ante (Luobbal Sámmol Sámmol Ánte). 2013b. Studies in Uralic etymology I: Saami etymologies. *Linguistica Uralica* XLIX. 161–174.

Aikio, Ante (Luobbal Sámmol Sámmol Ánte). 2014a. On the reconstruction of Proto-Mari vocalism. *Journal of Language Relationship* 11. 125–157.

Aikio, Ante (Luobbal Sámmol Sámmol Ánte). 2014b: Studies in Uralic etymology II: Finnic etymologies. *Linguistica Uralica* L. 1–19.

Aikio, Ante (Luobbal Sámmol Sámmol Ánte). 2014c: Studies in Uralic etymology III: Mari etymologies. *Linguistica Uralica* L. 81–93.

Aikio, Ante (Luobbal Sámmol Sámmol Ánte). 2015. The Finnic 'secondary *e*-stems' and Proto-Uralic vocalism. *Journal de la Société Finno-Ougrienne* 95. 26–66.

Aikio, Ante (Luobbal Sámmol Sámmol Ánte). 2018. Notes on the development of some consonant clusters in Hungarian. In Holopainen, Sampsa & Saarikivi, Janne (eds.), *Perì orthótētos etýmōn (Περὶ ὀρθότητος ἐτύμων): Uusiutuva uralilainen etymologia* (Uralica Helsingiensia 11), 77–90. Helsinki: Société Finno-Ougrienne.

Aikio, Ante (Luobbal Sámmol Sámmol Ánte). 2022. Proto-Uralic. In Bakró-Nagy, Marianne & Laakso, Johanna & Skribnik, Elena (eds.), *The Oxford guide to the Uralic languages*, 3–27. Oxford: Oxford University Press. https://doi.org/10.1093/oso/9780198767664.003.0001

Aikio, Ante (Luobbal Sámmol Sámmol Ánte). (In preparation.) *Uralic etymological dictionary*. Unpublished manuscript.

Alatalo, Jarmo (ed.). 2004. *Sölkupisches Wörterbuch aus Aufzeichnungen von Kai Donner, U. T. Sirelius und Jarmo Alatalo* (Lexica Societatis Fenno-Ugricae XXX). Helsinki: Société Finno-Ougrienne.

Anttila, Raimo. 1989. *Historical and comparative linguistics* (Current Issues in Linguistic Theory 6). Amsterdam & Philadelphia: John Benjamins.

Bakró-Nagy, Marianne. 2003. Egy inetimologikus hangról. In Bakró-Nagy, Marianne & Rédei, Károly (eds.), *Ünnepi kötet Honti László tiszteletére*, 27–45. Budapest: MTA.

Bakró-Nagy, Marianne. 2013. Mit tudunk az ugor történeti fonológiáról? A nyugati ótörök és magyar kapcsolatok hozadéka. In Agyagási, Klára & Hegedűs, Attila & É. Kiss, Katalin (eds.), *Nyelvelmélet és kontaktológia 2*, 173–189. Piliscsaba: PPKE BTK.

Bárczi, Géza. 1958. *Magyar hangtörténet*. Budapest: Tankönyvkiadó.

Beekes, Robert Stephen Paul. 2010. *Etymological dictionary of Greek*, I–II (Leiden Indo-European Etymological Dictionary Series, vol. 10). Leiden & Boston: Brill.

Benkő, Loránd. 1980. *Az Árpád-kor magyar nyelvű szövegemlékei*. Budapest: Akadémiai Kiadó.

Bereczki, Gábor. 2013. *Etymologisches Wörterbuch des Tscheremissischen (Mari): Der einheimische Wortschatz*. Nach dem Tode des Verfassers herausgegeben von Klára Agyagási und Eberhart Winkler (Veröffentlichungen des Societas Uralo-Altaica 86). Wiesbaden: Harrassowitz Verlag.

Campbell, Lyle. 1996. On sound change and challenges to regularity. In Durie, Mark & Ross, Malcolm (eds.), *The comparative method reviewed: Regularity and irregularity in sound change*, 72–89. New York: Oxford University Press.

Campbell, Lyle. 1999. *Historical linguistics: An introduction*. First MIT Press edition. Cambridge, Massachusetts: The MIT Press.

Clackson, James. 2007. *Indo-European linguistics*. Cambridge: Cambridge University Press.

Csúcs, Sándor. 2005. *Die Rekonstruktion der permischen Grundsprache*. Budapest: Akadémiai Kiadó.

Csúcs, Sándor & Honti, László & Salánki, Zsuzsanna & Varga, Judit. 1991. *Statistik der uralischen Lautentsprechungen* (Lingustica Series B Documenta, 1). Budapest: A Magyar Tudományos Akadémia Nyelvtudományi Intézete.

De Vaan, Michiel. 2008. Book review: James Clackson, *Indo-European linguistics: An introduction*, 2007. *Lingua* 118. 1228–1232.

E. Abaffy, Erzsébet. 2003. Az ősmagyar kor: Hangtörténet. In Kiss, Jenő & Pusztai, Ferenc (eds.), *Magyar nyelvtörténet*, 106–128. Budapest: Osiris Kiadó.

EWAia = Mayrhofer, Manfred. 1986–2001. *Etymologisches Wörterbuch des Altindoarischen* I–III. Heidelberg: Winter.

EWUng = Benkő, Loránd (ed.). 1993–1997. *Etymologisches Wörterbuch des Ungarischen*. Budapest: Akadémiai Kiadó.

Fox, Anthony. 1995. *Linguistic reconstruction: An introduction to theory and method*. Oxford & New York: Oxford University Press.

Gerstner, Károly. 2018. Hangtörténet. In Kiss, Jenő & Pusztai, Ferenc (eds.), *A magyar nyelvtörténet kézikönyve*, 103–128. Budapest: Tinta Könyvkiadó.

Hahmo, Sirkka-Liisa. 2003/2004. Zur finnisch-ugrischen Phonotaktik: Gab es Wörter mit inlautendem *r-* in der finnisch-ugrischen Grundsprache? *Ural-Altaische Jahrbücher, Neue Folge* 18. 1–23.

Häkkinen, Jaakko. 2007. *Kantauralin murteutuminen vokaalivastaavuuksien valossa*. Helsinki: University of Helsinki. (Master's thesis.)

Hasselbrink, Gustav. 1981–1985. *Oårj'elsaamien baaguog'ärjaa = Südlappisches Wörterbuch 1–3* (Skrifter utgivna genom Dialekt- och folkminnesarkivet i Uppsala, Ser. C. 4). Uppsala: Dialekt- och folkminnesarkivet.

Helimski, Eugene. 1997. *Die matorische Sprache: Wörteverzeichnis, Grundzüge der Grammatik, Sprachgeschichte*. Unter Mitarbeit von Beáta Nagy (Studia Uralo-Altaica 41). Szeged: JATE Finnugor Tanszék.

Helimski, Eugene. 1999. Basic vocabulary in PU and PFU: Remarks on etymology and reconstruction. (A handout delivered at the Budapest Uralic Workshop 2: Etymology, September 1999, organized by the Nyelvtudományi Intézet, Budapest.)

Helimski, Eugene. 2002. Zu den iranischen Lehnwörtern im Ungarischen. In Blokland, Rogier & Hasselblatt, Cornelius (eds.), *Finno-Ugrians and Indo-Europeans: Linguistic and literary contacts, Proceedings of the Symposium at the University of Groningen, November 22-24, 2001* (Studia Fenno-Ugrica Groningana 2), 105–111. Maastricht: Shaker.

Helimski, Eugene. Manuscript. *Marginalia ad UEW*. (Unpublished manuscript.)

Holopainen, Sampsa. 2022. Uralilaisen lingvistisen paleontologian ongelmia: Mitä sanasto voi kertoa kulttuurista? In Kaheinen, Kaisla & Leisiö, Larisa & Erkkilä, Riku & Qiu, Toivo (eds.), *Hämeenmaalta Jamalille: Kirja Tapani Salmiselle 07.04.2022*, 101–114. Helsinki: University of Helsinki.

Holopainen, Sampsa & Junttila, Santeri & Kuokkala, Juha. 2017. Indoeurópai jövevényszavak és a második szótagi labiális magánhangzók fejlődése az uráli nyelvekben. In Forgács, Tamás & Németh, Miklós & Sinkovics, Balázs (eds.), *Nyelvtudományi kutatások ujabb eredményei IX*, 109–136. Szeged: Szegedi Tudományegyetem.

Honti, László. 1982. *Geschichte des obugrischen Vokalismus der ersten Silbe*. Budapest: Akademiai Kiadó.

Honti, László. 2013. Comments on Uralic historical phonology. *Acta Linguistica Hungarica* 60. 1–68.

Honti, László. 2017. *A magyar és a nyugati ótörök szókészleti kapcsolatairól*. Budapest: Tinta Könyvkiadó.

Ikeda, Tetsuro. 2000. On some problems of the Pre-Settlement of Pre-honfoglalás Hungarian language. *Uralica* 12. 59–69.

Janhunen, Juha. 1977. *Samojedischer Wortschatz: Gemeinsamojedische Etymologien* (Castrenianumin toimitteita 17). Helsinki: Société Finno-Ougrienne.

Janhunen, Juha. 1981. Uralilaisen kantakielen sanastosta. *Journal de la Société Finno-Ougrienne* 77. 219–274.

Janhunen, Juha. 2020. Issues of comparative Uralic and Altaic studies (6): Uralic copulas and their analogues in other Eurasian languages. In Holopainen, Sampsa & Kuokkala, Juha & Saarikivi, Janne & Virtanen, Susanna (eds.), *Ёмас сымыӊ нэ̄кве во̄ртур э̄тпост самын патум: Scripta miscellanea in honorem Ulla-Maija Forsberg* (Mémoires de la Sociéte Finno-Ougrienne 275), 129–140. Helsinki: Société Finno-Ougrienne.

Kallio, Petri. 2012. The non-initial-syllable vowel reductions from Proto-Uralic to Proto-Finnic. In Hyytiäinen, Tiina & Jalava, Lotta & Saarikivi, Janne & Sandman, Erika (eds.), *Per Urales ad Orientem: Iter polyphonicum multilingue, Festskrift tillägnad Juha Janhunen på hans sextioårsdag den 12 februari 2012* (Mémoires de la Sociéte Finno-Ougrienne 264), 163–175. Helsinki: Société Finno-Ougrienne.

Kallio, Petri. 2018. Ensitavun diftongit kantasuomessa. In Holopainen, Sampsa & Saarikivi, Janne (eds.), *Perì orthótētos etýmōn* (*Περὶ ὀρθότητος ἐτύμων*): *Uusiutuva uralilainen etymologia* (Uralica Helsingiensia 11), 251–268. Helsinki: Société Finno-Ougrienne.

Katz, Hartmut. 2003. *Studien zu den älteren indoiranischen Lehnwörtern in den uralischen Sprachen*. Aus dem Nachlass herausgegeben von Paul Widmer, Anna Widmer und Gerson Klumpp (Indogermanische Bibliothek, 3. Reihe). Heidelberg: Winter.

Kiparsky, Paul. 2015. New perspectives in historical linguistics. In Bowern, Claire & Evans, Bethyn (eds.), *The Routledge handbook of historical linguistics*, 64–102. London & New York: Routledge.

Koivulehto, Jorma. 1986. Die Sieverssche Regel im Lichte der germanisch-finnischen Lehnbeziehungen. In Brogyanyi, Bela & Krömmelbein, Thomas (eds.), *Germanic dialects: Linguistic and philological investigations* (Current Issues in Linguistic Theory 38), 249–294. Amsterdam: John Benjamins.

Korompay, Klára. 2018. Helyesírás-történet. In Kiss, Jenő & Pusztai, Ferenc (eds.), *A magyar nyelvtörténet kézikönyve*, 83–98. Budapest: Tinta Könyvkiadó.

Kulonen, Ulla-Maija. 1993. Suomalais-ugrilaisista *mo-* ja *mu-*pronominivartaloista. In Kulonen, Ulla-Maija (ed.), *Festschrift für Raija Bartens zum 25.10.1993* (Mémoires de la Société Finno-Ougrienne 215), 195–201. Helsinki: Société Finno-Ougrienne.

Kümmel, Martin. 2019. Thoughts about Pre-Indo-European stop systems. In Kloekhorst, Alwin & Pronk, Tijmen (eds.), *The precursors of Proto-Indo-European: The Indo-Anatolian and Indo-Uralic hypotheses* (Leiden Studies in Indo-European 21), 115–130. Leiden: Brill.

Kümmel, Martin. 2020. Substrata of Indo-Iranic and related questions. In Garnier, Romain (ed.), *Loanwords and substrata: Proceedings of the Colloquium held in Limoges (5th–7th June, 2018)* (Innsbrucker Beiträge zur Sprachwissenschaft, Band 164), 237–277. Innsbruck: Institut für Sprachwissenschaft der Universität Innsbruck.

Labov, William. 2020. The regularity of regular sound change. *Language* 96. 42–59.

Läglos = Kylstra, A. D. & Hahmo, Sirkka-Liisa & Hofstra, Tette. 1991–2012. *Lexikon der älteren germanischen Lehnwörter in den Ostseefinnischen Sprachen* I–III. Amsterdam: Rodopi.

Maticsák, Sándor. 2020. *A magyar nyelv eredete és rokonsága*. Budapest: Gondolat Kiadó.

Metsäranta, Niklas. 2020. *Periytyminen ja lainautuminen: Marin ja permiläisten kielten sanastontutkimusta*. Helsinki: University of Helsinki. (Doctoral dissertation.)

Palló, Margit. 1982. *Régi török eredetű igéink* (Studia Uralo-Altaica Supplementa 1). Szeged: Szegedi tudományegyetem.

Pystynen, Juho. 2018. *Itämerensuomen pitkien vokaalien alkuperä*. Helsinki: University of Helsinki. (Master's thesis.)

Reshetnikov, Kirill & Zhivlov, Mikhail. 2011. Studies in Uralic vocalism II. *Journal of language relationship* 5. 96–109.

Ringe, Don. 2004. Reconstructed ancient languages. In Woodard, Roger D. (ed.), *The Cambridge encyclopedia of the world's ancient languages*, 1112–1128. Cambridge: Cambridge University Press.

Róna-Tas, András. 2017. Vitás magyar etimológiák: Válasz Honti Lászlónak. *Nyelvtudományi Közlemények* 113. 37–84.

Salmons, Joe. 2020. *Sound change*. Edinburgh: Edinburgh University Press.

Sammallahti, Pekka. 1979. Über die Laut- und Morphemstruktur der uralischen Grundsprache. *Finnisch-Ugrische Forschungen* 43. 22–66.

Sammallahti, Pekka. 1988. Historical phonology of the Uralic languages with special reference to Samoyed, Ugric and Permic. In Sinor, Denis (ed.), *The Uralic languages: Description, history and foreign influences* (Handbuch der Orientalistik 8: Handbook of Uralic Studies 1), 478–554. Leiden: E. J. Brill.

SSA = Itkonen, Erkki & Kulonen, Ulla-Maija (eds.). 1992–2000. *Suomen sanojen alkuperä: Etymologinen sanakirja* (Suomalaisen Kirjallisuuden Seuran Toimituksia 556 & Kotimaisten kielten tutkimuskeskuksen julkaisuja 62). Helsinki: Suomalaisen Kirjallisuuden Seura & Kotimaisten kielten tutkimuskeskus.

*Suomen vanhimman sanaston etymologinen verkkosanakirja* [Digital etymological dictionary of the oldest vocabulary of Finnish]. Junttila, Santeri & Holopainen, Sampsa & Kallio, Petri & Kuokkala, Juha & Pystynen, Juho (eds.). Helsinki. https://sanat.csc.fi/wiki/EVE-etusivu (Constantly updated. Last accessed 2023-03-24.)

Tálos, Endre. 1975. A magyar hang és alaktan néhány kérdéséhez. *Uralica* 3. 39–53.

Tálos, Endre. 1984. $\frac{\text{Vogul + osztjak}}{2}$. *Nyelvtudományi Közlemények* 86(1). 89–99.

UEW = Rédei, Károly. 1986–1991. *Uralisches Etymologisches Wörterbuch*. Band I–II: Unter Mitarbeit von Marianne Bakró-Nagy, Sándor Csúcs, István Erdélyi †, László Honti, Éva Korenchy †, Éva K. Sal und Edit Vértes, Band III: Register Zusammengestellt von Attila Dobó und Éva Fancsaly. Wiesbaden: Harrassowitz.

WOT = Róna-Tas, András & Berta, Árpád (with the assistance of László Károly). 2011. *West Old Turkic: Turkic loanwords in Hungarian*, I–II (Turcologica 84). Wiesbaden: Harrassowitz.

YSuS = Kallio, Petri. 2019. *Yhteissuomalainen sanasto* [Common Finnic vocabulary]. Online version. https://sanat.csc.fi/wiki/Etymologiadata:YSuS

Zhivlov 2006 = Живлов, Михаил. 2006. *Реконструкция праобско-угорского вокализма* [Reconstruction of Proto-Ob-Ugric vocalism]. Unpublished dissertation. Moscow: Russian State University for the Humanities.

Zhivlov, Mikhail. 2010. Studies in Uralic vocalism I: A more economical solution for the reconstruction of the Proto-Permic vowel system. *Journal of Language Relationship* 4. 167–176.

Zhivlov, Mikhail. 2014. Studies in Uralic vocalism III. *Journal of Language Relationship* 12. 113–148.

Zhivlov, Mikhail. 2015. Reflexes of Proto-Uralic velar nasal in Ugric: An attempt at a Neo-grammarian explanation. In Mantila, Harri & Sivonen, Jari & Brunni, Sisko & Leinonen, Kaisa & Palviainen, Santeri (eds.), *Congressus Duodecimus Internationalis Fenno-Ugristarum, Oulu 2015: Book of abstracts*, 482–483. Oulu: University of Oulu.

Zhivlov, Mikhail. 2023. Reconstruction of Proto-Uralic. In Abondolo, Daniel & Valijärvi, Riitta-Liisa (eds.), *The Uralic languages*, 117–175. 2nd edn. London & New York: Routledge.

*Niklas Metsäranta, Veronika Milanova & Terhi Honkola*

University of Helsinki    Austrian Academy    University of Helsinki
of Sciences

# Borrowability of kinship terms in Uralic languages

Kinship terms are assumed to be universal and central to social life, and consequently they are not particularly prone to borrowing. Borrowing of kinship terms does happen, however, and this provides us a lens with which to evaluate the nature and intensity of contact situations. In this study, we provide a general overview of the borrowability of kinship terms into the Uralic languages. We collected kinship terms from twenty Uralic languages and used a list of 146 kin categories total as the basis for our data collection. We found that affinal kin categories such as those denoting spouses, spouse's siblings, and sibling's spouses had the largest number of loanwords. However, among the kin categories with the largest number of loanwords were also consanguineal categories such as those of 'mother' and 'father'. We also found that the Uralic languages vary notably in how large a percentage of their kinship terminology has been borrowed: the Mordvin languages have borrowed the most, more than 40 percent of their kinship terms, while for many Samoyedic languages no loanwords were detected in their kinship terminology. In addition to the quantitative approach, we also delve into the kin categories with the largest number of loanwords and discuss the patterns of these loanwords in certain languages, and the occurrence of semantic change as a factor explaining the large number of loanwords of terms for 'husband' and 'wife'. All in all, borrowing of kin terms is a context-dependent process and it is challenging to make global generalizations. Nevertheless, we propose that borrowed kin terms could provide us the best possible material through which individual contact situations of the past could be studied. This study also summarizes the borrowed kin terms in the Uralic languages, brings the topic into the spotlight, and pinpoints cases where more research is needed.

## 1. Introduction

All languages have kinship terms to denote family relationships. These relationships range from the biologically closest one between a mother and her child to more remote ones, for example those between a father and his child's spouse's parents. The relationships described with kinship terms are the building blocks of kinship networks which are at the heart of social life in many societies. Along these networks, languages, genes, and cultures are transmitted both vertically from one generation to another and horizontally from one family to another. Conventionally, at least in the Western tradition, kin terms are viewed as part of the basic vocabulary[1] and central to social life, and especially terms denoting close kin are seen as resistant to borrowing, while borrowing of more distant kin terms is not such a rarity (Doerfer 1988: 98–99; Matras 2009: 169–171; 2010: 82). In recent research with a global sample, it was found that while terms denoting more

---

1. Many scholars studying basic vocabulary and lexical universals do not consider the majority of English (European) kinship concepts (even denoting such close kin as 'brother', 'sister', 'sibling', 'son', and 'daughter') to be basic or universal in the global perspective (Milanova et al. 2020: 345–347 with references to Swadesh & Sherzer 1971: 283, Goddard & Wierzbicka 2014: 22–54, and Wierzbicka 2016).

distant kin were borrowed more often, also terms denoting close relatives were borrowed, but they often coexisted with the native term (Honkola & Jordan, in press). Thus, it seems that the patterns of kin-term borrowing are not as simple as they may first seem.

Here we study how language contact has influenced the kinship terminology in the Uralic languages, i.e. what kinship terms are typically borrowed, whence, and when. The borrowed kin terms are, across a number of Uralic languages representing each main branch, examined *vis-à-vis* the loanword layers they are borrowed into. Some of the reasons why certain languages have borrowed kinship terms more readily than others are explored. In essence, we study the borrowability of words in a certain semantic group, namely kinship terms. The study combines etymology with loanword typology, albeit mostly on the level of a single language family, Uralic. The relevance to Uralistics comes first and foremost from the summary of the borrowed kin terms in Uralic languages; the paper also includes minor etymological remarks and additions (see Appendix 2) which hopefully will spark more interest on the topic. While the list of the borrowed kin terms is comprehensive, it also demonstrates the fact that the more western branches, mainly Finnic, Saami, and Hungarian, have been quite thoroughly studied etymologically, while the more eastern branches Mansi, Khanty, and Samoyedic have attracted noticeably less attention.

Uralic is a language family with ca. 40 languages, which today exist in very different sociolinguistic realities. Only Estonian, Finnish, and Hungarian are majority languages that have their own nation-states. The rest are minority languages spoken primarily in Russia as well as in some Nordic and Baltic countries. As these languages are and have been spoken in geographically distant locations for an extended period of time, it is perhaps stating the obvious to say that naturally the languages also differ in what languages they are and have been in contact with. For the Saami languages, Finnic and Germanic have been the two most prominent sources of loanwords, both in borrowed kin terms (cf. Appendix 2) and more generally. For Finnic the most prominently featured source for borrowed kinship terms is Baltic, although the amount of old Germanic loanwords is generally higher. The Uralic languages spoken in Central Russia around the Volga and its tributaries, i.e. Mordvin, Mari, and Udmurt, have borrowed heavily from different Turkic languages, mainly Chuvash and Tatar. Today, for the languages spoken in Russia, Russian is obviously a common source of loanwords in general, and kinship terms in particular.

Loanwords into Uralic have been a longstanding topic of research throughout the last century and even further back, starting with Thomsen 1870.[2] The long-lasting contacts between Indo-European and Uralic languages are somewhat of a given in Uralic studies (cf. e.g. Joki 1973), although the exact chronology of the most ancient loanwords from Indo-European into Uralic is open to interpretation (Aikio 2022: 25). Some have suggested that the earliest loanwords were borrowed already from Proto-Indo-European into Proto-Uralic (Koivulehto 1999: 207–211). The number of potential Proto-Indo-European loanwords in Proto-Uralic is relatively low, however, around a dozen or so, and not without their problems. For loanwords, they also suspiciously include many basic verbs such as 'bring, give', 'fear', 'wash', etc. (op. cit.) instead of nouns, which are more common among loanwords and vocabulary in general. Whatever the case may be, it seems that there are no kinship terms that were borrowed from Proto-Indo-European into Proto-Uralic.

The existence of old Indo-Iranian loanwords in Uralic languages is attested already by a clearly larger number of loanwords, some which can be reconstructed for Proto-Uralic and some which seem to postdate Proto-Uralic (Aikio 2022: 26). An old Indo-Iranian layer postdating Proto-Uralic is possibly where we find the oldest borrowed kinship term in Uralic, as at least MdE *sazor* M *sazər*, MariM *šüžar*, H *šə̂žar* and Udm *suzer* 'younger sister' were ultimately borrowed from a form closely resembling PII *\*swasar-* 'sister'; the borrowing of these words has been suggested to have taken place separately in the predecessors of Mordvin, Mari, and Permic (for further details see Metsäranta 2023: 162–167). Some other clearly prehistorical loanword layers that also include kinship terminology are Proto-Scandinavian and Old Norse loanwords in Proto-Saami and Baltic loanwords in Proto-Finnic. The main bulk of borrowed kinship terms in the Uralic languages are much later loanwords. Turkic languages, namely Tatar and Chuvash, typically started to assert their influence in the Volga-area languages after the Mongol conquest of Volga Bulgaria in AD 1236 (Bartens 1999: 16–17; Bereczki 1994: 14–16) and Russian even later, some of the languages having come into close contact with Russian only in the course of the 20th century.

The kinship terms of the Uralic languages and the nature of their connection to social realities have been a topic of study for more than a

---

2. Some of the loanword studies relevant for our purposes include Qvigstad 1893; Wichmann 1903; Räsänen 1920; 1923; Kálmán 1961; Csúcs 1990, etc.

century (Ahlqvist 1875; Smirnov 1889; 1891; 1895; Setälä 1900; Karjalainen 1913; Harva 1939–1940). From more recent and branch-specific research on the topic the following studies could be mentioned: Whitaker (1955; 1979), Fehlig (1981), Kejonen (2020), Næss et al. (2021) for Saami, Nirvi (1952) and ALFE 2 for Finnic, Szíj (1979; 1998) for Permic, Vavra (1965) and Bíró (2004) for Mansi, Sokolova (1974) for Ob-Ugric, Székely (2016) for Hungarian, and Simčenko (1974), Fainberg (1984), and Volzhanina (2011) for Samoyedic.

Another type of kinship research that has been done in the Uralic context is the reconstruction of Proto-Uralic kinship terms. While several core kin terms can be reconstructed for Proto-Indo-European (Milanova 2020), the situation for Proto-Uralic is quite different, as even the basic consanguineal kin terms such as 'father', 'brother', and 'sister' cannot be reliably reconstructed (Aikio 2022: 24). Equally peculiar is at least the seeming absence of 'child' in Proto-Uralic. Interestingly, most of the more securely reconstructable kinship terms are all terms for different in-laws, e.g. PU *ẹna 'mother-in-law', *ẹppə 'father-in-law', *ẹna-ẹppə 'parents-in-law', *mińä 'daughter-in-law', *wäŋəw 'son-in-law', *ańə 'sister-in-law', *käləw 'sister- or brother-in-law', *nataw 'sister- or brother-in-law', etc. (op. cit.; UED: 54). In general, it can be said the kin terms and the terminologies as a whole have changed notably since Proto-Uralic. It should become clear from the present paper that borrowing is a major contributing factor for these changes and for our inability to reconstruct many of the basic Proto-Uralic kin terms, although certainly not the only factor at play.

In sum, until now both kinship terminologies and loanwords in Uralic have been a topic of extensive research, and there have also been attempts to reconstruct Proto-Uralic kin terms. However, to the best of our knowledge, borrowing of kin terms in the Uralic languages has not been studied earlier in a holistic manner; Milanova et al. (2020) touches upon the topic, but in the current paper we aim at being more exhaustive.

Kinship terminologies can be structured according to various principles. The Uralic languages have a rich diversity of ways for how relatives can be classified.[3] For example, in Finnish there are separate terms for 'brother', 'mother's brother', and 'father's brother' (veli, eno, and setä respectively) whereas in Udmurt agaj denotes both 'elder brother' and 'father's brother' while there is a separate term for 'mother's brother' (čužmurt). One feature

---

3. The Uralic languages do not have grammatical gender, so the gender of the relative is most often marked lexically.

that has particularly rich diversity across the Uralic languages is the relative age distinction, that is, the existence of separate terms for example 'elder sister' and 'younger sister' (instead of having only one term for 'sister') and 'elder brother' and 'younger brother' (instead of having only one term for 'brother'). A complete relative age distinction of sibling terms (four terms) exists in Mordvin, Mari, Udmurt, Khanty, Mansi, and Hungarian, and it is partly present (three terms) in Nganasan and Tundra Nenets, while it is missing (two terms) from Finnic, Saami, and Komi. Languages with relative age distinction for siblings also often follow a similar pattern in other areas of kinship terminology as well, e.g. 'elder sister's husband' and 'younger sister's husband' or 'husband's elder brother' and 'husband's younger brother'. Contact is a likely explanation for the preservation of relative age distinction in at least some modern Uralic languages, but it is probable that already Proto-Uralic had relative age distinction in some capacity (Metsäranta et al. manuscript).

The principles for how kinship terminologies are structured show areal tendencies across language-family borders in general (Trautmann 2001: 282) and this is also seen in Northern Eurasia and Europe where Uralic languages are spoken. The kinship terminologies of Uralic languages spoken in Siberia share similarities with non-Uralic languages of the area, and the same is largely true for the Uralic languages of the Volga-Kama and Circum-Baltic regions. The notable exception to this geographical similarity tendency is Saami kinship terminology, which has some eastern Eurasian features as well as a pattern of alternate generation equivalence – a feature that does not exist in any other Uralic language or in their immediate contact languages.[4] The Saami languages have, however, borrowed several kin terms (Whitaker 1979; Kejonen 2020) and there is some indication that the Saami system has started to change in the same direction as the other Circum-Baltic kinship terminologies.

We studied the kin-term borrowability of twenty Uralic languages covering each main branch of the family. The more precise variety (see Section 2) was often chosen based on the availability of dictionaries and other literary sources. This was the case especially with the eastern Uralic languages.

---

4. Alternate generation equivalence refers to kin-term pairs where the same lexeme or a derivation thereof is used to denote certain pairs of relatives e.g. SaaN *eahki* 'father's elder brother' and *eahkit* '(younger) brother's child' (to their uncle). The closest analogues to this pattern are found in North America, India, South-East Asia, Papua New Guinea, and Australia (Dziebel 2007: 211–254, 322–324).

From these twenty languages, we collected kinship terms and their known etymologies. In the collection of kin terms we used a template list of 146 kin categories (for further details, see Section 2). The collection of the lexical information largely followed the guidelines of the collection of the data in the World Loanword Database (WOLD; Haspelmath & Tadmor 2009).

With this data we aim to answer two sets of questions: 1) *Which kin categories have loanwords in the Uralic languages?* We also divide the data into subgroups based on e.g. consanguinity and gender of the relative to study whether kin categories in one of these subgroups have pronounced numbers of loanwords. Additionally, we delve deeper into the kin categories with the largest number of loanwords and look at both the extra- and intralinguistic reasons as to why kin terms in these particular categories might be the most commonly borrowed. 2) *Which Uralic languages have borrowed kin terms?* We also study from which languages these terms have been borrowed. Furthermore, we discuss the occurrence of kin-term loans in the light of what is known about the contact situation in question and, conversely, what can be deduced about the contact situation based on the presence or absence of borrowed kin terms.

In what follows, in Section 2 we explain the principles of data collection and key concepts. We focus on explaining how the collection took place and some of the challenges our approach might entail. In Section 3, we present our results and discussion. We have subdivided this section based on the two research questions mentioned above. In 3.1, we found that terms denoting affinal relatives have been borrowed the most, but among the most borrowed ones were also terms denoting close relatives. We discuss the patterns of borrowing for some of these categories and highlight the cases where semantic change has likely played a role in the process. In 3.2 we see that the Uralic languages vary notably in how many kin terms they have borrowed depending on their contact history, but also of how well the languages in question have been studied. Finally, in Section 4 we conclude our study, summarize its main findings, and give an insight into our ongoing work, as well as discuss possible avenues for further study.

The paper also has four appendices. In Appendix 1, we list the kinship categories included in this study. In Appendix 2, we present the research material, i.e., the borrowed kin terms and their etymologies. In Appendix 3, we show the complete list of kin categories with the number of borrowing events, and in Appendix 4, we list kin categories for which no loanwords were detected in the Uralic languages.

## 2. Data collection

We collected kinship terms and their etymological information, including loanword status, from twenty languages covering each main branch of Uralic: Saami (South, North, Skolt), Finnic (Finnish, Veps, Estonian, Livonian), Mordvin (Erzya, Moksha), Mari (Hill, Meadow), Permic (Komi-Zyrian, Udmurt), Mansi (Sosva), Khanty (Kazym), Hungarian, and Samoyedic (Tundra Nenets, Forest Enets, Nganasan, Taz Selkup). Initially, we used a list of 115 kin categories total as the basis for the data collection. This list has been developed to collect kinship terminologies worldwide and it includes 88 categories of genealogical kin and 27 categories for kin by marriage (i.e. affinal relatives) (for further information see Passmore et al. 2023). We added 31 categories to the original list so that it would meet the needs of our project better when collecting kin-term data from Uralic languages.[5] These categories covered relative age distinction (elder/younger) of affinal relatives. For example, instead of having a category only for 'husband's sister' we added new categories for 'husband's elder sister' and 'husband's younger sister'. Thus, in total, data was collected from 146 kinship categories. The list of original and added categories can be seen in Appendix 1.

### 2.1. Kinship terms and their etymology

We considered a kinship term to exist in a language if it was found in a dictionary or other lexical source we used either as its own entry or, at minimum, as a part of another, as this would imply at least some level of convention. We included phrasal expressions only when they were found in a dictionary, as in those cases the expression could be considered to be fixed and conventionalized (following the guidelines of Haspelmath & Tadmor 2009: 11). This requirement was necessary, as all familial relationships can be described with phrasal expressions (e.g. the English kin term *uncle* can be described as parent's brother). In our data one kin category could have more than one kin term (e.g. in Komi both *ćoj* and *soć* denote sister) and one kin term could fill more than one kin category (i.e. polysemic terms, e.g. Meadow Mari *aka* is 'elder sister; parent's (father's or mother's) younger sister').

---

We aimed for collecting the standard modern varieties of the language, but this was not always possible. We collected kinship terms mainly from various dictionaries, both print dictionaries and those available online (a full list of the materials utilized can be found under Lexical sources). For some of the smaller and uniform languages in our study (such as Nganasan or Forest Enets), the choice of which dictionary to utilize was rather straightforward, as there simply are only a few dictionaries to choose from. In many cases bilingual dictionaries were used and also dialectal dictionaries were utilized if found necessary or otherwise helpful. We avoided using dialectal materials as primary sources mainly because the bulk of them were collected well over a century ago and we were generally aiming for the modern standard variety. In a few cases, in the absence of comparably comprehensive material resources this could not be feasibly avoided. For example, our Sosva Mansi material is based on *Wogulisches Wörterbuch* (WogWb), as alternatives of matching scope (i.e. Munkácsi & Kálmán 1986) are also dialectal and fairly similar in terms of when the materials were collected. Generally, we have striven to use primary sources. However, for a few languages or branches of languages, there exist comprehensive descriptions of their kinship terms along with etymological information (such as Karjalainen 1913 for Khanty, Mészáros 2001 for Mordvin), so we chose to use these sources as the basis for our data collection. The collected kinship terms and their references are part of the Kinbank database (kinbank.net; Passmore et al. 2023) and can be found online (Honkola et al. 2022; github.com/kinbank/kinura).[6]

After collecting the kinship terms, the task was to gather all the existing etymologies – that is, particularly to include information whether they are borrowed or not – for them. It bears repeating that within the Uralic language family the geographically more western languages have been the subject of more rigorous etymological research. Traveling from west to east, the amount of etymological research declines steadily. Estonian, Finnish, and Hungarian are the most thoroughly studied and there exist several etymological dictionaries of these languages. For languages which do not have etymological dictionaries of their own (Mansi, Tundra Nenets, Forest Enets, Nganasan, Taz Selkup), etymological notes from individual articles and studies were used as well as Uralic etymological dictionaries, e.g. the UEW and UED. The above-mentioned imbalance in the amount

---

6. With the exception of Taz Selkup, which was not added to Kinbank, as it is based on an unpublished source (Helimski 2007) not readily available.

of research into different languages has necessarily had an effect on our results, too; the lack of borrowed kin terms especially in the eastern Uralic languages may also at least partly be due to lack of research. Nevertheless, keeping this in mind, our paper provides a comprehensive list of borrowed kin terms in the Uralic languages.

## 2.2. Information about borrowing

For each kin term we defined whether the term was known to be borrowed or not. In cases when a term was borrowed, we also collected information about the source language and the time of the borrowing. As this task is not as straightforward as it may first seem, in the following sections we provide details about this procedure.

### 2.2.1. Defining a loanword and analyzability

A *loanword* is defined as a word that at some point in the history of a language entered its lexicon as a result of borrowing (Haspelmath 2009: 36). In this study *borrowing* is used to refer to the point in time when the transfer of lexical units happens and to denote this process in general, i.e. a loanword is the linguistic unit that is transferred, and borrowing is the process by which it is transferred. Loanwords are typically unanalyzable in the recipient language even if they are more complex in the source language (op. cit. 37). Hungarian *mostoha* 'stepmother' is a Slavic loanword, cf. Czech *macecha*, Slovak *macocha*, Bulg *мащеха*, Ru *мачеха*. This can be determined by analyzability, as the Slavic words are derived from the common Slavic word for 'mother', cf. Old Church Slavonic *mati*, Old Czech *máti*, Ru *матъ* < Proto-Slavic *\*màti* (Derksen 2008: 303), with the suffix *\*-juxa* (*-jexa*) 'step-' (Matasović 2014: 152). Such an analysis cannot be done for the Hungarian word, which is opaque in form and thus a loan. In general, analyzability is used as one of the criteria by which the direction of borrowing is determined.

The example above is a straightforward example of a loanword. Our data contains a rather large number of loanwords that have been further modified in the recipient language, usually by derivation or compounding. Somewhat typical examples of derivations are diminutive derivatives such as SaaSk *päärnaž* 'child' ← SaaSk *pä'rnn* 'son, boy' (< PS *\*pārnē*) ← Scand, cf. ON *barn* 'child' or Veps *baboi* 'grandmother' ← Ru *баба* 'old woman',

Veps *dedoi* 'grandfather' ← Ru *дед* 'id.', and compounds containing a loanword or loanwords, e.g. Komi *bať-mam* 'parents' (*bať* 'father' ← Ru *батя* and *mam* 'mother' < Proto-Permic *\*måm*), Udm *anaj-ataj* 'parents' (*anaj* 'mother' ← Ta *ana, änej* and *ataj* 'father' ← Ta *ätej, ataj*) (Csúcs 1990: 104, 112). According to a definition given in Haspelmath & Tadmor (2009: 12) and Haspelmath (2009: 37), if a word is analyzable in the recipient language and has in a way been "created" in the recipient language, then it is no longer treated as a loanword. Under this definition, SaaSk *päärnaž* 'child', analyzable as a diminutive derivation in Skolt Saami, is not a loanword, while *pä'rnn*, which is underived, is a Scandinavian loanword. In our study we follow the criteria set by Haspelmath & Tadmor (2009: 15) and only discuss kin terms which are loanwords in the strict sense, i.e. not further modified by derivation or compounding in the target language. We do this in order to keep the amount of data manageable and our dataset comparable with that of WOLD.

## 2.2.2. Certainty of borrowing

In this section, we will be discussing the etymological treatment of the data. One of the things we did was to try to evaluate the reliability of the etymologies that have been proposed in previous literature and the certainty of borrowing. To this end, each kinship term was assigned a value ranging from 4 to 0 following the five-point classification used in Haspelmath & Tadmor (2009: 12–13). We also follow Haspelmath & Tadmor (2009: 20) in that we consider words in classes 4 and 3 as loanwords and focus our discussion in this paper on those.

| | |
|---|---|
| 4 = clearly borrowed | 1 = very little evidence for borrowing |
| 3 = probably borrowed | 0 = no evidence for borrowing |
| 2 = perhaps borrowed | |

We will give some examples and try to formalize how words were divided into these categories. In general, the validity of proposed loan etymologies is evaluated based on matching phonological shape and matching meaning – these are factors on the lexical level. We also considered the validity of the suggested source language, i.e. are there other loanwords from the same source and how well established the prehistorical or historical contacts between the languages in question are.

Words that are clearly borrowed were assigned the value 4. These often include, among others, recent loanwords between languages that are

known to have been in contact from verifiable historical sources or are perhaps still in close contact. These are typically loanwords that can be spotted even by non-experts, as the borrowing has taken place in the not-too-distant past, so that the phonological shapes between them deviate only slightly or not at all and the meanings are similar enough to be recognizable. Words with value 4 include, for example, many of the recent Russian loans into individual Uralic languages, Komi *vnuk* 'grandchild' ← Ru внук, MdE *ďaďa* 'father's brother, uncle' ← Ru дядя, MdE Komi *šurin* 'wife's brother' ← Ru шурин, Komi *plemjannik* 'brother's or sister's son, nephew' ← Ru племянник. Minority-language speakers are nowadays almost uniformly bilingual in their native language and Russian, so these examples might even be difficult to distinguish from code-switching.

In the previous cases the phonological match is one to one, but this does not need to be the case and phonological substitutions and adaptations, if predictable, do not in our view change the level of certainty. Meadow Mari *ońo* and Hill Mari *ońə* 'father-in-law' can be explained as loanwords from Chuvash *χoń*, *χuń* (Räsänen 1920: 166), as zero substitution for Chuvash *χ*-is common in other Chuvash loans as well (although admittedly Hill Mari shows two substitution patterns, zero substitution and *χ*-).

Ideally, we would want to determine a chronologically clearly defined source for all loanwords. Failure to do so unequivocally does not automatically mean that the certainty of borrowing is any less, however. The relationship between the Uralic words MdE *sazor* 'younger sister', M *sazʕr*, MariM *šüžar*, H *šʕžar*, Udm *suzer* is phonologically ambiguous and difficult to interpret conclusively (see Metsäranta 2023 : 162–167). It is certain that no matter which specific chronology we settle for, the words are all certainly borrowed from an Indo-European word ultimately reflecting Proto-Indo-European *\*su̯ésor-* 'sister' (Milanova 2021: 113–117), although the interpretation we give them can have a profound impact on how we view the prehistory of these languages. The status of different 'sister' words as loanwords does not change even though there are many ways in which their internal relationship and chronology can be interpreted.

In sum, etymologies were assigned value 4 if they exhibited the following characteristics: 1) regular phonological match between the source and the target; 2) semantic match; 3) belonging to a known loanword layer, i.e. there are other words borrowed from the same source and not just the kinship term in question. If one of these criteria was not met or there was otherwise uncertainty related to the etymology, value 3 = probably borrowed

was used instead. For example, a case that was demoted due to a semantic mismatch between the source and target word is Meadow Mari *βate* 'wife' and Hill Mari *βätə*, which was deemed value 3 because the proposed loan original for the Mari words, Chuvash *vatậ*, means 'old' (Räsänen 1920: 120) rather than 'wife'. Although the difference in meaning between the Mari and Chuvash words can be explained through semantic change (see Section 3.1.2), the change itself does muddy the waters enough so that the case can no longer be viewed as "clearly" borrowed.

As another example of a case where value 3 was assigned instead of 4 we could mention Proto-Saami *\*muoϑā/ē* 'mother's younger sister' > SaaS *muahra*, N *muottá*, Sk *mueʹđđ*, which is thought to have been borrowed from Proto-Germanic *\*mōþō(n)* 'mother's sister' (Kümmel 2015: 121–129). Although we generally find the etymology plausible, the reasoning behind assigning it a value of 3, instead of 4, lies in the fact that the particular word form is not actually attested anywhere in Germanic (although similar derivatives do exist). The phonological correspondence between Proto-Saami and Proto-Germanic is expected, the meanings are a close match, and there are otherwise a large amount of Germanic loans in Saami, but borrowing from an otherwise unattested form does add a level of uncertainty, hence a value of 3.

Values 2 and 1 were assigned for poorly defined and uncompelling etymologies. In general, *Lallwörter* were assigned value 2, especially if the loan original could not be determined with any level of certainty. For example, Hungarian *papa* 'father's father, mother's father' can be a loanword from German *Papa* 'father', but this does not need to be the case and the Hungarian word can certainly have been borrowed from many other languages as well. Ill-defined etymologies were assigned value 1. For example, we have Komi *geṯịr* 'wife, spouse' that has been compared to German *Gatte* and related Germanic words, perhaps entering Komi from Old Norse through Finnic (KESKJa: 81). However, as there is no Finnic word that could be considered as the mediator and there is otherwise no known layer of Old Norse loans in Komi, there is no compelling reason to believe that the word in Komi is of Germanic origin.

Value 0 represents words with no evidence for borrowing. This is not to say that words with value 0 could not be loanwords, but rather that they have not been treated as loanwords and/or no credible loan etymologies have, to the best of our knowledge, been proposed for them in the etymological literature. This group of words is heterogeneous as it includes

1) inherited Uralic words; 2) words that can be reconstructed for different branch ancestors, e.g. Proto-Finnic, Proto-Permic, etc.; and 3) words of unknown origin, i.e. they are not known to be borrowed but only exist in individual modern languages, and therefore their origin is an open question. Finally, this group also contains words that were excluded from our study by the definition of a loanword used here, i.e. calques and loan blends (derivations and compounds); see Section 2.2.1.

In Appendix 2 we present the research material used in this study, that is, all the etymologies in our material that were deemed either value 4 or 3. This material was used to calculate the borrowability rates and properties tied to that. The material is organized by branch and has been chronologically ordered. The chronology provided is relative and there to give the reader a rough estimation as to which stage the borrowing occurred at. Unfortunately, an extensive etymological analysis of the research material is not possible here, but some brief etymological remarks are included in this appendix. There are a few rather major departures from the standard views expressed in the etymological literature regarding the chronology and validity of certain borrowed kinship terms; these are treated more closely in Metsäranta (2023).

## 3. Results and discussion

In what follows, we present our findings both from the perspective of kin categories (Section 3.1) and from the perspective of the languages studied (Section 3.2). In Section 3.1 our focus is first specifically on the number of borrowing events per kin category (Table 1), and after that on the number of loanwords (Table 2).

For Table 1 we calculated the number of borrowing events in two different ways. First, we counted separately all borrowing events. For example, if a term denoting a category (e.g. 'father's father') had in a language (e.g. Finnish) been borrowed twice (from Swedish both *pappa* and *vaari*), it was counted as two borrowing events. In the second, perhaps less intuitive, calculation we had a restriction that the maximum number of borrowing events per language / language stage is one.[7] Thus, in this calculation the

---

7. The proto-language stages considered here are the well-established branch ancestors, that is, Proto-Finnic, Proto-Saami, Proto-Mordvin, Proto-Mari, and Proto-Permic (see also Appendix 2).

two Swedish loanwords for 'father's father' in the above-mentioned example are counted as one.[8] The reason to limit the maximum number of borrowing events per language / language stage comes from the aim to keep the results stabilized, so that for example a large number of recent Swedish loans into Finnish would not distort calculations which are supposed to illustrate the frequency of borrowing throughout the Uralic family (something which is already somewhat distorted due to the lack of research into the easternmost Uralic languages). In addition, a similar principle of calculation was used in Honkola & Jordan (in press), making these numbers comparable with those when the same principle of calculation is followed.

In Table 1 we also present the total number of languages in which the kin category in question is occupied by a loanword in our material. It needs to be borne in mind when having several daughter languages of the same parent language in the sample, that the word was possibly borrowed already into the proto-language stage instead of the individual languages. For example, if a term was borrowed into Proto-Saami and it exists in the three modern Saami languages included in the sample, in terms of absolute numbers there are three loanwords as a result of one borrowing event. As we are interested in the borrowability of kin terms instead of the absolute number of loanwords in our sample, we have focused on the number of borrowing events, as that would seem to give a more reliable picture of the actual borrowability.

In the calculations presented in Section 3.1 the relative age distinction of affinal relatives (i.e. the additional categories mentioned in Section 2 and listed in Appendix 1) are merged into their main categories. That is, for example, the merged category of 'wife's brother' includes also terms for 'wife's younger brother' and 'wife's elder brother'. Merging of categories was done as although age distinction is rather prevalent, it is not a universal feature of kinship terms in Uralic languages. In other words, if the additional categories would have been kept separate in our calculations, it would have automatically excluded a number of languages by definition

---

8.  If a term has been borrowed both into the proto-language and into the individual modern languages in a certain branch, these were counted separately. For example, a term for 'husband' has been borrowed into Proto-Saami from Proto-Norse, and later again into South Saami from Scandinavian and into North Saami from Finnic, resulting in three instances of borrowing for the category of 'husband'.

from certain categories – something which should be avoided when summarizing large datasets. When calculating the number of kin categories with loanwords per language (results presented in Section 3.2) the categories of relative age distinction for affinal relatives were kept separated. This way we obtained a more realistic picture about the number of kin categories existing in each language.

## 3.1. Which kin categories have loanwords?

In total 68 kin categories had borrowing events and thus also loanwords in Uralic languages. The distribution of loanwords into these categories is very uneven, however, as loanwords in the 18 top categories listed in Table 1 covered 57.4% of all the loanwords (in total 157 loanwords, see Appendix 2). Terms denoting certain affinal relatives – that is, the more distant relatives – have been borrowed the most. A similar pattern of borrowing has been suggested earlier (Doerfer 1988: 98–99; Matras 2009: 169–171; 2010: 82) and has also been found from the global dataset of WOLD (Honkola & Jordan, in press). The categories with the largest number of borrowed terms include mainly 'sibling's spouse' ('sister's husband') and 'spouse's siblings' ('husband's brother', 'wife's brother', and 'wife's sister'). In many languages, the kin terms in these affinal categories are polysemous. For example, in most languages a term for 'sister's husband' also means something else, such as 'wife's brother' (e.g. in South Saami *maake*, Finnish *lanko*, and Hungarian *sógor*) or 'daughter's husband' (e.g. Erzya *ezna* and Komi *źať*). Loanwords into these affinal categories are a topic of closer inspection in Section 3.1.1.

While the kin categories which most often have loanwords in the Uralic languages are affinal, also kin terms denoting the closest familial relationships such as 'father', 'mother', 'child', 'elder sister', 'husband', and 'wife' have been borrowed in several languages (Table 1).[9] In the global study made by Honkola & Jordan (in press) with the WOLD dataset, it was found

---

9.  We use here a slightly modified version of the close/distant categorization used in Honkola & Jordan (in press) and consider the categories 'mother', 'father', 'sister', 'brother', 'son', 'daughter', 'child', 'wife', and 'husband' as "close" kin categories ('child' was not included in the list of Honkola & Jordan) and all other kin categories as "distant". This kind of a binary division may feel artificial, but it is used here to capture the main axis of kinship interaction.

Table 1: Kin categories with the largest number of borrowing events in Uralic languages. Boldface font indicates the close kin categories. Parameter column introduces the abbreviations of the kin categories which are used later in the text. Parentheses indicate categories where the age distinction (e = elder, y = younger) has been merged into the main category. The table has been sorted by the total number of borrowing events. The column with a maximum of one borrowing event per language has a § in cases when a term in the kin category in question has been borrowed both into proto-language and into the individual languages in one of the subgroups; in the case of merged categories the § sign indicates that two categories exist in one language (e.g. Erzya has both WZ and WeZ) and these have been counted separately. Number of languages with a loanword shows the total number of languages in which the kin term in question is a loanword. Asterisk in the Polysemy column indicates that at least in one of the languages the kin term is polysemous (i.e. linked to more than one kin category). The number in the Coexistence column indicates in how many languages the borrowed kin term coexists with a term with no evidence of borrowing. The full list is given in Appendix 3.

| Kin category | Parameter | # of borr. events (total) | # of borr. events (max. 1 / lang.) | # of lang. with a loan-word | Poly-semy | Co-ex. |
|---|---|---|---|---|---|---|
| sister's husband | (e)ZH | 11 | 9 § | 10 | * | 3 |
| husband's brother | H(e/y)B | 10 | 10 | 12 | * | 0 |
| wife's brother | W(e/y)B | 10 | 10 § | 10 | * | 1 |
| **wife** | **W** | **9** | **7 §** | **8** | | **4** |
| grandchild | CC | 9 | 7 | 7 | | 2 |
| **husband** | **H** | **8** | **8 §** | **10** | | **3** |
| wife's sister | W(e)Z | 8 | 8 § | 8 | * | 0 |
| wife's sister's husband | W(e)ZH | 7 | 7 | 8 | * | 0 |
| **child** | **C** | **7** | **7 §** | **7** | | **4** |
| father's father | FF | 7 | 5 | 5 | * | 3 |
| father's mother | FM | 6 | 6 | 6 | * | 3 |
| **elder sister** | **eZ** | **6** | **5** | **6** | * | **1** |
| **father** | **F** | **6** | **5** | **6** | | **2** |
| **mother** | **M** | **5** | **5** | **6** | | **2** |
| daughter's husband | DH | 5 | 5 | 5 | * | 4 |
| mother's mother | MM | 5 | 5 | 5 | * | 2 |
| sister's son | ZS | 5 | 4 | 5 | * | 1 |
| sister's daughter | ZD | 5 | 4 | 5 | * | 2 |

that in cases where a term denoting a close kin category was borrowed, the borrowed term often coexisted with the variant with no evidence of borrowing. This was especially the case with 'father' and 'mother', as in all languages where a term for mother was borrowed, it coexisted with a non-borrowed term; with 'father' this was also the case except in the two languages where the term also denoted father's brother. This kind of a pattern is not, however, seen in our data, as in the categories of both 'mother' and 'father' in three out of five borrowing events the borrowed term has replaced the native variant. Terms for both 'mother' and 'father' have been replaced in Erzya, and in Hill Mari and Meadow Mari (borrowing took place in Proto-Mari). In addition, a term for 'mother' has been replaced in Finnish and a term for 'father' in Komi.[10]

Of the 146 kin categories used as the template in our data collection, 20 did not have any loanwords despite being a relatively common category in the Uralic language family (Appendix 4). As a criterion for being a "relatively common category", a term for that category needed to exist in three or more languages covering more than one subgroup of the Uralic family. These categories include, for example, the age distinction of siblings when denoting the nephews and nieces (e.g. 'younger brother's son', 'elder sister's daughter') and terms for grandchildren (e.g. 'son's son', 'daughter's son'). The reason why these categories appear not to be affected by borrowing is that these categories typically include phrasal expressions, e.g. MdE *ťejťeŕeń ćora* 'daughter's son', that are not considered strictly speaking loanwords in this study even though they might involve borrowing in some way (see Section 2.2.1).

We also studied the number of loanwords when dividing the data into certain subgroups based on consanguinity, generation, and gender (Table 2). For these calculations, we used a so-called "balanced" version of the data, in which the focus is on the loanwords instead of kin categories. The difference between these two approaches is clarified with the following example. When calculating the number of borrowing events in Table 1, each language that had a borrowed term in the kin category in question was counted as one borrowing event. For example, as Finnish *mummu* 'mother's mother' is a loanword from Swedish, it is counted as one borrowing

---

10. The inherited variant for 'mother' could still exist in the language in question but with a different meaning. For example, in Finnish *emä*, the Uralic variant for 'mother', denotes animal mother instead of human mother.

event for this MM category. However, as *mummu* denotes also father's mother it is counted as a loanword also for the FM category. Now, when we are interested in the total number of actual loanwords, taking these values directly would give us an impression that *mummu* would have been borrowed twice, which is not the case. Therefore, to compare loanwords in different groups in a more realistic way, we made a balanced data where we counted the borrowed kin terms instead of having the focus on the categories. This means that as we have one term *mummu*, which covers two categories (MM and FM), each category was given a value of 0.5 so that those sum up equal with the fact that we are now talking about one single loanword. A similar practice was followed when a kin term covered more than two categories. For example, Veps *bratan* 'male cousin' fills four categories ('father's brother's son', 'mother's brother's son', 'father's sister's son', 'mother's sister's son'). As a result, in the balanced data for Veps each of these categories got a value of 0.25.

When comparing different subgroups with each other, the results are the following. We found that the largest number of borrowings had taken place in consanguineal categories (Table 2). This is however due to the fact that there are more consanguineal categories in our data. Affinal categories had a higher number of loanwords per category (2.4 loanwords/category) than what consanguineal categories had (1.5 loanwords/category). When counting the number of kin categories, which had a large number of loanwords (five or more loanwords; three or more loanwords), the affinal categories had a larger percentage of categories with a large number of loanwords. This can be seen also from Table 1.

When comparing subgroups divided by the generation of the relative in question, we found that terms denoting relatives in ego's generation (e.g. 'sister', 'husband') were borrowed the most when taking into account the number of categories in each of the subsets (Table 2). This is contrary to what was found with the global dataset of WOLD, as there terms denoting elder relatives were borrowed the most (Honkola & Jordan, in press). Finally, when comparing the subgroups which were divided by the gender of the relative in question, we found that terms denoting male relatives were borrowed slightly more than terms denoting females. This is different from what was found in Honkola & Jordan (in press) where no difference between these groups was found. Terms denoting a category without a specific gender (e.g. 'sibling') had a large number of loanwords based on the average and the percentage of categories, which had three or

more loanwords (28.6% in both). This is because there were two categories with a large number of loanwords ('grandchild' and 'child'), whereas other general categories had less than three loanwords (Table 2).

Table 2: Comparison of subgroups divided by consanguinity, generation, and gender. Number of kin categories in each subgroup is shown in the first column including the relatively common categories with no detected loanwords (Appendix 4); in the second column is the percentage of borrowed kin terms in the subgroup in question when calculated from the balanced data. This percentage does not take into account the different sizes of the subgroups. In the third column we show the average number of loanwords per kin category in the subgroup in question. This value takes into account the different sizes of the subgroups. In the fourth and in the fifth column we present the percentage of categories with a large number of loanwords. In the fourth column the threshold of what is considered a large number of loanwords is five; in the fifth column this threshold is three.

| | # of kin categories | % of loanwords | Average number of loanwords / category | % of categories with ≥5 loanwords | % of categories with ≥3 loanwords |
|---|---|---|---|---|---|
| **Consanguinity** | | | | | |
| Consanguineal | 58 | 55.4 | 1.5 | 8.6 | 20.7 |
| Affinal | 30 | 44.6 | 2.4 | 16.7 | 33.3 |
| Total | 88 | | 1.8 | | |
| **Generation** | | | | | |
| Older than ego | 30 | 28.6 | 1.5 | 10.0 | 16.7 |
| Ego's generation | 33 | 47.7 | 2.3 | 15.2 | 33.3 |
| Younger than ego | 25 | 23.7 | 1.5 | 8.0 | 24.0 |
| Total | 88 | | 1.8 | | |
| **Gender** | | | | | |
| Male | 41 | 47.6 | 1.8 | 12.2 | 26.8 |
| Female | 40 | 39.1 | 1.5 | 7.5 | 22.5 |
| General | 7 | 13.3 | 3.0 | 28.6 | 28.6 |
| Total | 88 | | 1.8 | | |

### 3.1.1. Borrowed terms for 'spouse's siblings' and 'sibling's spouses'

Loanwords occur most commonly in kinship categories for in-laws, with terms for brother-in-law being the most commonly borrowed, cf. ZH 'sister's husband' (11), HB 'husband's brother', (10), and WB 'wife's brother' (10). Terms for WZ 'wife's sister' are also among the top most borrowed kin terms, with eight instances of borrowing. HZ 'husband's sister' is clearly the least borrowed of the terms, as in our materials there are only two loanwords in this category (see Appendix 3). The terms for ZH 'sister's husband' (11) are clearly more likely to be borrowed than the terms for BW 'brother's wife' (4) (Appendix 3). Terms denoting these categories have been borrowed in all subgroups of the Uralic family with the exception of the Samoyedic languages, which have very few detected loanwords in their kinship terminology overall. The majority of the loanwords are found in Mordvin, Mari, and Komi, due to which in the following we focus our discussion on those languages.

Some general tendencies can be found from the loans in the above-mentioned categories. Borrowed affinal kinship terms for 'spouse's sibling' and 'sibling's spouse' rarely if ever come from anything other than similar kinship terms in the source languages. Semantically, borrowed kinship terms for 'spouse's siblings' and 'sibling's spouses' form a sort of continuum. On the one end of this semantic continuum, there are words expressing quite general meanings of affinal relatedness that have perhaps come to cover several different categories, and at the other end are highly specific kinship terms that usually occupy only one category or two at maximum. Even within the more polysemic affinal kinship terms, whatever polysemy exists there is cross-polysemy between the categories ZH, HB, WB, WZ, HZ, and BW.

Quite specific meanings of affinal relatedness seen in modern languages are oftentimes the result of semantic narrowing. An example of a loanword that referred to relatives in general and later narrowed in meaning is Finnish *lanko* 'brother-in-law', which in the standard literary language covers categories WB, HB, and ZH. This constellation of meanings is evidently a Finnish-specific development. In old literary Finnish, the oldest source from 1637 defines *lanko* as 'cognatus, frände, ein Blutfreund; cognata, frändka, die Freundin' (VKS: s.v. *lanko*). The meaning 'brother-in-law' is first encountered in Daniel Juslenius' dictionary from 1745. Also the cognates of Fi *lanko* in other Finnic languages more commonly refer to an affinal relative in general, cf. Karelian *lanko* 'affinal relative, esp. sister's

husband', Est *lang* 'affinal relative' (SSA 2: 44–45). The Finnic word is usually thought to be a Germanic loanword, cf. PGrm *\*ga-langaz* > Old High German *gilang* 'brother-in-law, relative' / PGrm *\*bi-langaz* > Old Low German *bilang* 'related; joined, connected' (LägLoS: 167; SSA 2: 44–45).

Another rather similar case of narrowing of a borrowed term with a more general meaning in the source language is SaaS *siʹbjege* 'en manns el. gutts eldre brors el. fetters kone; ens ektemanns yngre bror el. fetter' (eBW = HyB) and SaaN *sivjjot* 'affinal relative of the opposite sex; husband's brother or male relative, sister's husband or male relative (to their wife's sister); wife's sister or female relative, brother's wife (to their husband's brother)' (HB = fZH = WZ = mBW). Both terms have been borrowed from Scandinavian, cf. ON *sifjungr* 'verwandt', which refers to affinal relatives in general. The polysemy eBW = HyB presently found in Saami is probably largely due to the Old Norse loanword replacing a Proto-Saami reciprocal pattern still found in Kildin and Ter Saami (PS *\*oańē* > Kildin *vuəńń*, Ter *vi̯eńńe* 'elder brother's wife' → PS *\*oańēp* > Kildin *vuəńńev* and Ter *vi̯eńńev* 'husband's younger brother') (Bergsland 1942: 176–177; Itkonen 1958 [2011]: 789).

When talking about highly specific kinship terms, if the loan original features relative age distinction, this distinction is often present in the target language as well. In addition, it seems likely that the borrowing of kinship terms with age distinction has at least partly resulted in the development or reinforcement of a similar dichotomy also in the target language. Mari, which has borrowed most of its kinship terms for elder affinals, is a prime example:

1. Spouse's elder sibling or elder sibling's spouse in Mari

- WeB = HeB MariM *ońô̂ska*, H *ońô̂ska* ← Chu dial. *χońô̂ska*, cf. lit. *хунчӑкам* 'brother-in-law (wife's elder brother)'
- WeZ = HeZ MariM *ońaka*, H *ońaka* ← Chu dial., cf. lit. *хунакам* 'sister-in-law (wife's elder sister)'
- eBW MariM *jeŋga*, H *jeŋgä* ← Ta *җиңгии, җиңгә(й)* 'elder brother's wife' perhaps via Chu, cf. *инке* 'sister-in-law (elder brother's wife)'
- eZH MariM *kurska*, H *kô̂rska* (←)[11] Chu *kərü* 'Bräutigam, Schwager'

---

11. The Mari word is structurally obscure, but the stem is ultimately from Chuvash. We have not included the word in the research material used in this study.

2. Spouse's younger sibling or younger sibling's spouse in Mari

- WyB = HyB MariM *pörâž*, H *pörâž* (not known to be borrowed)
- WyZ = HyZ MariM *nuδo*, H *nuδâ* (inherited Uralic word)
- yBW MariM *šeške*, H *šeškə* (not known to be borrowed)
- yZH MariM *βeŋe*, H *βiŋgə* (inherited Uralic word)

The terms for 'spouse's elder siblings' and 'elder sibling's spouses' are typically borrowed in Mari, whereas the terms for their younger counterparts are usually either inherited Uralic words (in the case of WyZ and yZH) or at least are not known to have been borrowed from anywhere (in the case of WyB and yBW). One could perhaps argue that the borrowing of elder affinals from Chuvash (and in a single case perhaps from Tatar) introduced the relative age distinction to Mari, after which this feature spread to existing native terms and produced the dichotomy we find in Mari today. It might also very well be that contact with the Turkic languages only reinforced a distinction already present in Mari.

The same pattern repeats between eZ 'elder sister' and yZ 'younger sister'. The word for 'elder sister' in Mari, M *aka*, H *äkä*, was borrowed from Chuvash and the existing word, MariM *šüžar* 'younger sister', MariH *šâžar* 'id.' (itself originally a loanword, but probably predating the Chuvash contacts) in turn presumably narrowed in meaning from 'sister' to 'younger sister'. This pattern does not hold all the way through borrowed consanguineal and affinal kinship terms in Mari, the most notable counterexample being eB 'elder brother' and yB 'younger brother'. The term used for eB is possibly an old Uralic kinship term or at least does not have a credible loan etymology, M *iza*, H *əzä* < PM \**ĭćä* 'elder brother; father's younger brother' (UED: 30–32), whereas the term for yB, M *šoľo*, H *šoľa, šoľâ*, is a loanword from Chuvash. Despite this one obvious counterexample, the pattern is rather pervasive. However, it must be said that Mari is rather the exception in the symmetry and uniformity of its borrowed kinship terms in comparison to, for example, the Mordvin languages, where the kinship system seems to have been more in flux with different layers of loanwords present.

The most frequently borrowed single term for 'wife's brother' is Russian *шурин* 'brother-in-law (wife's brother)' that has been independently borrowed by both the Mordvinic languages Erzya and Moksha and by Komi. It is perhaps interesting to note that none of the languages in our material has borrowed the "mirror image" of this kinship term, i.e. Russian *золовка* 'sister-in-law (husband's sister)', which seems to correlate with the fact that

'husband's sister' is clearly the category with the smallest number of loan-words of spouse's siblings. This seems to be an even more general trend. In general, the Russian kinship terms denoting the wife's relatives have been more attractive objects of borrowing, even so much so that the Russian words for the husband's relatives, *свёкор* 'husband's father', *свекровь* 'husband's mother', *деверь* 'husband's brother', and *золовка* 'husband's sister' do not seem to have been borrowed by any of the Uralic languages we surveyed. Even languages such as Komi, which has borrowed all of its kinship terms denoting the wife's relatives from Russian, has not borrowed any of the terms denoting the husband's relatives from Russian and instead uses native derivatives and compounds.

3.  Wife's parents and siblings in Komi

- *t'est'* 'father-in-law (wife's father)' ← Russian *тесть* 'id.'
- *t'ęša* 'mother-in-law (wife's mother)' ← Russian *тёща* 'id.'
- *šurin* 'wife's brother' ← Russian *шурин* 'id.'
- *svest'a* 'wife's sister' ← Russian *свестья* 'id.'

4.  Husband's parents and siblings in Komi

- *ajka* 'father-in-law (husband's father)'
- *eńka* 'mother-in-law (husband's mother)'
- *piver* 'husband's brother'
- *ajaņiv* 'husband's sister'

It is not immediately obvious why this should be the case. In general, it is difficult to determine conclusively how certain features have come about in kinship terminology, especially as correlation does not necessarily mean causation. The fact that wife's relatives in Komi are borrowed from Russian might mean that Komi men having Russian spouses was what first introduced these terms to Komi, and intermarriage between Komi-speaking men and Russian-speaking women was more common than intermarriage between Komi-speaking women and Russian-speaking men. Any such analysis is, however, bound to be speculative in nature and also too simplistic to accurately describe the contact situation as a whole in any meaningful way. Whatever the explanation, the dichotomy is surprisingly clear-cut in Komi.

Whereas some of the languages, like Mari and Komi, have only one clear source for the borrowed kinship terms, Chuvash in the case of Mari

and Russian in the case of Komi, the situation in Mordvin is more het-
erogeneous. Russian loans represent only the most recent layer; before
that Mordvin was in contact with the Turkic languages of the Volga area,
(Mishar) Tatar and Chuvash, whence it also borrowed several kinship
terms. Different words for 'wife's brother' illustrate the influences and
overall situation quite well.

5.  Wife's brother in Mordvin

- WeB MdE *aľa*, M *aľǵä* (no convincing loan etymology)
- WyB MdE *baľźa*, M *baźä*, *päľźä*, *paźä* ← Mishar Tatar *baźa* 'Mann der
  Frauenschwester'
- WB MdE *šurin*, M *šuŕin*, *šuŕəń* ← Russian *шурин* 'id.'

The most recent of the loanwords is Russian *шурин* 'brother-in-law (wife's
brother)'. Like the Russian loan original, the word has no relative age
distinction in Mordvin, despite the fact that age distinction is otherwise
heavily featured in Mordvin affinal and consanguineal kinship terms
alike. The Russian loanword has not replaced the earlier terms for 'wife's
brother' that have relative age distinction, but it has superimposed itself
onto the existing system that consisted of WeB MdE *aľa*, M *aľǵä* and WyB
MdE *baľźa*, M *baźä*, *päľźä*, *paźä*, a loan from Mishar Tatar. We have no
information on how stable such a system is, whether the Russian term is,
for example, edging out the more specific terms or whether they still serve
a useful purpose not being full synonyms with the newest addition. It is,
in any event, interesting to note how clearly the different contacts are still
visible within the Mordvin kinship terminology.

It is unfortunately not possible to fully compare our numbers against
WOLD's borrowed scores, as they only include a single kin category 'sib-
ling-in-law' with a borrowed score of 0.12 (see Tadmor 2009 for further
details on how the borrowed scores were calculated in WOLD). Even as a
collapsed category, 'sibling-in-law' does not come anywhere close to be-
ing the kinship category with the largest number of loanwords among the
world's languages.[12]

---

12. In WOLD in the semantic field of kinship, the categories which have the high-
    est borrowed scores are 'family' (0.42), 'relatives' (0.40), and 'descendants'
    (0.35). Specific kin categories with the highest borrowed scores are 'uncle'
    (0.30), 'father's sister' (0.28), and 'grandmother' (0.27).

### 3.1.2. Borrowed terms for 'husband' and 'wife'

Among the categories where borrowing most often occurs, are the categories for 'husband' and 'wife'. It is perhaps interesting to ponder reasons as to why these particular categories have one of the largest number of loanwords. *A priori*, it seems reasonable to assume that the reasons are either 1) language-external social factors, i.e. explained by intermarriage between two linguistic groups or by otherwise close linguistic contact or 2) language-internal as in due to polysemy and semantic change etc. In the following, we will examine how many of the loanwords for 'husband' and 'wife' can be explained with language-internal factors in our material. The effects of semantic change have often been ignored in loanword typology.

The word for 'husband' in our material, whether borrowed or not, is rarely if ever strictly monosemic. Words denoting men often develop polysemy that 'husband' is a part of and the most common cross-linguistic semantic patterns, including synchronic polysemy and semantic change (perhaps better characterized as diachronic polysemy), are found in our material. We should first take a look at the most common tendencies concerning semantic shifts and borrowability in the world's languages.

In the *Database of Semantic Shifts in Languages of the World* (DatSem-Shift), 'husband' most commonly co-occurs with 'man' (43 languages). In the 20 cases for which the direction of the semantic shift is reported in DatSemShift, 'husband' always develops secondarily from 'man'. The borrowed scores of 'man' and 'husband' in WOLD are 0.10 for 'man' and 0.20 for 'husband'. To try to put the WOLD's borrowed scores into perspective, of the 41 languages, the word for 'husband' was reportedly either clearly or probably borrowed in 13 languages. The word 'man' was either clearly or probably borrowed in 5 languages. Considering that words meaning 'man' are two times less likely to be borrowed than the words meaning 'husband', the cross-linguistically common semantic shift from 'man' to 'husband' is probably not a very significant factor alone in explaining how 'husband' is one of the categories with the largest number of loanwords in the Uralic languages.

Another co-occurring polysemy according to DatSemShift that is also found in our material is that between 'husband' and 'old man'. In DatSemShift 'old man' and 'husband' are polysemous in 28 languages, with the meaning 'husband' developing from the meaning 'old man' in 15 languages. In general, 'husband' and 'old man' have very similar borrowed

scores with the latter closely edging out the former with 0.23 according to WOLD. The third quite clear polysemic pattern is between 'husband' and words denoting men of some kind of elevated social standing, i.e. 'master', 'owner', 'head of household', 'host', 'lord', etc. In DatSemShift there are 38 languages where this kind of polysemy occurs (examples can be found under 'owner'). The borrowed score of these words varies, but it is generally either equal with 'husband' (0.20), cf. 'host' (0.21) or even noticeably higher, cf. 'chieftain' (0.34), 'master' (0.38).

Perhaps unsurprisingly, 'wife' exhibits most of the same semantic patterns symmetrical with 'husband' in that the most common polysemy is between 'wife' and 'woman' (44 languages in DatSemShift). 'Wife' also shows a rather similar borrowed score of 0.18 in WOLD to 'husband's' 0.20. Words for 'woman' are borrowed slightly more often (0.16) than words for 'man' (0.10). Similarly to the polysemy of 'old man, husband', 'old woman, wife' is also a fairly typical meaning pair according to Dat-SemShift (15 languages). The borrowed score of 'old woman' in WOLD is 0.13. The third most common polysemic co-occurrence is between 'wife' and 'house, dwelling' (10 languages). This pattern, although interesting in itself, is not found in our material.[13] The fourth and fifth most common polysemy, again quite similarly to 'husband', is between 'wife' and 'mistress of a house' (5 languages) and 'owner' (4 languages).

In many cases, it is not easy to untangle the historical meanings from one another. It can, however, be established through etymological research that at least some of the terms for 'husband' and 'wife' in Kinura's materials were not initially kinship terms (at least four cases for 'husband' and three for 'wife'). They acquired their current meanings 'husband' and 'wife' only after having been borrowed, following some common cross-linguistic semantic shifts mentioned above (e.g. the meaning 'husband' being born out of 'man', 'old man', or 'head of the house'/'master'/'host'), and in this sense we are actually not dealing with the borrowability of kinship terms *per se* but rather with borrowing in general.

At least in four cases in our material 'husband' is clearly a secondary development brought on by semantic change. It perhaps comes as no surprise that these secondary developments follow more or less the same

---

13. According to DatSemShift, instances of polysemy between 'wife' and 'house, dwelling' are found in India, China, Korea, and Japan.

semantic patterns as listed above, the meaning 'husband' being born out of 'man', 'old man', or 'head of the house'/'master'/'host'.

In the first two cases the secondary development of 'husband' is from 'man' (in Mordvin) or from '(young) man' (in Mari). MdE *miŕďe*, M *miŕďä* has acquired its current meaning 'husband' through semantic change and at an earlier date the word probably was polysemous meaning 'man, husband' (or 'human; man, husband') considering that the loan original is nowadays usually identified as Pre-Indo-Iranian *\*mérto-* > Old Indic *márta-* 'mortal, human' (Holopainen 2019: 143–146). This interpretation is not without its difficulties, however.

The second case of secondary development of 'husband', also from '(young) man' in our material is in MariM *marij* 'Mari; man, husband', MariH *marə̂*. The word is used as an endonym by the Mari people themselves, so it seems reasonable to assume that the word initially referred to 'man' as ethnonyms developing from words meaning 'man, human' are a common occurrence unlike ethnonyms developing from words meaning 'husband'. The Mari word is usually thought to have been borrowed from an Iranian source representing a reflex of PI/PII *\*mar(H)ya-* > Young Avestan *mairiia* 'Schurke, Bube', Old Indic *márya-* 'Jungmann, Jüngling' (Holopainen 2019: 135–137), although given that the Mari vowel is atypical of Pre-Proto-Mari lexicon, it is unlikely that Proto-Indo-Iranian and Proto-Iranian could be the source. Rather, the word has entered the Mari language rather recently either directly or indirectly from an unidentified Iranian source.

The third case of secondary development of a term for 'husband' comes from Udmurt. This time the polysemy is between 'husband' and 'old man', cf. Udm *kart* 'муж, супруг' ← Ta *kart* 'старик' (Csúcs 1990: 207). No further examples of a shift from 'old man' to 'husband' can be found in Kinura's research materials. Referring to (one's) husband as 'old man' is noted to be somewhat widespread in Finnish dialects as well, cf. Fi *äijä* generally 'old man; geezer, gaffer', in western dialects also 'husband', *ukko* generally 'old man, gaffer', dialectally also 'husband', and *faija* both 'old man' and 'husband' in the Porvoo area, etc. (Nirvi 1952: 18–32). These are often used in a playful manner, but also partly as euphemisms brought on by affection. Words denoting 'old man' and especially 'old woman' have a tendency to undergo pejoration and it is usually dependent on the level of pejoration whether or not any given word denoting 'old man' or 'old woman' can be used to refer to one's husband or wife (op. cit.).

In the fourth case the meaning 'husband' developed secondarily from a meaning referring to a man with an elevated social status: North Saami *isit* 'head of the household; man, husband' is borrowed from Finnish *isäntä* 'master, lord (of a household); host' (Sammallahti 1998: 240). It is difficult to ascertain at what point exactly the word came to mean 'husband' in North Saami, as the word has been used for one's husband or father-in-law, usually in a way to convey a certain dignity, reverence, and distance in Finnish dialects whence the North Saami word was originally borrowed. Nowadays referring to one's husband as *isäntä* in Finnish is done more tongue-in-cheek.

'Wife' has also developed secondarily. 'Wife' and 'old woman' intertwine in the same way 'old man' and 'husband' do and often undergo pejoration as mentioned. At first glance, the idea that MariM *βate* 'wife', MariH *βätə* was borrowed from Chuvash *вамă* 'old, old (person)' (Räsänen 1920: 120) strikes one as semantically peculiar. Given that the polysemy 'old woman, wife' is commonplace (see above), the semantics become less of an issue.[14]

In DatSemShift, the fourth and fifth most commonly occurring polysemy is between 'wife' and 'mistress of a house' or 'owner'. There are at least two cases of borrowing in our material where 'wife' has clearly been borrowed from a word that primarily refers to a woman of elevated social standing, cf. SaaN *eamit* 'housewife; (female) owner; solitary woman who does her own cooking; wife' ← Finnic, Fi *emäntä* 'lady (of the house); housewife; hostess, matron; wife, spouse' (SSA 1: 104–105; Sammallahti 1998: 240) and MdE *koźejka*, *koźajka* 'wife' ← Russian *хозяйка* 'hostess', coll. 'wife' (Mészáros 2001: 174). In the case of North Saami *eamit*, it is yet again difficult to discern whether the meaning 'wife' developed independently after the borrowing or whether it was influenced by the similar use present in Finnish dialects or even borrowed as such.

In sum, the amount of loanwords in the categories 'husband' and 'wife' is at least partly explained by purely language-internal factors, i.e. by the fact that 'husband' and 'wife' often develop secondarily from a number of primary meanings such as 'man', 'old man', 'head of the household', 'host'

---

14. One might add that context matters here. Chuvash loanwords are a common occurrence in Mari, in fact Mari has borrowed roughly 10% of its vocabulary from Chuvash. If this was a more isolated loan etymology, we would perhaps be less convinced of its validity.

or 'woman', 'old woman', 'mistress', 'hostess'. Some of these primary meanings are also clearly more likely to be borrowed than 'husband' and 'wife' themselves.

It is worth pointing out here that simply looking at raw numbers and answering the question "is the word currently occupying this semantic slot a loan?" will yield a number that might not be all that representative of the borrowability of a given kinship term. For example, in our material it happens that 'sister' has been borrowed once or twice (PF *sisar/*sęsar ← Baltic), while 'younger sister' has been borrowed at least three times (MdE *sazor* 'younger sister', M *sazər*, MariM *šüžar* 'younger sister', MariH *šȃžar*, Udm *suzer* 'younger sister' are separate loanwords from (Indo-)Iranian *swasar-* 'sister'). The numbers in Uralic languages seem to stand in stark contrast to the borrowability rates one finds globally.

In WOLD 'sister' has a borrowed score of 0.12 while 'younger sister' only has a borrowed score of mere 0.01. Interestingly, however, words in Uralic languages in all likelihood came to mean 'younger sister' only secondarily. It has been argued that the words initially did not have age distinction and simply meant 'sister', and only developed their current meaning as a result of contact with the Turkic languages (Holopainen 2019: 224; Metsäranta 2023: 162–167). In other words, if we account for semantic change, the Uralic languages are actually not an anomaly when it comes to borrowability rates between 'sister' and 'younger sister' but rather conform to global tendencies. Ideally, it would be good to take the possibility of semantic change into account when examining the borrowability rates of all items but, unfortunately, this is often too laborious and open to interpretation in practice.

### 3.1.3. Borrowed terms for 'grandchild'

One of the most commonly borrowed kinship terms in our material is that for 'grandchild'. In many of the categories, borrowed kinship terms include different terms borrowed from various languages at different times; in the case of 'grandchild' the borrowability in our material is entirely due to the same, originally Slavic, kinship term being borrowed congruently into Uralic languages. Komi *vnuk*, MariM *unȃka*, H *ȃnȃka*, MdE *nuka*, M *unək* and Veps *vunuk* are borrowed from Russian *внук* '(male) grandchild'. Hungarian *unoka* is likewise a Slavic loanword, cf. Serbo-Croatian *unuk* 'grandchild' (EWUng: 1578). The Russian word has been very

expansive and has been borrowed into other non-Uralic languages as well, cf. Ta *оныk*, Chu *мăнук*. Alongside the masculine *внук*, Komi seems to be the only one that has also borrowed the feminine version from Russian, Komi *vnučka* '(female) grandchild' ← Ru *внучка*. In general, the word denotes grandchildren of both sexes in the target languages.

Despite the fact that most of the words have ultimately been borrowed from Slavic, it seems unlikely that this modern Russian word is straightforwardly the loan original for most of the Uralic words, although it is often stated as such: Russian *внук* → MdE *nuka*, M *unǝk* (Mészáros 2001: 171) and MariM *unǝka*, H *ǝnǝka* (Savatkova 1969: 92). Only the Komi word has quite clearly been borrowed from a form that is identical with the modern Russian word and as such can be a very late addition to the lexicon. The rest of the words are more open to interpretation. The Mordvin words quite clearly represent two separate loanwords, as the word-initial cluster *vn-* has been substituted in two different ways, in Erzya by simplification and in Moksha by vocalization of *v* to *u*. Vocalized forms, such as *unuk*, are found in various Russian dialects as well. Their distribution (Pskov, Kaluga, and Stavropol) does not seem to coincide closely enough with any of the vocalized forms on the Uralic side to have any direct connection.

The wide distribution in Mari dialects, as well as the vowel correspondences regularly reflecting Proto-Mari first-syllable *ŭ* (with the exception of NW *onoka*) (TschWb: 873), seem to indicate that at the time of the borrowing of this particular word, Mari was still a fairly uniform language, phonologically perhaps even identical with Proto-Mari. If Russian *внук* were borrowed separately into already diverged Mari dialects, one would expect to find more variation in the way in which *vn-* was substituted like we find in Mordvin between Erzya and Moksha. It should also be said that there seems to be no phonological objection to regarding the Proto-Mari *ŭnǝka* as an even earlier loanword similar to Old East Slavic *vŭnukŭ* (PM *w-* at least is irregularly dropped in certain native words as well, cf. PU *worka-* 'sew' > PM *ŭrge-*, PU *workama* > PM *wŭrgem* 'clothes'). Veps *vunuk* is also difficult to interpret. Perhaps the word was borrowed from Russian with the initial *vun-* being a substitution for *vn-*, or it was borrowed from an earlier form more similar to Old East Slavic. The final *-a* in Erzya and Mari might indicate that the words were borrowed from the genitive form *внука* instead of the nominative.

Why, then, has 'grandchild' been such an attractive term for borrowing? Questions of why are always difficult to answer definitively when it

comes to borrowing. One could perhaps claim that the fact that 'grand-child' specifically has been so prone to borrowing somehow reflects the reality in which these minority languages exist, and where language change to Russian has been a common occurrence, resulting in a situation where the grandparents speak the minority languages while their grandchildren are often Russian-speakers (see also Section 3.1.4 about the borrowing of a term for 'grandmother' from Russian). Although this might be partly true for the current situation, it would be rather anachronistic to project the current situation of these languages that far back into the past for it to be relevant for most of the loanwords in question. It seems that many of the languages that have borrowed the Russian word also have had an existing word for 'grandchild' that has not completely fallen out of use, cf. MdE *bujo* 'grandchild', Komi *peľen šuɨś* 'grandchild' (*peľ-en* grand-father-INSTRUMENTAL *šu-ɨś* call-PARTICIPLE.PRESENT, lit. "someone who calls someone else grandfather"). Thus, they seem to have had no real need to borrow the word and the concept could have been expressed some other way.

One explanation could be that the Russian word is just convenient. It is short and phonologically simple, or at least easily adaptable, and as such quite easily adopted even with limited proficiency in the language. It is however difficult or even impossible to assess such nebulous concepts as convenience, especially in a historical context. All in all, the borrowability of 'grandchild' is most likely due to a combination of sociolinguistic and practical factors.

### 3.1.4. Borrowed terms for 'grandmother'

A term for 'father's mother' (FM) is seemingly borrowed more often than a term for 'mother's mother' (MM), but on closer inspection this difference turns out to be nonexistent. According to Mészáros (2001: 170), which we have used as the source for our data on Mordvin, the word *baba* is in both Erzya and Moksha used only of 'father's mother', which is true for a number of Erzya and Moksha variants. However, in many other variants no such distinction is made and the word simply refers to 'grandmother' in general (MdWb: 108). Both Erzya and Moksha have terms which specifically refer to 'mother's mother', in Erzya *vasolbaba* or *mazibaba* (*vasolo* means 'distant, far away' and can be used as a way to specify relatives on the mother's side, cf. *vasolboďa* 'mother's father', *mazi*

'beautiful')[15] and in Moksha *ščava* (an (obscured) compound, the first component of which is cognate with E *čiče* 'older sister's husband' and the second with *ava* 'woman, wife; mother, mother-in-law').[16] In other words, *baba* can mean either 'father's or mother's mother' in both Erzya and Moksha, but if the distinction needs to be made, it is done to denote mother's mother. In Erzya it is done by qualifying the word *baba* with *vasol-* or *mazi-* and in Moksha by using an independent lexeme. Perhaps this is why Mészáros reports *baba* to mean 'father's mother', since as a standalone word it can refer primarily to 'father's mother' but also more broadly to 'grandmother'. In any event, we can conclude that despite the different numbers, there is no actual discrepancy in the borrowability of 'father's mother' and 'mother's mother' and the distinction we find in some Erzya and Moksha variants has come about secondarily.

Most of the Uralic languages in our data set do not make a distinction between FM and MM. This is the case, for instance, with SaaS *aahka* 'bestemor; gammel kvinne; (dial.) woman, wife' (< PS *ākkā* 'wife; old woman' ← Finnic, cf. Fi *akka* 'wife; old woman') and Finnish *mummo*, *mummu* 'grandmother; old woman' (← Swe dial. *mummu*, *mumm* = *mormor* 'mother's mother'), but also in more eastern languages discussed in more detail below.

As with 'grandchild', 'grandmother' also appears in our materials as one of the most borrowed categories largely due to the popularity of a single term borrowed from Russian (*баба* '(informal) old woman, (Lallwort) grandmother, (coll.) woman'): Hill Mari *papa* 'grandmother; old woman', Komi *bab* 'grandmother; old woman', MdE M *baba* with varying dialectal meanings 'old woman; wife; grandmother; father's mother; mother-in-law; father's uncle's wife; midwife', and Veps *bab* 'grandmother; old woman; midwife'. With the exception of Meadow Mari, the languages that have borrowed the term 'grandchild' from Russian have borrowed the term for

---

15. Interestingly, a similar expression exists in Meadow Mari, *koβa* 'grandmother', *mündôr koβa* 'mother's mother' (*mündôr* 'distant, far away'). This in all likelihood reflects the fact that a wedded couple tended to live with the husband's family away from at least the immediate vicinity of the mother's relatives, making them more "distant" geographically speaking.

16. Similarly to Moksha, reflexes of PU *čečä* are used to convey the meaning 'relative of mother's side' in compounds also elsewhere in Uralic, cf. MariH *papa* 'grandmother', *čəžə papa* 'mother's mother', Udm *čuž-anaj* (*anaj* 'mother') 'mother's mother', *čuž-murt* 'mother's brother' (*murt* 'human').

'grandmother' from Russian as well. The overlap and correlation is interesting, but as discussed in the previous section, causal links are difficult to establish.

'Grandmother' being featured among the kinship categories with the largest number of loanwords in the Uralic languages is in line with what has been found in earlier typological surveys with a global sample. In WOLD, 'grandmother' has a relatively high borrowed score of 0.27 and it is among the kin categories with the largest number of loanwords in the WOLD dataset (Honkola & Jordan, in press, Table 1). Words denoting 'grandmother' can change rather quickly with one term falling out of use and being replaced by another term regardless of whether they are borrowed or not. One possible explanation for this relatively high turnover is that words meaning 'grandmother' are often polysemous with 'old woman'. Words denoting women often can go through pejoration, as a result of which they are no longer suited to be used as kinship terms.

One of the Proto-Finnic terms for 'grandmother' was probably *ämmä*. Some of its reflexes still mean 'grandmother' in Finnish dialects and in Karelian. The semantics of Vote *ämmä* 'mother-in-law' and Est *ämm* 'id.' can be explained as secondary, as it can be reliably assumed that *anoppi* was the term for 'mother-in-law' in Proto-Finnic based on its wide distribution elsewhere in Uralic. The meaning of Livonian *āma* 'mother' (< PF *ämmä*) might be due to confusion with *jemā* 'mother' (< PF *emä* 'mother'). In standard Finnish *ämmä* is nowadays categorically derogatory, cf. *pahasuinen ämmä* 'foul-mouthed bitch' and consequently has fallen out of favor as a kinship term. The Finnish word for 'mother' *äiti* is a Germanic loanword (cf. PGrm *aiþīn ~ aiþōn* 'mother' > Gothic *aiþei* 'mother', ON *eiða* 'id.'). It has largely replaced the earlier PF *emä* 'mother' (> Est *ema* 'mother', Liv *jemā* 'id.') in this semantic slot in Finnish, although this is not the case throughout Finnish dialects where *emä* was a neutral term for mother, especially in the Karelian isthmus and neighboring areas (SMS: s.v. *emä*).

Pejoration has probably played a part in why *äiti* came to replace *emä*. It is unlikely that the meaning 'mother' could produce pejoration, rather the explanation for why *emä* in some Finnish dialects is negatively connotated and occurs as part of pejorative expressions is probably due to the word's polysemy, which covers both 'mother' and 'female animal', i.e. "animal mother". Words meaning 'female animal' are often used as insults or in an otherwise pejorative manner, cf. Russian *сука* 'bitch (female

dog); (vulgar) bitch (contemptible person)' (the Russian word is ultimately cognate with Swe *hund* 'dog', Latin *canis* 'id.', etc.). Similar usage where 'female dog' is used as an insult is found in many languages, cf. English *bitch*, Finnish *narttu*. Not all words denoting women are likely to undergo pejoration; it in large part depends on the polysemy of the word. It is unlikely that a word meaning only 'grandmother' could come to be used as a derogatory term on its own, rather any such use results from the polysemy of the word, most often from co-occurring meanings such as 'old woman'. If pejoration occurs, the resulting lexical gap can be filled with a loanword, but pejoration and borrowing are two independent processes.

## 3.2. Which languages have borrowed kin terms?

The Uralic languages vary notably in how many kin categories they have a borrowed kin term. The Mordvin languages and Hill Mari have the largest number of categories with loans (ca. 20 categories in each) while the Samoyedic languages Tundra Nenets, Taz Selkup, and Nganasan do not have any, and Forest Enets has only one (Table 3). The same general pattern holds also when we take into account the size of the kin category inventory in a language, that is, when calculating the percentage of kin categories with loanwords from all kin categories existing in the language in question (Table 3). Taking the inventory size into account is necessary, as the number of kin categories in a language may vary depending, for example, whether there is relative age distinction for sibling terms (i.e. instead of having a term only for 'sister' and 'brother' – two categories – there are terms for 'elder sister', 'younger sister', 'elder brother', and 'younger brother', so four categories).

On average 20.4% of kin categories have known loanwords in our sample of Uralic languages (with a median of 19.1%). This means that, on average, one fifth of kin categories in the Uralic languages has a loanword. In eight languages the percentage is as high as ca. thirty or more; these eight languages include languages from all the western subgroups of the Uralic family (Finnic: Veps 31.0%, Saami: North Saami 33.3% and South Saami 29.2%, Mordvin: Erzya 40.7% and Moksha 53.5%, Mari: Hill Mari 36.7%, and Meadow Mari 29.1%, Permic: Komi-Zyrian 39.4%; Table 3). The average percentage (20.4%) is higher than what was found with the global dataset of WOLD. In the calculations made by Tadmor (2009) the semantic field of kinship was found to have 15% of loanwords, whereas in Honkola & Jordan

Table 3: Frequency of borrowed kin terms in Uralic languages. # of kin categories with loanwords presents the raw count of categories with loanwords in each studied language. % of kin categories with loanwords shows the proportion of kin categories with loanwords from all kin categories documented in a language. The table has been sorted by values in the latter column.

|  | # of kin categories with loanwords | % of kin categories with loanwords |
|---|---|---|
| Moksha Mordvin | 23 | 53.5 |
| Erzya Mordvin | 22 | 40.7 |
| Komi-Zyrian | 15 | 39.4 |
| Hill Mari | 18 | 36.7 |
| North Saami | 14 | 33.3 |
| Veps | 9 | 31.0 |
| South Saami | 14 | 29.2 |
| Meadow Mari | 16 | 29.1 |
| Udmurt | 13 | 22.6 |
| Skolt Saami | 10 | 19.2 |
| Sosva Mansi | 7 | 18.9 |
| Finnish | 6 | 14.6 |
| Livonian | 5 | 13.5 |
| Hungarian | 5 | 12.8 |
| Estonian | 3 | 6.8 |
| Forest Enets | 1 | 4.5 |
| Kazym Khanty | 1 | 2.8 |
| Tundra Nenets | 0 | 0 |
| Taz Selkup | 0 | 0 |
| Nganasan | 0 | 0 |

(in press), who used a subset of the kin categories included in the WOLD data, the number of loanwords dropped to 13.6%. It is likely that the differences in kin categories included in these studies partly explain the differences in the percentages, but as the data-collection procedure and criteria of what counts as a loanword is the same, these numbers should roughly be comparable with each other.

In what follows we will be looking at the possible reasons why kinship terms have been so readily borrowed in some of the languages of our sample and, conversely, why in some of the languages in our sample there are so few borrowed kinship terms. These questions obviously are very much tied to the historical sociolinguistic situation, which we do not and cannot know in detail with any level of certainty. For that reason, much of what follows will be inherently speculative. We can make certain assumptions about the nature of contacts, for example, between Mari and Chuvash based on the amount and type of linguistic influence observed as a result of those contacts, but we cannot, for example, know what the percentage of medieval Mari people bilingual in Chuvash was. Comparisons are also drawn to the overall number of loanwords in a selection of languages in order to provide a wider overall perspective.

As mentioned earlier, the Mordvin languages, Erzya and Moksha, have the largest number of kin categories with loanwords (Table 3). The effect of borrowing on Mordvin kinship terminology has been a long, chronologically layered, and gradual process. In Mordvin, the majority of these loanwords are from Russian, but the Mordvin languages have also borrowed kin terms from (Mishar) Tatar and Chuvash. Mordvin also has a few loanwords that clearly predate what could feasibly be considered Mordvin and were likely borrowed already into Pre-Mordvin, i.e. a dialect of Proto-Uralic that would eventually give rise to Proto-Mordvin and the modern Mordvin languages. These loanwords include E *miŕďe* 'husband', M *miŕďä* 'id.' < Pre-Mordvin \*mertä ← PII \*m̥rtá- 'dead' or Pre-Indo-Iranian \*mérto- 'mortal, human' (Holopainen 2019: 143–146), E *sazor* 'younger sister', M *sazər* 'id.'< Pre-Mordvin \*sasar ← PII/PI \*swasar- 'sister' (Mészáros 2001: 172; Holopainen 2019: 222–224; Metsäranta 2023: 162–167) and E *ťejťeŕ* 'daughter', M *śťiŕ* < Pre-Mordvin \*tüktärə ← Baltic, cf. Lith *duktė* : *dukter-* (Mészáros 2001: 171; Metsäranta 2023: 167–172) (cf. also Appendix 2). All the different prehistorical and historical loanword layers, (Indo-)Iranian, Baltic, Chuvash, Tatar, and Russian (Bartens 1999: 13–19; Grünthal 2012: 307) are more or less featured also in borrowed kinship terminology, although for certain words to be counted as kinship terms is more a historical happenstance brought on by semantic change rather than anything else, cf. E *miŕďe*, M *miŕďä* (see Section 3.1.2).

The presence of kin terms among some of these earlier loanword layers is somewhat surprising. Independent early (Indo-)Iranian loans in Mordvin are few (7 to be exact) and typically denote either material and

immaterial culture: 'goat', 'god', 'gold', 'husk of grain', and 'iron' (Holopainen 2019: 339). Given the amount of words and their semantics, one could justifiably argue that Pre-Mordvin only had casual, perhaps even indirect, trade contacts with (Indo-)Iranian. The word denoting sister fits poorly into this scenario, as its presence would imply a familial relationship. Perhaps the remaining few (Indo-)Iranian words represent just the tip of the iceberg, i.e. perhaps the bulk of loanwords disappeared and the relationship between the Pre-Mordvin and (Indo-)Iranian was closer than the amount of loanwords suggests (although it will become clear in later paragraphs that the amount of loanwords is a poor metric in trying to define the sociolinguistic nature of contacts). It is also entirely possible that the context for borrowing was not familial, but rather the word which nowadays means 'sister' was earlier used in addressing or showing reverence, as such polysemy does occur in world's languages, e.g. Brahui *addī* 'sister; term of address to a woman' (DatSemShift).

Geography is probably the main answer as to why Mordvin has borrowed words from different sources at different times. It is thought that Mordvin and its linguistic predecessor have been spoken in the area between Oka and Sura rivers for thousands of years (Bartens 1999: 13). This location, close enough to the European steppes to expose the Mordvin speakers to different linguistic influences but far away enough to prevent them from getting drawn into the nomadic lifestyle of the steppes, has undoubtedly been a major contributor. Other contributing factors are less easy to identify.

The other major branch of Uralic that has borrowed its kinship terms *en masse* is Mari. For Mari, the most central lexifier has been the Chuvash language. If in Mordvin the process of borrowing has been more gradual in nature, in Mari the effect of borrowing has been more abrupt. As attested by a number of Middle Mongolian loanwords that have found their way into Mari via Chuvash, the contacts between Mari and Chuvash can hardly predate the Mongol conquest of Volga Bulgaria in 1236 (Bereczki 1994: 14–16).

In general, Chuvash loanwords are common throughout the Mari lexicon. The overall amount of Chuvash loanwords in Mari, calculated from *Tscheremissisches Wörterbuch* (TschWb), is around 500 (Saarinen 2010: 339). Chuvash loanwords have a wide distribution, which is usually thought to signify that they were borrowed before the disintegration of the Mari-speaking area starting in the late 16th century. We can therefore

state with some level of confidence that in the course of around 300 years, Mari borrowed around 10% of its basic vocabulary from Chuvash. In total the percentage of borrowed kinship terms is around 30% (Table 3), slightly higher in Hill Mari, and all the terms, with the exception of MariM *marij* 'Mari; man, husband', MariH *marâ* (< Iranian) and perhaps MariM *jeŋga* 'elder brother's wife', MariH *jeŋgä* (< Tatar), have been borrowed from Chuvash. If anything, borrowing is even more prevalent in kinship terms than elsewhere in the lexicon.

It is interesting to contrast the Chuvash numbers and situation against the Tatar loanwords in Mari. The amount of Tatar loanwords in TschWb is around 700 (op. cit.). There are significant dialectal differences, however, with the amount of loanwords varying from around 200 in Hill Mari to 2100 in the eastern dialects. The uneven distribution quite clearly means that the Tatar loanwords were, for the most part, borrowed only after the Mari people had already become geographically dispersed.

What the Tatar loanwords in Mari demonstrate especially when contrasted with Chuvash, is that the total amount of loanwords, the "raw numbers", is a poor predictor for the borrowing of kinship terms and instead the right type of contacts are needed for the transmission of kinship terms. What exactly was the deciding factor with the Mari contacts between Chuvash and Tatar that resulted in kinship terms being borrowed on a large scale in one but not in the other? One could hypothesize that the pervasiveness of Chuvash influence throughout the Mari lexicon, including kinship terminology, could not have happened without large-scale bilingualism and intermarriage between linguistic groups. This might be true as a general rule, but although bilingualism might be a predictor for the borrowing of kinship terms, it does not necessarily result in borrowing of kinship terms, as we can learn from the Forest Enets example below.

There are probably several historical, geographical, and sociolinguistic factors – and even reasons related to the history of science and availability of research materials – as to why there at least appears to be so few borrowed kinship terms in the Siberian Uralic branches Mansi, Khanty, and Samoyedic. The history of science reason we are alluding to here is that there has been far less etymological research done on Mansi, Khanty, and especially Samoyedic. Although the Samoyedic languages have come more into focus in historical phonology in the last four decades and etymological research has been carried out into their lexicon (Janhunen 1977; 1981; Sammallahti 1988; Aikio 2002; 2006), it is still heavily centered on

inherited vocabulary, i.e. identifying cognates for Samoyedic words in other branches of Uralic and elsewhere in Samoyedic. Language-specific etymology has been almost non-existent and none of the Samoyedic languages have their own etymological dictionaries. The lack of etymological research combined with the fact that lexical sources, i.e. dictionaries, are often modest in scope, at least partly explains why Samoyedic languages seemingly have so few borrowed kinship terms. For example, the dictionary we used for Nganasan, Kosterkina et al. (2001), has around seven thousand words, while dictionaries for many of the minority languages on the European side, North Saami, Mari, Komi, and Udmurt are in the thirty thousand to forty thousand range. This state of affairs obviously has an impact.

Mansi, Khanty, and the Samoyedic languages are spoken in relatively isolated and sparsely populated areas in northwestern Siberia. While these languages are not (necessarily) nowadays as geographically isolated as they were earlier, it must be borne in mind that e.g. Russian – a major source of kin terms for many of the Uralic languages on the European side – has come in contact with these languages only relatively recently, for example with Forest Enets only since the 1950s (Siegl 2013: 36). Thus, Russian has had the potential to be a significant lexifier for many of these languages only for a short while and the result of these contacts has often been rapid language shift to Russian rather than lexical borrowing.

Before intensive contacts with Russian started, Tundra Nenets was the dominant language of the region and Forest Enets speakers usually had good skills in Tundra Nenets. Intermarriage between Tundra Nenets and Forest Enets speakers was common and their progeny was brought up bilingually (op. cit.). Forest Enets has borrowed lexemes from Tundra Nenets, but despite widespread bilingualism, this has not resulted in an influx of loanwords at least in the realm of kinship terminology from one language to another. Additionally, it is interesting to note that even though Forest Enets speakers have also been in contact with Evenki and Dolgan speakers, these contacts have left no linguistic traces, because a Russian-based pidgin, Govorka, was used for interethnic communication with non-Samoyedic peoples (op. cit.). Nowadays, the situation is very different, however. After World War II the use of Forest Enets rapidly declined and language change has taken place, resulting in the younger generation of Forest Enets people being functionally monolingual in Russian (op. cit. 51–55). Thus, it is likely that Forest Enets simply has not had the

time necessary for Russian loans to be disseminated through its lexicon before the language change, and for that reason it does not have a significant amount of such loans.

From Mansi we have a few more examples of how long-lasting contacts and the number of loanwords in themselves are not reliable predictors for the presence of borrowed kinship terms in individual languages. The Mansi people have been in at least some kind of contact with Russians for some 1000 years[17] and even under their rule for 400 years. Despite the fact that there are at least 500 Russian loanwords spread across the Mansi dialects, only one of them is a kinship term (a term for 'father'). A Russian loan for 'father' is found in more than one dialect (Kálmán 1961: 16–24; 129), but not from the variety we collected kinship terms from (Sosva).

The situation between Mansi and Komi is rather similar to that between Mansi and Russian. A significant part of the Mansi populace still resided west of the Ural mountains until the 15th century, in an area between the rivers Kama and Chusovaya (Rédei 1970: 76–77). This is the earliest the Mansi could have been in contact with the speakers of Komi. Considering the fact that 85% of Komi loanwords in Mansi (338 in total) are found in northern Mansi (of which 138 exclusively there), it seems likely that most of the Komi loanwords were introduced into Mansi in Siberia. The majority of the loanwords were probably introduced into northern Mansi by Komi speakers who migrated there from west of the Urals in the course of the 18th and 19th centuries, with the epicenter in northwestern Siberia. All in all, the contacts between Mansi and Komi have lasted for centuries and Komi loanwords in Mansi number in the hundreds. Nevertheless, similarly to Russian loans in Mansi, the number of borrowed kin terms from Komi to Mansi is very low, as the only kinship term we find among them is MsSo *ōjka* 'man, husband, old man' ← Komi, cf. Komi-Permyak *ajka* 'man, husband', Komi-Jaźva *ajka* 'husband', Komi-Zyrian *ajka* 'husband's father' (Rédei 1970: 91). Russian and Komi loans in Mansi show that even otherwise extensive borrowing does not necessarily manifest itself in kinship terminology, i.e. not all kinds of contacts are conducive for the borrowing of kinship terminology.

---

17. Although this contact for several hundred years seems to have been mainly "mute fur trade" (Kálmán 1961: 18).

## 4. Conclusion and future directions

In this paper, we have examined the borrowability of kinship terms in Uralic languages. We found that the kinship categories with the most loanwords are affinal categories, such as those denoting spouse's siblings and sibling's spouses as well as husband and wife. The consanguineal category with the largest number of loanwords was 'grandchild', and terms denoting mother and father were among the most borrowed ones. The Uralic languages with the largest number of borrowed kinship terms were the Mordvin languages Erzya and Moksha, Komi-Zyrian, and Hill Mari. There were several Uralic languages such as Tundra Nenets and Nganasan with zero borrowed kinship terms.

We examined certain contact situations more closely in order to try to determine why some languages have borrowed kinship terms and why some have not. We have not attempted to answer only the question "is the word occupying category X a loanword in language Y?" but have looked at some of the intralinguistic reasons affecting each category such as semantic change as well as extralinguistic factors at play. Lexicon is a huge open-ended system within a system, and researching even a small sliver of it on the level of a whole language family is a tremendous undertaking. As the formation of any part of the lexicon is a long-lasting historical process that is always unique to a specific language, it is very difficult to make generalizations about it. Either the generalizations are too vague to be meaningful, or they are generally true on the macro level but not necessarily true for a specific language. This can be seen from our sample as well. There are languages that have borrowed a large part, half or more than one third, of their kinship terms, there are languages that have borrowed none and everything in between, but this can usually only be determined by taking a closer look at a specific language. We also found that neighboring, understood rather loosely here, languages often are quite similar when it comes to the borrowing of kinship terms.

It was noted earlier (see Section 1) that the most securely reconstructable kin terms in Uralic languages are terms for different in-laws, e.g. PU *ęppə* 'father-in-law', *mińä* 'daughter-in-law', *wäŋəw* 'son-in-law', etc. At first glance, this etymological observation might seem to be at odds with our general findings that kinship categories such as 'wife's brother', 'sister's husband', 'wife's sister', etc. were among the categories with largest number of borrowing events (Section 3). Given that for the reconstruction of

kinship terms (or any word for that matter), it is not necessary for them to survive in every single Uralic language or even every branch of Uralic, both statements can be true at the same time. As a result, in our language-family-wide study of kinship terms affinal categories had the largest number of loanwords while being more reconstructable for Proto-Uralic than consanguineal categories. This simply means that some languages (especially true for the branches at the far ends of the Uralic continuum, i.e. Saami, Finnic, Mansi, Khanty, and Samoyedic) have preserved the earlier terms while others have borrowed new ones that might or might not have replaced the earlier kinship terms.

The effect of language contact does not limit itself purely to loanwords. Due to the definition of loanword applied here, we had to exclude calques as well as loan blends (derivations and compounds) and phrasal expressions from our examination, although the individual parts they are composed of can be and often are borrowed. We include loanwords, calques, and loan blends in our study where we specifically focus on convergences in Circum-Baltic languages. In that paper we found that certain semantic borrowings actually have a larger spread than loanwords (Milanova et al. submitted manuscript). This emphasizes the need to include also this material when studying how borrowing has affected the kinship terminologies of the Uralic languages as a whole.

Borrowability of kinship terms is only one aspect of kinship terminology that can be studied with our dataset and with the Kinbank database (Passmore et al. 2023). Another aspect that can be studied and on which we currently focus, is to try to reconstruct the Proto-Uralic kinship system or at least parts thereof (Metsäranta et al. manuscript). Until now, Proto-Uralic kinship has mainly been studied from an etymological point of view with the focus on the reconstruction of the kinship terms themselves, while the structure (i.e. which kin categories exist separately and which are merged) has received very little attention. Proto-Uralic kinship can, however, also be examined from the point of view of different structural patterns that can persist even when the term itself is replaced due to borrowing.

As an example of an intriguing point that can be discussed with this approach is the case of terms for 'brother' and 'sister' in Proto-Uralic. By saying that these terms cannot be reconstructed for Proto-Uralic (Aikio 2022: 24), it is already implied that we should be able to reconstruct them. However, distinction of one's siblings by sex into 'brother' and 'sister' is

a distinctly European pattern, and fewer than 20 percent of the world's languages have this type of classification (Murdock 1968: 4). In the Uralic language family the westernmost languages have this European pattern, possibly as a result of borrowing, while in the eastern languages there is more variation including relative age distinction, e.g. 'elder brother/sister', 'younger brother/sister' or 'elder brother/sister' and 'younger sibling'. A culturally unbiased comparative treatment of Uralic kinship terms is needed, as it is not immediately obvious which of the sibling patterns observed in modern Uralic languages is what Proto-Uralic is expected to have had.

Exploring the historical development of kinship patterns is the current focus of our project, but it goes without saying that there is still plenty left to be explored in Uralic kinship terminology and its system within the scope of linguistics, but also with a multidisciplinary approach where genetic, archaeological, and anthropological evidence including residence and marriage patterns are taken into account. For example, studies on how kinship terminologies change both intrinsically and as a result of external contact are needed, and the ongoing change in the Saami system (see Section 1) provides an exceptional possibility to study the topic. While contact studies can give us insight into the (relative) time of contact, a multidisciplinary approach with archaeogenetic research can link the research with both time and space, such as the Proto-Finnic homeland, which has been a hotspot for prehistorical kin-term borrowing. In this way the research which started from individual kinship terms, can give us insight into the lives of the actual people that lived in the past.

## Acknowledgments

## Language abbreviations

| | | | |
|---|---|---|---|
| Bulg | Bulgarian | NenT | Tundra Nenets |
| Chu | Chuvash | ON | Old Norse |
| EnF | Forest Enets | PGrm | Proto-Germanic |
| Est | Estonian | PF | Proto-Finnic |
| Fi | Finnish | PI | Proto-Iranian |
| Hung | Hungarian | PII | Proto-Indo-Iranian |
| KhKaz | Kazym Khanty | PM | Proto-Mari |
| KhN | North Khanty | PN | Proto-Norse |
| Latv | Latvian | PS | Proto-Saami |
| Lith | Lithuanian | PSam | Proto-Samoyedic |
| Liv | Livonian | Ru | Russian |
| LPF | Late Proto-Finnic | SaaN | North Saami |
| MariH | Hill Mari | SaaS | South Saami |
| MariM | Meadow Mari | SaaSk | Skolt Saami |
| MdE | Erzya Mordvin | Scand | Scandinavian |
| MdM | Moksha Mordvin | Swe | Swedish |
| MPF | Middle Proto-Finnic | Ta | Tatar |
| MsN | North Mansi | Udm | Udmurt |
| MsSo | Sosva Mansi | | |

## Kin term abbreviations

| | | | |
|---|---|---|---|
| f | female (speaker) | HyZ | husband's younger sister |
| m | male (speaker) | HZ | husband's sister |
| | | M | mother |
| BW | brother's wife | MM | mother's mother |
| C | child | W | wife |
| CC | grandchild | WB | wife's brother |
| DH | daughter's husband | WeB | wife's elder brother |
| eB | elder brother | WeZ | wife's elder sister |
| eBW | elder brother's wife | WyB | wife's younger brother |
| eZ | elder sister | WyZ | wife's younger sister |
| eZH | elder sister's husband | WZ | wife's sister |
| F | father | yB | younger brother |
| FF | father's father | yBW | younger brother's wife |
| FM | father's mother | yZ | younger sister |
| H | husband | yZH | younger sister's husband |
| HB | husband's brother | ZD | sister's daughter |
| HeB | husband's elder brother | ZH | sister's husband |
| HeZ | husband's elder sister | ZS | sister's son |
| HyB | husband's younger brother | | |

## References

Ahlqvist, August. 1875. *Die Kulturwörter der Westfinnischen Sprachen: Ein Beitrag zu der Älteren Kulturgeschichte der Finnen*. Helsingfors: Verlag der Wassenius'schen Buchhandlung.

Aikio, Ante. 2002. New and old Samoyed etymologies. *Finnisch-Ugrische Forschungen* 57. 9–57.

Aikio, Ante. 2006. New and old Samoyed etymologies (Part 2). *Finnisch-Ugrische Forschungen* 59. 9–34.

Aikio, Ante. 2009. *The Saami loanwords in Finnish and Karelian*. Oulu: University of Oulu. (Doctoral dissertation.)

Aikio 2014 = Luobbal Sámmol Sámmol Ánte (Ante Aikio). 2014. The Uralic-Yukaghir lexical correspondences: Genetic inheritance, language contact or chance resemblance? *Finnisch-Ugrische Forschungen* 62. 7–76. https://doi.org/10.33339/fuf.86078

Aikio 2020 = Luobbal Sámmol Sámmol Ánte (Ante Aikio). 2020. Loanwords from unattested Nordic source forms in Saami. *Finnisch-Ugrische Forschungen* 65. 5–24. https://doi.org/10.33339/fuf.87404

Aikio 2022 = Luobbal Sámmol Sámmol Ánte (Ante Aikio). 2022. Proto-Uralic. In Bakró-Nagy, Marianne & Laakso, Johanna & Skribnik, Elena (eds.), *The Oxford guide to the Uralic languages*, 3–27. Oxford University Press. https://doi.org/10.1093/oso/9780198767664.003.0001

ALFE 2 = Tuomi et al. (eds.). 2007. *Atlas Linguarum Fennicarum*. Helsinki: Suomalaisen Kirjallisuuden Seura & Kotimaisten kielten tutkimuskeskus.

Bartens, Raija. 1999. *Mordvalaiskielten rakenne ja kehitys* (Mémoires de la Société Finno-Ougrienne 232). Helsinki: Suomalais-Ugrilainen Seura.

Bereczki, Gábor. 1994. *Grundzüge der tscheremissischen Sprachgeschichte* I (Studia Uralo-Altaica 35). Szeged.

Bergsland, Knut. 1942. Det samiske slektkaps- og svogerskapsordsystem. *Norsk tidsskrift for sprogvidenskap* 13. 148–198.

Bíró, Bernadette. 2004. A vogul rokonságnevek rendszere és morfológiai felépítése. *Nyelvtudományi Közlemények* 101. 32–56.

Csúcs, Sándor. 1990. *Die tatarischen Lehnwörter des Wotjakischen*. Budapest: Akadémiai Kiadó.

DatSemShift = Zalizniak, Anna A. et al. 2002–2022. *Database of semantic shifts in the languages of the world*. Moscow: Institute of Linguistics & Russian Academy of Sciences. https://datsemshift.ru/ (Accessed 2023-03-03.)

Derksen, Rick. 2008. *Etymological dictionary of the Slavic inherited lexicon* (Leiden Indo-European Etymological Dictionary Series Volume 4). Leiden & Boston: Brill.

Doerfer, Gerhard. 1988. *Grundwort und Sprachmischung: Eine Untersuchung an Hand von Körperteilbezeichnungen*. Stuttgart: Franz Steiner Verlag Wiesbaden GmbH.

DZIEBEL, GERMAN VALENTINOVICH. 2007. *The genius of kinship: The phenomenon of human kinship and the global diversity of kinship terminologies*. Youngstown: Cambria Press.

EVE = JUNTTILA, SANTERI & HOLOPAINEN, SAMPSA & KALLIO, PETRI & KUOKKALA, JUHA & PYSTYNEN, JUHO (eds.). 2018–. *Suomen vanhimman sanaston etymologinen verkkosanakirja* [Digital etymological dictionary of the oldest vocabulary of Finnish]. Helsinki. https://sanat.csc.fi/wiki/EVE-etusivu (Constantly updated. Accessed 2023-02-01.)

EWUng = BENKŐ, LORÁND & GERSTNER, KÁROLY & HÁMORI, ANTÓNIA S. & ZAICZ, GÁBOR (eds.). 1993–1997. *Etymologisches Wörterbuch des Ungarischen* I–III. Budapest: Akadémiai Kiadó.

FAINBERG, L. A. 1984. Kinship terminology of the Nganasan as a historical source. In Dragadze, Tamara (ed.), *Kinship and marriage in the Soviet Union*, 220–237. London: Routledge & Keegan Paul.

FEHLIG, BODO. 1981. Zur Altersklasseneinteilung in der lappischen Verwandtschaftsterminologie. *Finnisch-Ugrische Mitteilungen* 5. 61–73.

GODDARD, CLIFF & WIERZBICKA, ANNA. 2014. *Words and meanings: Lexical semantics across domains, languages, and cultures*. Oxford: Oxford University Press.

GRÜNTHAL, RIHO. 2012. Baltic loanwords in Mordvin. In Grünthal, Riho & Kallio, Petri (eds.), *A linguistic map of prehistoric Northern Europe* (Mémoires de la Société Finno-Ougrienne 266), 297–343. Helsinki: Suomalais-Ugrilainen Seura.

HARVA, UNO. 1939–1940. Der Bau des Verwandtschaftsnamensystems und die Verwandtschaftsverhältnisse bei den Fenno-Ugriern. *Finnisch-Ugrische Forschungen* 26. 91–120.

HASPELMATH, MARTIN. 2009. Lexical borrowing: concepts and issues. In Haspelmath, Martin & Tadmor, Uri (eds.), *Loanwords in the world's languages: A comparative handbook*, 35–54. Berlin: De Gruyter Mouton.

HASPELMATH, MARTIN & TADMOR, URI. 2009. The Loanword Typology Project and the World Loanword Database. In Haspelmath, Martin & Tadmor, Uri (eds.), *Loanwords in the world's languages: A comparative handbook*, 1–34. Berlin: De Gruyter Mouton.

HOLOPAINEN, SAMPSA. 2019. *Indo-Iranian borrowings in Uralic: Critical overview of sound substitutions and distribution criterion*. Helsinki: University of Helsinki. (Doctoral dissertation.)

HONKOLA, TERHI & JORDAN, FIONA M. In press. Kin term borrowings in the world's languages. *Journal of Language Contact*.

HONKOLA, TERHI & METSÄRANTA, NIKLAS & MILANOVA, VERONIKA & PASSMORE, SAM & JORDAN, FIONA M. 2022. Kinura: A database of kinship terminology from the Uralic language family. Zenodo. https://doi.org/10.5281/zenodo.6673055https://github.com/kinbank/kinura

ITKONEN, T. I. 1943. Karjalaiset ja Kuolan-Lappi. *Kalevalaseuran vuosikirja* 22. 40–55.

Itkonen, T. I. 1958 [2011]. *Koltan- ja kuolanlapin sanakirja = Wörterbuch des Kolta- und Kolalappischen* (Lexica Societatis Fenno-Ugricae XV). 3rd edition. Helsinki: Suomalais-Ugrilainen Seura.

Janhunen, Juha. 1977. *Samojedischer Wortschatz: Gemeinsamojedische Etymologien* (Castrenianumin toimitteita 17). Helsinki: Société Finno-Ougrienne.

Janhunen, Juha. 1981. Uralilaisen kantakielen sanastosta. *Journal de la Société Finno-Ougrienne* 77. 219–274.

Joki, Aulis J. 1973. *Uralier und Indogermaner* (Mémoires de la Société Finno-Ougrienne 151). Helsinki: Société Finno-Ougrienne.

Junttila, Santeri. 2015. *Tiedon kumuloituminen ja trendit lainasanatutkimuksessa: Kantasuomen balttilaislainojen tutkimushistoriaa.* Helsinki: University of Helsinki. (Doctoral dissertation.)

Juutinen, Markus. 2022. List of Russian loan etymologies in Skolt Saami. Online appendix. Main article: Juutinen, Markus. 2022. Russian loanwords in Skolt Saami. *Finnisch-Ugrische Forschungen* 67. 75–126.
https://doi.org/10.33339/fuf.110737

Kalima, Jalo. 1910. *Die Russischen Lehnwörter im Syrjänischen* (Mémoires de la Société Finno-Ougrienne XIXX). Helsinki: Société Finno-Ougrienne.

Kallio, Petri 2018. Ensitavun diftongit kantasuomessa. In Holopainen, Sampsa & Saarikivi, Janne (eds.), *Perì orthótētos etýmōn (Περὶ ὀρθότητος ἐτύμων): Uusiutuva uralilainen etymologia* (Uralica Helsingiensia 11), 251–268. Helsinki: Suomalais-Ugrilainen Seura.

Kálmán, Béla. 1961. *Die russischen Lehnwörter im Wogulischen.* Budapest: Akadémiai Kiadó.

Karjalainen, K. F. 1913. Wie EGO im ostjakischen die verwandten benennt. *Finnisch-Ugrische Forschungen* 13. 207–295.

Kejonen, Olle. 2020. On the development of some Saami kinship terms. In Blokland, Rogier & Valijärvi, Riitta-Liisa (eds.), *Där Östersjön är Västersjön = Seal, kus Läänemeri on Idameri: Festskrift till Virve och Raimo Raag = Pühendusteos Virve ja Raimo Raagile*, 43–53. Uppsala universitet.

KESKJa = Лыткин, В. И. & Гуляев, Е. С. 1999. *Краткий этимологический словарь коми языка.* Переиздание с дополнением. Сыктывкар: Коми книжное издательство.

Koivulehto, Jorma. 1999. Varhaiset indoeurooppalaiskontaktit: Aika ja paikka lainasanojen valossa. In Fogelberg, Paul (ed.), *Pohjan poluilla: Suomalaisten juuret nykytutkimuksen mukaan* (Bidrag till kännedom av Finlands natur och folk 153), 207–236. Helsinki: Finska Vetenskap-Societeten = Suomen tiedeseura.

Korenchy, Éva. 1972. *Iranische Lehnwörter in den obugrischen Sprachen.* Budapest: Akadémiai Kiadó.

Korhonen, Mikko. 1977. Karjalaisten ja saamelaisten yhteydet kielentutkimuksen valossa. In Saloheimo, Veijo (ed.), *Karjalan synty: Symposio 30.6.–2.7.1976 Joensuu, Alustukset*, 75–82. Joensuu: Joensuun korkeakoulu.

Kosterkina et al. 2001 = Костеркина, Н. Т. & Момде, А. Ч. & Жданова, Т. Ю. 2001. *Словарь нганасанско-русский и русско-нганасанский.* Санкт-Петербург: Филиал издательства Просвещение.

Kroonen, Guus. 2013. *Etymological dictionary of Proto-Germanic* (Leiden Indo-European Etymological Dictionary Series 2). Leiden & Boston: Brill.

Kümmel, Martin Joachim. 2015. Die Schwestern der Eltern: Samische Evidenz für germanische Wortbildung? In Krisch, Thomas & Niederreiter, Stefan & Crombach, Michael (eds.), *Diachronie und Sprachvergleich: Beiträge aus der Arbeitsgruppe „historisch-vergleichende Sprachwissenschaft" bei der 40, Österreichischen Linguistiktagung 2013 in Salzburg* (Innsbrucker Beiträge zur Sprachwissenschaft 150), 121–129. Innsbruck: Institut für Sprachen und Literaturen der Universität Innsbruck.

Kuokkala, Juha. 2018. Finnic-Saamic labial vowels in non-initial syllables: An etymological evaluation. In Holopainen, Sampsa & Saarikivi, Janne (eds.), *Perì orthótētos etýmōn (Περὶ ὀρθότητος ἐτύμων): Uusiutuva uralilainen etymologia* (Uralica Helsingiensia 11), 11–74. Helsinki: Suomalais-Ugrilainen Seura.

LägLoS = Kylstra, A. D. & Hahmo, Sirkka-Liisa & Hofstra, Tette & Nikkilä, Osmo (eds.). 1991–2012. *Lexikon der älteren germanischen Lehnwörter in den ostseefinnischen Sprachen* I–III. Amsterdam: Rodopi.

LW = Kettunen, Lauri. 1938. *Livisches Wörterbuch mit grammatischen Einleitung* (Lexica Societatis Fenno-Ugricae V). Helsinki: Suomalais-Ugrilainen Seura.

Matasović, Ranko. 2014. *Slavic nominal word-formation: Proto-Indo-European origins and historical development.* Heidelberg: Universitätsverlag Winter.

Matras, Yaron. 2009. *Language contact.* Cambridge: Cambridge University Press.

Matras, Yaron. 2010. Contact, convergence and typology. In Hickey, Raymond (ed.), *The handbook of language contact*, 66–85. Chichester: Wiley.

MdWb = Heikkilä, Kaino & Kahla, Martti (eds.). 1990–1999. *H. Paasonens Mordwinisches Wörterbuch* (Lexica Societatis Fenno-Ugricae XXIII). Helsinki: Suomalais-Ugrilainen Seura.

Mészáros, Edit. 2001. A mordvin rokonságnevek eredet szerinti vizsgálata. *Acta Universitatis Szegediensis* 41. 169–184.

Metsäranta, Niklas. 2023. 'Sister', 'daughter' and 'brother': Etymological discussion of a few borrowed Uralic kinship terms. *Linguistica Uralica* 59. 161–180. https://dx.doi.org/10.3176/lu.2023.3.01

Metsäranta, Niklas & Milanova, Veronika & Honkola, Terhi. Manuscript. *Reconstruction of Proto-Uralic kinship system.*

Milanova, Veronika. 2020. *Kinship and affinity in Indo-European: Universal, inherited, contact, and area specific features of Indo-European kinship terms (with a special focus on the Iranian branch).* Vienna: University of Vienna. (Doctoral dissertation.) https://doi.org/10.25365/thesis.65526

Milanova, Veronika & Holopainen, Sampsa & Bradley, Jeremy. 2020. Contact phenomena in Indo-European kinship and social terms and beyond (a pilot study with special focus on the Iranian and Uralic languages in the context of Central Eurasia). In Garnier, Romain (ed.), *Loanwords and substrata: Proceedings of the Colloquium held in Limoges (5th–7th June, 2018)*, 345–366. Innsbruck: Innsbrucker Beiträge zur Sprachwissenschaft.

Milanova, Veronika & Metsäranta, Niklas & Honkola, Terhi. Submitted manuscript. *Kinship terminologies of the Circum-Baltic linguistic area: convergences and area-specific structural properties.*

Munkácsi, Bernát & Kálmán, Béla. 1986. *Wogulisches Wörterbuch.* Gesammelt von Bernát Munkácsi, geordnet, bearbeit und herausgegeben von Béla Kálmán. Akadémiai Kiadó: Budapest.

Murdock, G.P. 1968. Patterns of sibling terminology. *Ethnology* 7. 1–24.

Næss, Marius Warg & Fisktjønmo, Guro Lovise Hole & Bårdsen, Bård-Jørgen. 2021. The Sami cooperative herding group: the siida system from past to present. *Acta Borealia* 38 (2). 81–103.
https://doi.org/10.1080/08003831.2021.1972265

Nirvi, R. E. 1952. *Synonyymitutkimuksia sukulaisnimistön alalta.* Helsinki: Suomalaisen Kirjallisuuden Seura.

Paasonen, Heikki. 1897. Die türkischen Lehnwörter im Mordwinischen. *Journal de la Société Finno-Ougrienne* XV. 3–64.

Passmore, Sam & Barth, Wolfgang & Greenhill, Simon & Quinn, Kyla & Sheard, Catherine & Argyriou, Paraskevi & Birchall, Joshua & Bowern, Claire & Calladine, Jasmine & Deb, Angarika & Diederen, Anouk & Metsäranta, Niklas P. & Araujo, Luis Henrique & Hickey-Hall, Jo & Honkola, Terhi & Mitchell, Alice & Poole, Lucy & Rácz, Péter & Roberts, Seán G. & Ross, Robert M. & Thomas-Colquhoun, Ewan & Evans, Nicholas & Jordan, Fiona M. 2023. Kinbank: A global database of kinship terminology. *PLOS ONE* 18(5): e0283218.
https://doi.org/10.1371/journal.pone.0283218

Pelto, Pertti J. 1962. *Individualism in Skolt Lapp society* (Kansatieteellinen arkisto 16). Helsinki: Suomen muinaismuistoyhdistys.

Qvigstad, J. K. 1893. *Nordische Lehnwörter im Lappischen.* Christiania: Jacob Dybwad.

Räsänen, Martti. 1920. *Die tschuwassischen Lehnwörter im Tscheremissischen* (Mémoires de la Société Finno-Ougrienne 48). Helsinki: Société Finno-Ougrienne.

Räsänen, Martti. 1923. *Die tatarischen Lehnwörter im Tschremissischen* (Mémoires de la Société Finno-Ougrienne 50). Helsinki: Société Finno-Ougrienne.

Rédei, Károly. 1970. *Die syrjänischen Lehnwörter im Wogulischen.* Bloomington: Indiana University & The Hague: Mouton & Co.

Saarinen, Sirkka. 2010. Marin sanaston alkuperästä. In Saarinen, Sirkka & Siitonen, Kirsti & Vaittinen, Tanja (eds.), *Sanoista kirjakieliin: Juhlakirja Kaisa Häkkiselle 17. marraskuuta 2010* (Mémoires de la Société Finno-Ougrienne 259), 335–341. Helsinki: Suomalais-Ugrilainen Seura.

Sammallahti, Pekka. 1988. Historical phonology of the Uralic languages: With special reference to Samoyed, Ugric and Permic. In Sinor, Denis (ed.), *The Uralic languages: Description, history and foreign influences*, 478–554. Leiden: Brill.

Sammallahti, Pekka. 1998. *The Saami languages: An introduction.* Kárášjohka: Davvi Girji.

Savatkova 1969 = Саваткова, А. А. 1969. *Русские заимствования в марийское языке*. Йошкар-Ола: Марийское книжное издательство.

Setälä, E. N. 1900. I. N. Smirnow's Untersuchung über die Ostfinnen. *Journal de la Société Finno-Ougrienne* 17. 17–28.

Siegl, Florian. 2013. *Materials on Forest Enets, an indigenous language of Northern Siberia* (Mémoires de la Société Finno-Ougrienne 267). Helsinki: Société Finno-Ougrienne.

Simčenko 1974 = Симченко. Ю. Б. 1974. Терминология родства ненцев, энцев, нганасан и юкагиров [Kinship terminology of the Nenets, Enets, Nganasans, and Yukaghirs]. In Гурвич, И. С. (ed.), *Социальная организация и культура народов Севера* [Social organisation and cultures of the Arctic peoples.], 270–291. Moscow: Nauka.

SKES = Toivonen, Y. H. & Joki, Aulis J. & Itkonen, Erkki (eds.). 1955–1982. *Suomen kielen etymologinen sanakirja* (Lexica Societatis Fenno-Ugricae XII). Helsinki: Suomalais-Ugrilainen Seura.

Smirnov 1889 = Смирновъ, Иван Н. 1889. *Черемисы: Историко-этнографическій очеркъ*. Казань: Типографія Императорскаго Университета.

Smirnov 1891 = Смирновъ, Иван Н. 1891. *Пермяки: Историко-этнографическій очеркъ*. Казань: Типографія Императорскаго Университета.

Smirnov 1895 = Смирновъ, Иван Н. 1895. *Мордва: Историко-этнографическій очеркъ*. Казань: Типографія Императорскаго Университета.

Sokolova 1974 = Соколова, З. П. 1974. Социальная организация обских угров [Social organization of the Ob-Ugrians]. In Гурвич, И. С. (ed.), *Социальная организация и культура народов Севера* [Social organization and cultures of the Arctic peoples], 82–132. Москва: Наука.

SSA = Itkonen, Erkki & Kulonen, Ulla-Maija (eds.). 1992–2000. *Suomen sanojen alkuperä: Etymologinen sanakirja* (Suomalaisen Kirjallisuuden Seuran Toimituksia 556 & Kotimaisten kielten tutkimuskeskuksen julkaisuja 62). Helsinki: Suomalaisen Kirjallisuuden Seura & Kotimaisten kielten tutkimuskeskus.

Székely, Gábor. 2016. Rokonsági és családterminusok a magyar nyelvben [Kinship and family terms in Hungarian]. In Hidasi, Judit & Osváth, Gábor & Székely, Gábor (eds.), *Család és rokonság nyelvek tükrében* (Segédkönyvek a nyelvészet tanulmányozásához 186), 19–57. Budapest: Tinta Könyvkiadó.

Szíj, Enikő. 1979. A permi rokonságnevek rendszeréről, I: A kérdőíves anyaggyűjtés eredményei. *Nyelvtudományi Közlemények* 81. 247–272.

Szíj, Enikő. 1998. *Термины родства и свойства в удмуртском языке* [Terms of kinship and affinity in the Udmurt language]. Budapest: Akadémiai Kiadó.

Tadmor, Uri. 2009. Loanwords in the world's languages: Findings and results. In Haspelmath, Martin & Tadmor, Uri (eds.), *Loanwords in the world's languages: A comparative handbook*, 55–75. Berlin: de Gruyter Mouton.

Thomsen, Vilhelm. 1870. *Über den einfluss der germanischen sprachen auf die finnisch-lappischen: Eine sprachgeschichtliche untersuchung*. Halle: Verlag der Buchhandlung des Waisenhauses.

Trautmann, Thomas R. 2001. The whole history of kinship terminology in three chapters: Before Morgan, Morgan, and after Morgan. *Anthropological Theory* 1. 268–287.

TschWb = Moisio, Arto & Saarinen, Sirkka (eds.). 2008. *Tscheremissisches Wörterbuch*. Aufgezeichnet von Volmari Porkka, Arvid Genetz, Yrjö Wichmann, Martti Räsänen, T. E. Uotila und Erkki Itkonen (Lexica Societatis Fenno-Ugricae XXXII). Helsinki: Suomalais-Ugrilainen Seura & Kotimaisten kielten tutkimuskeskus.

UED = Luobbal Sámmol Sámmol Ánte (Aikio, Ante). 2020. *Uralic etymological dictionary*. Draft version of entries A–Ć published online on Academia.edu.

UEW = Rédei, Károly. 1986–1991. *Uralisches Etymologisches Wörterbuch*. Band I–II: Unter Mitarbeit von Marianne Bakró-Nagy, Sándor Csúcs, István Erdélyi †, László Honti, Éva Korenchy †, Éva K. Sal und Edit Vértes, Band III: Register Zusammengestellt von Attila Dobó und Éva Fancsaly. Wiesbaden: Harrassowitz.

Vavra, Klára. 1965. Principles of classifying the names of relationship in Mansi. *Советское финно-угроведение* I (2). 107–115.

Veršinin 2009 = Вершинин, В. И. 2009. *Этимологический словарь мордовских (эрзянского и мокшанского) языков*, Том IV. Йошкар-Ола.

VKS = *Vanhan kirjasuomen sanakirja* (Kotimaisten kielten keskuksen verkkojulkaisuja 38). 2022. Helsinki: Kotimaisten kielten keskus. (Updating publication. Last updated 25.11.2022.) https://kaino.kotus.fi/vks/ (Accessed 2023-09-02.)

Volzhanina, Elena A. 2011. The Interpretation of Nenets demography in the first third of the twentieth century. In David G. Anderson (ed.), *The 1926/27 Soviet Polar Census Expeditions*, 67–96. Oxford: Berghahn Books.

Whitaker, Ian. 1955. *Social relations in a nomadic Lappish community*. Oslo: Norsk Folkemuseum.

Whitaker, Ian. 1979. The kinship and affinal terminology of the Karesuando Lapps. *Finnisch-Ugrische Forschungen* 43. 80–110.

Wichmann, Yrjö. 1903. *Die tschuwassischen Lehnwörter in den permischen Sprachen* (Mémoires de la Société Finno-Ougrienne 21). Helsinki: Société Finno-Ougrienne.

Wierzbicka, Anna. 2016. Back to 'mother' and 'father': Overcoming the Eurocentrism of kinship studies through eight lexical universals. *Current Anthropology* 57 (4). 408–429. http://dx.doi.org/10.1086/687360

Winkler, Eberhard. 2014. Loanword strata in Livonian. *Eesti ja soome-ugri keeleteaduse ajakiri* 5(1). 215–227. http://dx.doi.org/10.12697/jeful.2014.5.1.11

WogWb = Moisio, Arto (ed.). 2013. *Wogulisches Wörterbuch*. Gesammelt und geordnet von Artturi Kannisto, bearbeitet von Vuokko Eiras, herausgegeben von Arto Moisio (Lexica Societatis Fenno-Ugricae XXXV & Kotimaisten kielten tutkimuskeskuksen julkaisuja 173). Helsinki: Société Finno-Ougrienne & Kotimaisten kielten keskus.

WOLD = Haspelmath, Martin & Tadmor, Uri (eds.). 2009–2022. *World loanword database*. Leipzig: Max Planck Institute for Evolutionary Anthropology. https://doi.org/10.5281/zenodo.11137 (Accessed 2022-08-11.)

WOT = Róna-Tas, András & Berta, Árpád (with the assistance of László Károly). 2011. *West Old Turkic: Turkic loanwords in Hungarian*, I–II (Turcologica 84). Wiesbaden: Harrassowitz.

WotjWsch = Korhonen, Mikko (ed.). 1987. *Wotjakischer Wortschatz*. Aufgezeichnet von Yrjö Wichmann, bearbeitet von T. E. Uotila und Mikko Korhonen, herausgegeben von Mikko Korhonen (Lexica Societatis Fenno-Ugricae XXI). Helsinki: Suomalais-Ugrilainen Seura.

## Lexical sources

Andrianova et al. 2019 = Андрианова, Е. М. & Гаврилова, В. Г. & Гусева, Э. В. и др. 2019. *Русско-марийский словарь*, Том I: А–Ою. Йошкар-Ола: МарНИИЯЛИ.

Barceva et al. 2019 = Барцева, Л. И. & Гаврилова, В. Г. & Ипакова, М. Т. и др. 2019. *Русско-марийский словарь*, Том II: П–Я. Йошкар-Ола: МарНИИЯЛИ.

Bergsland, Knut & Magga, Lajla Mattsson. 1993. *Åarjelsaemien-daaroen baakoegærja = Sydsamisk-norsk ordbok*. Alta: Idut.

Besnosikova et al. 2000 = Бесносикова, А. М. & Айбабина, Е. А. & Коснырева, Р. И. 2000. *Коми-роч кывчукӧр = коми-русский словарь*. Сыктывкар: Коми книжное издательство.

DeWoS = Steinitz, Wolfgang. 1966–1993. *Dialektologisches und etymologisches Wörterbuch der ostjakischen Sprache*. Berlin: Akademie-Verlag.

*Eesti keele seletav sõnaraamat*. Eesti Keele Instituut. (Online dictionary.) https://www.eki.ee/dict/ekss/ (Accessed 2022-08-14.)

*Eesti-soome sõnaraamat*. Eesti Keele Instituut. (Online dictionary.) https://www.eki.ee/dict/efi/ (Accessed 2022-08-14.)

ETY = Metsmägi et al. (eds). 2012. *Eesti etümoloogiasõnaraamat*. Eesti Keele Instituut. (Online dictionary.) https://www.eki.ee/dict/ety/ (Accessed 2023-02-01.)

Galkin, I. S. (Галкин, И. С.) et al. (eds.) 1990–2005. *Словарь марийского языка* 1–10. Йошкар-Ола: Марийское книжное издательство.

Helimski 2007 = Хелимский, Е. 2007. *Северноселькупский словарь*. (Unpublished manuscript.)

Karjalainen, K. F. 1913. Wie EGO im ostjakischen die verwandten benennt. *Finnisch-Ugrische Forschungen* 13. 207–295.

*Kielitoimiston sanakirja*. 2022–. Kotimaisten kielten keskus & Kielikone Oy. (Online dictionary.) https://www.kielitoimistonsanakirja.fi/#/ (Accessed 2022-08-14.)

Kirillova 2008 = Кириллова, Л. Е. 2008. *Удмурт-ӟуч кыллюкам = Удмуртско-русский словарь*. Ижевск.

Kosterkina et al. 2001 = Костеркина, Н. Т. & Момде, А. Ч. & Жданова, Т. Ю. 2001. *Словарь нганасанско-русский и русско-нганасанский*. Санкт-Петербург: Филиал издательства Просвещение.

*Livonian-Estonian-Latvian dictionary*. UL Livonian Dictionary. (Online Dictionary.) http://www.livones.net/lili/en/vardnica/ (Accessed 2022-08-14.)

LW = Kettunen, Lauri. 1938. *Livisches Wörterbuch mit grammatischen Einleitung* (Lexica Societatis Fenno-Ugricae V). Helsinki: Suomalais-Ugrilainen Seura.

Mészáros, Edit. 2001. A mordvin rokonságnevek eredet szerinti vizsgálata. *Acta Universitatis Szegediensis* 41. 169–184.

Moshnikoff, Jouni & Moshnikoff, Satu. 2022. *Suomi–koltansaame-sanakirja = Lää'dd-sää'm sää'nnǩe'rjj.* Lehtinen, Miika & Koponen, Eino & Fofonoff, Merja & Lehtola, Raija (eds.). Aanar: Sää'mte'ǧǧ = Saamelaiskäräjät.

Papp, István & László, Jakab. 1993. *Magyar-finn szótar.* Második kiadás. Budapest: Akadémiai Kiadó.

Passmore, Sam & Barth, Wolfgang & Greenhill, Simon & Quinn, Kyla & Sheard, Catherine & Argyriou, Paraskevi & Birchall, Joshua & Bowern, Claire & Calladine, Jasmine & Deb, Angarika & Diederen, Anouk & Metsäranta, Niklas P. & Araujo, Luis Henrique & Hickey-Hall, Jo & Honkola, Terhi & Mitchell, Alice & Poole, Lucy & Rácz, Péter & Roberts, Seán G. & Ross, Robert M. & Thomas-Colquhoun, Ewan & Evans, Nicholas & Jordan, Fiona M. 2023. Kinbank: A global database of kinship terminology. *PLOS ONE* 18(5): e0283218. https://doi.org/10.1371/journal.pone.0283218

Pelto, Pertti J. 1962. *Individualism in Skolt Lapp society* (Kansatieteellinen arkisto 16). Helsinki: Suomen muinaismuistoyhdistys.

*Russko-gornomarijskij slovar' = Русско-горномарийский словарь.* FU-Lab. (Online dictionary.) https://dict.fulab.ru/dict?id=253705 (Accessed 2022-08-14.) Compiled from three dictionaries: 1) Юадаров, К. Г. 1979. *Русско-марийский словарь для учащихся горномарийских средних школ.* Йошкар-Ола: Марийское книжное издательство. 2) Дмитриев, С. Д. 2013. *Русско-марийский словарь для школьников.* Йошкар-Ола: Марийское книжное издательство. 3) Васикова, Л. П. 1994. *Кырык марла орфографи лӹмдер. Орфографический словарь горномарийского языка.* Йошкар-Ола: Республиканское газетно-журнальное издательство «Периодика Марий Эл».

Sammallahti, Pekka. 1989. *Sámi-suoma sátnegirji = Saamelais-suomalainen sanakirja.* Ohcejohka: Jorgaleaddji Oy.

Savatkova 2008 = Саваткова, А. А. 2008. *Словарь горномарийского языка.* 2nd ed. Йошкар-Ола: Марийское книжное издательство.

Sorokina & Bolina 2009 = Сорокина, И. П. & Болина, Д. С. 2009. *Энецкий словарь с кратким грамматическим очерком.* Санкт-Петербург.

Steinitz, Wolfgang. 1966–1993. *Dialektologisches und etymologisches Wörterbuch der ostjakischen Sprache.* Berlin: Akademie-Verlag.

Tereščenko 1965 = Терещенко, Н. М. 1965. *Ненецко-русский словарь.* Москва: Советская Энциклопедия.

TschWb = Moisio, Arto & Saarinen, Sirkka (eds.). 2008. *Tscheremissisches Wörterbuch.* Aufgezeichnet von Volmari Porkka, Arvid Genetz, Yrjö Wichmann, Martti Räsänen, T. E. Uotila und Erkki Itkonen (Lexica Societatis Fenno-Ugricae XXXII & Kotimaisten kielten tutkimuskeskuksen julkaisuja 151). Helsinki: Suomalais-Ugrilainen Seura & Kotimaisten kielten tutkimuskeskus.

WogWb = Moisio, Arto (ed.). 2013. *Wogulisches Wörterbuch*. Gesammelt und geordnet von Artturi Kannisto, bearbeitet von Vuokko Eiras, herausgegeben von Arto Moisio (Lexica Societatis Fenno-Ugricae XXXV & Kotimaisten kielten tutkimuskeskuksen julkaisuja 173). Helsinki: Société Finno-Ougrienne & Kotimaisten kielten keskus.

WotjWsch = Korhonen, Mikko (ed.). 1987. *Wotjakischer Wortschatz*. Aufgezeichnet von Yrjö Wichmann, bearbeitet von T. E. Uotila und Mikko Korhonen, herausgegeben von Mikko Korhonen (Lexica Societatis Fenno-Ugricae XXI). Helsinki: Suomalais-Ugrilainen Seura.

Zaiceva, Nina. 2010. *Uz' vepsä-venäläine vajehnik*. Petroskoi: Periodika.

Zaiceva, Nina & Mullonen, Irma. 2007. *Uz' venä-vepsläine vajehnik*. Petroskoi: Periodika.

## Appendix 1: Kin categories

Table 1: Original 115 kin categories used in the data collection. Distinction based on the gender of the speaker (male/female) was used in the data collection but is not presented here. The list was developed in the Kinbank project (see Passmore et al. 2023 for further details). The categories which do not have kin terms in Uralic languages (cousins by parent's sibling's age, cousins by the relative age of the cousin and co-spouses) are indicated with an asterisk (*).

| Description | Parameter |
|---|---|
| sibling | G |
| brother | B |
| sister | Z |
| elder brother | eB |
| younger brother | yB |
| elder sister | eZ |
| younger sister | yZ |
| parent | P |
| father | F |
| mother | M |
| child | C |
| son | S |
| daughter | D |
| ancestor | A |
| grandparent | PP |
| father's father | FF |
| father's mother | FM |
| mother's father | MF |
| mother's mother | MM |
| grandchild | CC |
| son's son | SS |
| son's daughter | SD |
| daughter's son | DS |
| daughter's daughter | DD |
| father's brother | FB |
| father's sister | FZ |
| mother's brother | MB |
| mother's sister | MZ |
| father's older brother | FeB |
| father's younger brother | FyB |
| father's older sister | FeZ |
| father's younger sister | FyZ |

| Description | Parameter |
| --- | --- |
| mother's older sister | MeZ |
| mother's younger sister | MyZ |
| mother's older brother | MeB |
| mother's younger brother | MyB |
| brother's son | BS |
| brother's daughter | BD |
| sister's son | ZS |
| sister's daughter | ZD |
| older brother's son | eBS |
| younger brother's son | yBS |
| older brother's daughter | eBD |
| younger brother's daughter | yBD |
| older sister's son | eZS |
| younger sister's son | yZS |
| older sister's daughter | eZD |
| younger sister's daughter | yZD |
| father's sister's daughter | FZD |
| father's brother's daughter | FBD |
| mother's brother's daughter | MBD |
| mother's sister's daughter | MZD |
| father's brother's son | FBS |
| father's sister's son | FZS |
| mother's brother's son | MBS |
| mother's sister's son | MZS |
| *father's older brother's son | FeBS |
| *father's younger brother's son | FyBS |
| *father's older sister's son | FeZS |
| father's younger sister's son | FyZS |
| *father's older brother's daughter | FeBD |
| *father's younger brother's daughter | FyBD |
| *father's older sister's daughter | FeZD |
| father's younger sister's daughter | FyZD |
| *mother's older brother's son | MeBS |
| *mother's younger brother's son | MyBS |
| *mother's older sister's son | MeZS |
| *mother's younger sister's son | MyZS |
| *mother's older brother's daughter | MeBD |
| *mother's younger brother's daughter | MyBD |
| *mother's older sister's daughter | MeZD |
| *mother's younger sister's daughter | MyZD |
| *father's brother's older son | FBeS |

| Description | Parameter |
|---|---|
| *father's brother's younger son | FByS |
| *father's sister's older son | FZeS |
| *father's sister's younger son | FZyS |
| *father's brother's older daughter | FBeD |
| *father's brother's younger daughter | FByD |
| *father's sister's older daughter | FZeD |
| *father's sister's younger daughter | FZyD |
| *mother's brother's older son | MBeS |
| *mother's brother's younger son | MByS |
| *mother's sister's older son | MZeS |
| *mother's sister's younger son | MZyS |
| *mother's brother's older daughter | MBeD |
| *mother's brother's younger daughter | MByD |
| *mother's sister's older daughter | MZeD |
| *mother's sister's younger daughter | MZyD |
| spouse | E |
| husband | H |
| wife | W |
| husband's father | HF |
| husband's mother | HM |
| wife's father | WF |
| wife's mother | WM |
| brother's wife | BW |
| sister's husband | ZH |
| wife's brother | WB |
| wife's sister | WZ |
| husband's brother | HB |
| husband's sister | HZ |
| *co-wife | co-W |
| *co-husband | co-H |
| father's wife (not mother) | FW(notM) |
| mother's husband (not father) | MH(notF) |
| son's wife | SW |
| son's wife's mother | SWM |
| son's wife's father | SWF |
| daughter's husband | DH |
| daughter's husband's mother | DHM |
| daughter's husband's father | DHF |
| father's sister's husband | FZH |
| father's brother's wife | FBW |
| mother's sister's husband | MZH |
| mother's brother's wife | MBW |

Table 2: Additional categories, i.e. categories which were not included in the Kinbank template but which were added here. The categories are sorted from most to least common in the Uralic family.

| Description | Parameter |
| --- | --- |
| younger brother's wife | yBW |
| elder brother's wife | eBW |
| husband's elder brother | HeB |
| wife's younger brother | WyB |
| husband's younger sister | HyZ |
| elder sister's husband | eZH |
| wife's elder brother | WeB |
| wife's elder sister | WeZ |
| husband's younger brother | HyB |
| husband's elder sister | HeZ |
| wife's sister's husband | WZH |
| younger sister's husband | yZH |
| wife's brother's wife | WBW |
| wife's younger sister | WyZ |
| husband's brother's wife | HBW |
| husband's sister's husband | HZH |
| husband's elder brother's wife | HeBW |
| father's elder brother's wife | FeBW |
| mother's elder brother's wife | MeBW |
| father's younger brother's wife | FyBW |
| father's elder sister's husband | FeZH |
| mother's younger brother's wife | MyBW |
| mother's elder sister's husband | MeZH |
| wife's elder sister's husband | WeZH |
| husband's younger brother's wife | HyBW |
| husband's younger sister's husband | HyZH |
| father's younger sister's husband | FyZH |
| mother's younger sister's husband | MyZH |
| elder brother's wife (f speaker) | feBW |
| wife's younger brother's wife | WyBW |
| wife's younger sister's husband | WyZH |

**Appendix 2: Loanwords by branch**[1]

Saami

Proto-Saami loans

1. SaaN *áhkká* 'wife; old woman', S *aahka* 'grandmother; old woman', Sk *ä'ǩǩ* 'wife' < PS *\*ākkā* 'wife; old woman' ← Finnic, cf. Fi *akka* 'wife; old woman' (Aikio 2009: 244)
2. SaaN *bárdni* 'son, boy', S *baernie* 'boy, (unmarried) son', Sk *pä'rnn* 'boy' < PS *\*pārnē* ← Scand, cf. ON *barn* 'child', Swe *barn* (Qvigstad 1893: 101; Sammallahti 1998: 230)
3. SaaS *dektier*, *daktere* '(married) daughter' < PS *\*tẹktēr*, *\*tẹktārẹ* ← PN *\*duhter-* (> ON *dóttir* 'daughter') (Aikio 2020: 17; Metsäranta 2023: 167–172)
4. SaaN *eanu* 'uncle (mother's brother; mother's male cousin; nephew or niece (to their uncle), cousin's child (to their mother's male cousin))', S *jyöne* 'mother's brother or male cousin', Sk *jään*, *jeän* 'maternal uncle' < PS *\*eanōj* ← PF *\*enoi* 'maternal uncle' (Kuokkala 2018: 39; Aikio 2020: 45–47)
5. SaaN *gállis* 'husband; old man', S *gaellis* 'husband', Sk *kää'lles* 'old man; husband' < PS *\*kāllēs* ← PN *\*karlaz* > ON *karl* 'man; man of the common people' > Elfdalian *kall* 'man; husband' (Qvigstad 1893: 161–162; Sammallahti 1998: 128–129; Kroonen 2013: 285)
6. SaaN *guoibmi* 'comrade, companion; husband, wife; neighbor' < PS *\*kuojmē* ← PF *\*kaima* 'namesake' (SSA 1: 276)

---

1. A few notes on the material presented in Appendix 2. First, the appendix does not aim to provide an exhaustive list of all the possible cognates in a given branch, but rather only includes the languages we surveyed and is there to inform the reader about the lexical basis of our calculations. As a result, the Saami forms for example only include North, South, and Skolt Saami. If the cognates are not kin terms, i.e. not belonging to any of the categories listed in Appendix 1, they have not been listed Appendix 2. This means for example that although Fi *lanko* 'sister's husband; spouse's brother' has a cognate in Estonian *lang* 'affinal relative', the Estonian word is not mentioned in the material, as it does not refer to any specific kin-term category.

7. SaaN *máhka* 'brother-in-law, uncle, husband of a relative that is not *spile* or *sivjjot*', S *maake* 'man that is married to one's older female relative: aunt's husband; older sister's (or female cousin's) husband; man of one's wife's kin that is younger than her: one's wife's younger brother or male cousin; one's wife's brother's (or sister's) son', Sk *maakk* 'older sister's husband; father's sister's husband' < PS *\*mākẹ* ← Scand, cf. ON *mágr* 'male in-law' (Qvigstad 1893: 229–230; Pelto 1962: 68–69; Sammallahti 1998: 129)

8. SaaN *mánná* 'child', S *maana* 'id.' < PS *\*mānā* ← Scand, cf. Swe *man* 'man' (Sammallahti 1998: 253)

9. SaaN *muottá* 'mother's younger sister or female cousin; nephew or niece (older sister's child to their aunt)', S *muahra* 'mother's younger sister or female cousin', Sk *mueʹđđ* 'mother's younger sister' < PS *\*muoϑā/ē* ← Germanic, cf. PGrm *\*mōþō(n)* 'mother's sister' (unattested derivation) ← *\*mōder-* 'mother' (Kümmel 2015: 121–129)

10. SaaN *neahpi* 'nephew or niece (to their uncle); cousin's child (female cousin's, to their mother's male cousin)', S *neapede* '(man's or boy's) sister's or female cousin's son or daughter; (man's) sister's grandchildren' < PS *\*neapē / \*neapātẹ* ← Finnic, PF *\*nẹpada* > Fi (dial.) *nepaa* 'cousin, cousin's child', Est *nõbu* 'cousin' (← Baltic) (SSA 2: 213)

    It is not entirely certain that the Saami group of words was borrowed from Finnic in its entirety. It has been suggested that at least SaaN *neahpi* could have been borrowed from Germanic, cf. ON *nefi* 'nephew, cousin's son; relative' (SKES: 373)

11. SaaN *viellja* 'brother', S *vielle* 'id.', Sk *villj* 'id.' < PS *\*vieljẹ* ← PF *\*velji* 'brother' (Aikio 2014: 68)

Post-Proto-Saami loans

1. SaaS *gaalla* 'husband' ← Scand, cf. Swe *karl* 'man, husband', Jämtlandic, Elfdalian *kall* 'man, husband'

2. SaaN *gálgu* 'wife; old woman' ← ? Scand, ON *\*kelg* < *kelling* 'wife' (Qvigstad 1893: 161)

3. SaaS *geellege* 'wife' ← ? Scand, cf. ON *kelling* 'wife'

    According to Qvigstad SaaN *gálgu* 'wife; old woman' is borrowed from *\*kelg* < *kelling* 'wife'. He does not mention the South Saami word, but it seems quite probable that the South Saami word is not entirely unrelated.

4. SaaN *eamit* 'housewife; (female) owner; solitary woman who does her own cooking; wife' ← Finnic, Fi *emäntä* 'lady (of the house); house-wife; hostess, matron; wife, spouse' (SSA 1: 104–105; Sammallahti 1998: 240–241)

5. SaaS *eejhtegh* 'parents' ← Scand, cf. ON *ættingi* 'relatives' (Qvigstad 1893: 86)

6. SaaN *isit* 'head of the household; man, husband' ← Finnic, Fi *isäntä* 'master, lord (of a household; host' (SSA 1: 229; Sammallahti 1998: 249)

7. SaaN *máttar* 'forefather, ancestor' < PS *\*māndẹr* ← Finnic, cf. PF *\*mandẹr* > Fi *manner* 'mainland, continent' (SSA 2: 147)

    The somewhat peculiar looking semantic connection can probably be explained through metaphor, cf. SaaS *maadtere* 'hut wall, especially the bottom part', Ter *māndạ r* 'mainland; strip of land on which the hut wall is erected and embankment on the outer wall'.

8. SaaS *måare*, *määre* 'mother' ← Scand, cf. Swe *mor* 'mother'

9. SaaSk *nee'vesǩ* 'son's wife' ← Karelian *ńeveskä* (Itkonen, T. I. 1943: 52; Korhonen 1977: 80)

10. SaaN *sivjjot* 'relative of the opposite sex through marriage; husband's brother or male relative; sister's husband or male relative (to their wife's sister); wife's sister or female relative; brother's wife (to their husband's brother)', SaaS *sïbjege* 'man's older brother's or male cousin's wife; husband's younger brother or cousin (in relation to older brother's or male cousin's wife)' ← Scand, cf. ON *sifjungr* 'relative' (Qvigstad 1893: 288)

11. SaaN *spile* 'wife's sister's husband, brother-in-law; wife's brother's wife, sister-in-law' ← Scand, cf. ON *svilar* 'brothers-in-law' (Qvigstad 1893: 314)

12. SaaSk *svaaik*, *svaajâk* 'wife's sister's husband' ← Russian *свояк* 'broth-er-in-law' (Juutinen 2022: 43)

13. SaaN *váhnemat*, *vánhemat* 'parents' ← Finnic, cf. Fi *vanhemmat* 'par-ents' (Sammallahti 1998: 265)

14. SaaSk *zee'tt* 'son-in-law' ← Russian *зять* 'son-in-law, brother-in-law' (Juutinen 2022: 54)

Finnic

Proto-Finnic loans

1. Fi *lanko* 'sister's husband; spouse's brother' < PF *\*lanko* ← Germanic, cf. PGrm *\*ga-langaz* > Old High German *gilang* 'brother-in-law, relative', PGrm *\*bi-langaz* > Old Low German *bilang* 'related; joined, connected' (LägLoS II: 167; SSA 2: 44–45)

2. Veps *mur′zain* '(young) wife' < LPF *\*morcijan* < MPF *\*mortijami* ← Baltic, cf. (accusative forms) *\*martjan ~ \*martjam*, Lith *martì* 'bride; young wife (until the birth of first child); daughter-in-law; sister-in-law', Latv *mārša* 'brother's wife', Old Prussian *martin*, *martan* 'bride' (SSA 2: 172, Junttila 2015: 56; 94; 145; EVE: s.v. *morcijan*)

3. Est *nõbu* 'cousin' < LPF/MPF *\*nepada* ← Baltic, cf. Old Lithuanian *nepotis*, *nepuotis*, *nepatis* 'grandchild; nephew; cousin's son' (SSA 2: 213; Junttila 2015: 89; EVE: s.v. *nëpada*)

4. Fi *sisar* 'sister', Veps *sizar* 'id.' < LPF/MPF *\*sisar*, Est *sõsar* 'sister', Liv *sõzār* < LPF/MPF *\*sęsar* ← Baltic, cf. Lith *sesuõ : seserᷠs* 'sister' (SSA 3: 187; Junttila 2015: 95; Kallio 2018: 255)

   Finnish and Veps, on one hand, and Estonian and Livonian, on the other, reflect slightly different PF proto-forms, which is perhaps due to parallel borrowing.

5. Fi *tytär* 'daughter', Est *tütar* 'id.', Veps *tütär* 'id.', Liv *tidār* 'id.' < LPF *\*tüťtär* < MPF *\*tüttäri* ← Baltic, cf. Lith *duktė : dukter-* (Junttila 2015: 96; 176; Metsäranta 2023: 167–172)

   The Finnic words have traditionally been considered cognate with SaaS *dektier*, *daktere* '(married) daughter' and MdE *ťejťeŕ* 'daughter', M *śťiŕ* (SSA 3: 349). The words were thought to reflect a common proto-form, e.g. West Uralic *\*tüktärə*, but there is no way of regularly deriving the Finnic geminate *\*-tt-* from an earlier *\*-kt-*. The Saami word has also been explained as a separate Scandinavian loan (Aikio 2020: 17), so it is perhaps best to approach the words as separate loanwords in Saami, Finnic, and Mordvinic.

6. Fi *äiti* 'mother' < LPF *\*äitei* ← Germanic, cf. PGrm *\*aiþīn ~ aiþōn* 'mother' > Gothic *aiþei* 'mother', ON *eiða* 'id.' (LägLoS III: 429–430; Kroonen 2013: 15)

Post-Proto-Finnic loans

1. Veps *bab* 'grandmother' ← Ru *баба* 'grandmother'
2. Veps *bratan* 'male cousin' ← Ru *братан* 'male cousin'
3. Veps *däd'* 'uncle' ← Ru *дядя* 'uncle'
4. Liv *māršša* 'daughter-in-law' ← Latv *mārša* 'sister-in-law' (LW: 218)
5. Fi *mummo, mummu* 'grandmother; old woman' ← Swe dial. *mummu, mumm = mormor* 'mother's mother' (SSA 2: 178)
6. Veps *mužik* 'man, husband' ← Ru *мужик* 'boor, peasant; man, old man'
7. Fi *pappa* 'grandfather' ← Swe *pappa* 'father' (SSA 2: 311)
8. Veps *sestrii* 'female cousin' ← Ru *(двоюродная) сестра* 'female cousin'
9. Liv *švōgõr* 'brother-in-law' ← Latv *švager* 'brother-in-law' (LW: 398; Winkler 2014: 219)
10. Fi *vaari* 'grandfather' ← Swe *far* 'father' (SSA: 386)
11. Veps *vunuk* 'grandchild' ← Ru *внук* 'male grandchild'
12. Liv *znūot* 'son-in-law' ← Latv *znuōts* 'son-in-law; brother-in-law' (LW: 401)

## Mordvin (Erzya and Moksha)

Pre-Proto-Mordvin loans

1. E *miŕďe* 'husband', M *miŕď'ä* 'id.' < Pre-Proto-Mordvin *\*mertä* ← PII *\*m̥rtá-* 'dead' or Pre-Indo-Iranian *\*mérto-* 'mortal, human' (Holopainen 2019: 143–146)

   Usually thought to have been borrowed either from PII *\*m̥rtá-* 'dead' or Pre-Indo-Iranian *\*mérto-* 'mortal, human' as PU *\*mertä*, whence also Udm *murt* 'human, person; alien, strange, foreign', Komi *mort* 'human' (< Proto-Permic *\*mɔrt*). The Permic words could be interpreted as later loans from PI *\*marta-*, although the phonological match between Mordvin and Permic is formally flawless.

2. E *sazor* 'younger sister', M *sazər* 'id.' < Pre-Proto-Mordvin *\*sasar* ← PII/PI *\*swasar-* 'sister' (Mészáros 2001: 172; Holopainen 2019: 222–224; Metsäranta 2023: 162–167)

3. E *ťejťeŕ* 'daughter', M *śťiŕ* 'id.' < Pre-Proto-Mordvin *\*tüktärə* ← Baltic, cf. Lith *duktė : dukter-* (Mészáros 2001: 171)

Post-Proto-Mordvin loans

1. M *aka* 'elder sister; father's or mother's younger sister' ← Chu *aɢi̯*, *aɢi*, *akka*, *aɢaj* 'elder sister' (Paasonen 1897: 27; Mészáros 2001: 172)
2. M *akĺä* 'husband's elder sister' ← Ta, cf. dial. *aɣî la* 'mother-in-law', 'aunt' (Paasonen 1897: 27; Mészáros 2001: 176)
3. E *aťa* 'paternal grandfather; old man', M *aťä* 'paternal grandfather' ← Turkic

    According to Mészáros (2001: 169) of Finno-Ugric origin, but Turkic origin is more probable, cf. Ta *ata*, *äti* 'father', Chu *aťɛ* 'father'
4. E *ava* 'mother', M *ava* 'wife' ← Chu *ави*, *аваj* 'mother' (Mészáros 2001: 169; 17)
5. E *baba* 'old woman; wife (old); grandmother; father's mother', M *paba*, *baba* 'father's mother' ← Ru *баба* 'old woman' (MdWb: 110)

    Mészáros (2001: 170) lists the Mordvin word as a Tatar loan without citing a loan original, but according to MdWb it comes from Russian; the latter possibility seems more likely.
6. E *balduz*, *paldus* 'wife's sister', M *baldəz*, *paldəs* 'id.' ← Ta *balduz* 'sister-in-law; wife's (younger) sister' (Paasonen 1897; Mészáros 2001: 176)
7. E *baĺźa* 'wife's younger brother', M *baźä*, *paĺźä*, *paźä* 'id.' ← Mishar Tatar *baźa* 'wife's sister's husband' (Paasonen 1897: 43; Mészáros 2001: 176)
8. M *baťka* 'wife's father' ← Ru *батька* 'dad, father' (Mészáros 2001: 174)
9. E *brat* 'brother', M *brat* 'id.' ← Ru *брат* 'brother' (Mészáros 2001: 171)
10. E *ćora* 'son', M *ćora*, *śora* 'id.' ← Chu *čora*, *čura* 'servant, slave' (Paasonen 1897: 48; Mészáros 2001: 171)
11. E *ďeda* 'father's father' ← Ru *дед* 'grandfather; old man' (Mészáros 2001: 169)
12. E *duga*, *dugan* 'younger brother', M *duga* 'id.' ← Ta *tugan* 'one's own; relative; brother' (Paasonen 1897: 56; Mészáros 2001: 172)
13. E *ezna* 'older sister's husband; son-in-law', M *äzna* 'id.' ← Ta *jiznä* 'older sister's husband' (Paasonen 1897: 33; Mészáros 2001: 174; 176)
14. E *koźejka*, *koźajka* 'wife' ← Ru *хозяйка* 'hostess' (Mészáros 2001: 174)
15. E *mačka* 'wife's mother' ← Ru *мачка* 'mom' (Mészáros 2001: 174)
16. M *matka* 'wife's mother' ← Ru *матка* '(animal) mother' (Mészáros 2001: 174)
17. E *nuka* '(male) grandchild', M *unək* 'id.' ← Ru *внук* '(male) grandchild' (Mészáros 2001: 171)

18. E *nućka* 'grandchild' ← Ru *внучек* (genitive *внучка*) 'grandson' (MdWb: 1363)

    According to Mészáros (2001: 171) the Erzya word was borrowed from Russian *внучка* '(female) grandchild'. This remains a possibility, but MdWb expresses a different view, according to which the word was borrowed from the diminutive of *внук* i.e. *внучек*.

19. E *paťa* 'elder sister; father's or mother's sister', M *paťä* 'elder brother' ← Slavic, cf. Ru *батя* 'father', Bulg *баща*, *бате*, *батьо*, *бачо* 'elder brother; uncle'

    According to Mészáros (2001: 171–172) either of Finno-Ugric or Baltic origin. Neither option strikes one as probable. Instead, the Mordvin word is probably of Slavic origin, cf. Hungarian *bátya* 'elder brother' that was borrowed from Slavic as well.

20. E *pľemjańńik* 'nephew', M *pľəmäńńək* 'id.' ← Ru *племянник* 'nephew' (Mészáros 2001: 173)

21. E *pľemjańńića* 'niece', M *pľəmäńńəća* 'id.' ← Ru *племянница* 'niece' (Mészáros 2001: 173)

22. E *svojačeńica* 'wife's older sister' ← Ru *свояченица* 'sister-in-law' (Mészáros 2001: 176)

23. E *svojak* 'wife's (older) sister's husband', M *svə̂jak*, *svajak* 'id.' ← Ru *свояк* 'brother-in-law' (Mészáros 2001: 177)

24. E *šuŕiń*, *šuŕeń* 'brother-in-law, wife's brother', M *šuŕəń* 'id.' ← Ru *шурин* 'brother-in-law, wife's brother' (Mészáros 2001: 176)

25. E *ťeťa* 'father', M *ťäťä* 'id.' ← Ru *тятя* 'father' (Veršinin 2009: 436)

    According to Mészáros (2001: 169), the Mordvin word is of Finno-Volgaic origin, but this explanation is treated as uncertain even by the author herself. Veršinin (2009: 436) suggests that the Erzya word was borrowed from Russian. A Russian origin for the Mordvin words seems much more likely considering that even a back-vocalic variant *taťa* is found in a number of Erzya dialects (MdWb: 2396). Also the fact that intervocalic *ť* has escaped voicing between vowels indicates that the word is a rather recent addition to the Mordvin lexicon.

26. E *ťošča* 'wife's mother' ← Ru *тёща* 'wife's mother' (Mészáros 2001: 174)

## Mari

### Pre-Proto-Mari loans

1. MariM *šüžar* 'younger sister', MariH *šə̂žar* 'id.' < PM *\*sŭzar* < ? Pre-Proto-Mari *\*susar* ← PII/PI *\*swasar-* 'sister' (Holopainen 2019: 222–224; Metsäranta 2023: 162–167)

### Proto-Mari or Common Mari loans

1. MariM *aβa* 'mother', MariH *äβä* 'id.' < PM *\*äwä* ← Chu *aʙi, aʙaj* 'mother' (Räsänen 1920: 109)
2. MariM *ača* 'father', MariH *ät́a* 'id.' < PM *\*ät́a* ← Chu *att́ɛ* 'father' (Räsänen 1920: 239)

   Räsänen classifies the Mari word in a category of words common to both Mari and Chuvash whose origin is unknown. Considering that Chuvash loans are rather ubiquitous in Mari kinship terminology and the fact that the word is phonologically clearly a rather recent addition to Mari vocabulary, it is still quite probably a loan from Chuvash to Mari.
3. MariM *aka* 'elder sister; parent's (father's or mother's) younger sister', MariH *äkä* 'elder sister; mother's sister; father's younger sister' < PM *\*äkä* ← Chu *aɢi̭, aɢi, akka, aɢaj* 'elder sister' (Räsänen 1920: 112)
4. MariM *aza* 'child', MariH *äźä* 'id.' < PM *\*äćä* ← Chu *aᴅźa* 'child' (Räsänen 1920: 111)
5. MariM *βate* 'wife', MariH *βätə* 'id.' < PM *\*wätə* ← Chu *vaᴅə̂* 'old' (Räsänen 1920: 120)
6. MariM *marij* 'Mari; man, husband', MariH *marə̂* id.' < PM *\*marə(j)* ← Iranian, cf. Old Indian *márya-* 'young man' (Holopainen 2019: 135–136)
7. MariM *ońo* 'father-in-law', MariH *ońə̂* 'id.' < PM *\*ońə* ← Chu *χonə̂m* 'my father-in-law', *χuńə̂m* (Räsänen 1920: 166)

   In Hill Mari *ońə̂* exists alongside the inherited Uralic *oβə̂* 'father-in-law' (< PU *\*ęppə* 'father-in-law')
8. MariM *ońaka* 'husband's or wife's elder sister', MariH *ońaka* 'id.' < PM *\*ońaka* ← Chu *χuńaɢa* 'wife's elder sister' (Räsänen 1920: 166)
9. MariM *ońə̂ska* 'husband's or wife's elder brother', MariH *ońə̂ska* 'id.' < PM *\*ońəska* ← Chu *χońə̂ska* 'wife's brother' (Räsänen 1920: 166)

10. MariM *posana* 'wife's sister's husband', MariH *pasana* 'id.' < PM \**påśana*
    ← Chu *požana*, *pužana* (Räsänen 1920: 180), cf. Udmurt *buśon(o)*
11. MariM *šoľo* 'younger brother', MariH *šoľa*, *šoľə̂* 'id.' < PM \**šoľə* ← Chu
    *šə̂llə̂m*, *šə̂llu*, *šə̂ľľə* '(my) younger brother' (Räsänen 1920: 211)
12. MariM *tulače* 'son's wife's mother; daughter's husband's mother', Mar-
    iH *tə̂klacə̂*, *tə̂γə̂lacə̂* 'id.' < PM ?\**tŭ(k)lačə̂* ← Chu *tə̂χlaɒźə̂* 'wife's or hus-
    band's sister or other female relative' (Räsänen 1920: 225)
13. MariM *tular* 'son's wife's father; daughter's husband's father', MariH
    *tə̂klar* 'id.' < PM ?\**tŭ(k)lar* ← Chu \**tə̂γlar* (Räsänen 1920: 225)

Post-Proto-Mari loans

1. MariM *jeŋga* 'elder brother's wife', MariH *jeŋgä* 'id.' ← (Chu) ← Ta
   *žiŋgä* 'daughter-in-law' (Räsänen 1923: 30)
2. MariM *ńoγa* 'child' (cf. Mari dial. *ńuγar*, *ńoγar* 'the small ones') ← ? Ta,
   cf. Kyrgyz *nögör* 'servant' (Räsänen 1923: 47)
3. MariH *roδə̂* 'family, stock; grandfathers, great-grandfathers' ← Ru *род*
   'family, kin'
4. MariH *ťeťä* 'child' ← Ru *дитя* 'child' (dated)
5. MariH *ťoťa* 'grandfather' ← Ru *тятя* 'father'
6. MariM *unə̂ka* 'grandchild', MariH *ə̂nə̂ka* 'id.' ← Russian *внук* 'male
   grandchild'

Permic

Pre-Proto-Permic loans

1. Udm *suzer* 'younger sister' (~ Komi *sozor* 'Fadenbruch im Gewebe')
   < Proto-Permic \**sɔzVr* < Middle Proto-Permic \**sasar* ← PII/PI \**swasar-*
   'sister' (Metsäranta 2023: 162–167)

Komi

1. Komi *bab* 'grandmother' ← Ru *баба* 'grandmother' (Kalima 1910: 37)
2. Komi *bať* 'father' ← Ru *батя* 'father' (Kalima 1910: 40)
3. Komi *bratan* 'male cousin' ← Ru *братан* 'male cousin'
4. Komi *ćeľaď* 'children' ← Ru *челядь* 'servants' (Kalima 1910: 150)

5. Komi *ďaď* 'uncle (father's or mother's brother, aunt's husband); step-father' ← Russian *дядя* 'uncle' (Kalima 1910: 45)
6. Komi *ďed* 'grandfather' ← Ru *дед* 'grandfather' (Kalima 1910: 46)
7. Komi *plemjannik* 'nephew' ← Ru *племянник* 'nephew' (Kalima 1910: 106)
8. Komi *plemjannića* 'niece' ← Ru *племянница* 'niece' (Kalima 1910: 106)
9. Komi *svesťa* 'wife's sister' ← Ru *свесьтя* 'wife's sister' (Kalima 1910: 135)
10. Komi *ťesť* 'wife's father' ← Ru *тесть* 'wife's father' (Kalima 1910: 145)
11. Komi *ť'ët* 'aunt, father's or mother's sister; stepmother' ← Ru *mëma*, *mëmя* 'aunt' (Kalima 1910: 147)
12. Komi *ťęša* 'wife's mother' ← Ru *теща* 'wife's mother' (Kalima 1910: 147)
13. Komi *vnuk* 'male grandchild' ← Ru *внук* 'male grandchild' (Kalima 1910: 167)
14. Komi *vnučka* 'female grandchild' ← Ru *внучка* 'female grandchild' (Kalima 1910: 167)
15. Komi *źať* 'sister's husband; daughter's husband' ← Ru *зять* 'son-in-law, brother-in-law' (Kalima 1910: 171)
16. Komi *šurin* 'wife's brother' ← Ru *шурин* 'brother-in-law (wife's brother)' (Kalima 1910: 142)

Udmurt

1. Udm *agaj* 'elder brother; father's brother' ← Ta *agaj* 'uncle (honorific term used of older men)' (Csúcs 1990: 95)
2. Udm *aka* 'elder sister; father's sister, uncle's wife' ← Chu *aɢi̮*, *aɢi*, *akka* 'elder sister' (Wichmann 1903: 38)
3. Udm *anaj* 'mother' ← Ta *ana* 'mother', *änej* 'Mütterchen', *anaj* (Csúcs 1990: 104)
4. Udm *apaj* 'elder sister' ← Ta *apaj* 'honorific vocative for elder sister' (Csúcs 1990: 105–106)
5. Udm *ataj* 'father' ← Ta *ätej*, *ataj* 'Väterchen, lieber Vater' (Csúcs 1990: 112)
6. Udm *brat* 'brother' ← Ru *брат* 'brother'
7. Udmurt *bulti̮r* 'wife's sister; second wife (after the death of the first)' ← Chu *pulo̭i̮r*, *polo̭i̮r* 'wife's younger brother or sister; husband's younger brother or sister' (Wichmann 1903: 49)
8. Udm *buśon(o)* 'wife's sister's husband' ← Chu *poźana*, *puźana* (Wichmann 1903: 52), cf. MariM *posana*, H *pasana*

9. Udm *d́ad́a*, *d́ad́aj* 'father; uncle' ← Ru *дядя* 'uncle'
10. Udm *kart* 'husband' ← Ta *kart* 'old man' (Csúcs 1990: 207)
11. Udm *ken* 'son's wife' ← Chu *kin*, *ḱiń* 'son's wife, younger brother's wife' (Wichmann 1903: 71–72)
12. Udm *kudo* 'daughter's husband's father; son's wife's father' ← Chu *χοᴅа* 'Freiwerber; son-in-law's or daughter-in-law's father' (Wichmann 1903: 77–78)
13. Udm *tukľaći* 'daughter's husband's mother; son's wife's mother' ← Chu *toχlåč̌* 'Brautwerberin' (Wichmann 1903: 109)
14. Udm *šịdnar* 'husband's brother' ← Chu (WotjWsch: 247)

## Khanty

1. KhKaz *aki* 'father's older brother; mother's older brother; husband's father; husband's older brother; father's older sister's husband; mother's older sister's husband' ← MsN, cf. MsSo *aki* 'term of respect for an older relative; father-in-law; father's older brother; mother's older brother; husband's older brother; husband's father's older or younger brother or husband's mother's older or younger brother; grandfather' (UED: 42–43)

## Mansi

1. MsSo *āpš̌i* 'younger brother; brother's son; (younger male) cousin' ← KhKaz *apśi* 'younger brother; father's younger brother; older brother's son (younger than ego); father's brother's son (younger than ego)' (DeWoS: 149–150)
2. MsSo *jaɣlaŋ* 'stepfather; mother's sister's husband' ← KhN, cf. KhKaz *jĭwłeŋ* 'stepfather; mother sister's husband' (DeWoS: 326–327; Karjalainen 1913: 230)
3. MsSo *kil* 'wife's (older or younger) sister; sister's husband; wife's sister's husband' ← KhN, cf. KhKaz *kĭłi* 'wife's older or younger sister; wife's older or younger brother's daughter; wife's older or younger sister's husband' (DeWoS: 620)
4. MsSo *nij* 'mother's sister' ← Khanty, cf. KhKaz *nĭj* 'mother's younger sister' (WogWb: 497)

5. MsSo *ńāwram* 'child' ← Khanty, cf. KhKaz *ńawrɛm* 'child' (WogWb: 524; DeWoS: 1038)
6. MsSo *ɔ̄jka* 'man, husband, old man' ← Komi, cf. Komi-Permyak *ajka* 'man, husband', Komi-Jaźva *ajka* 'husband', Komi-Zyrian *ajka* 'husband's father' (Rédei 1970: 91)
7. MsSo *pānt* 'husband's younger brother; sister's son (name used by mother's sister)' ← Middle Iranian *\*band-* 'to bind' > Old Persian *baⁿdaka-* 'vassal' (Korenchy 1972: 63; Holopainen 2019: 183–184)

## Hungarian

1. Hung *bátya* 'elder brother' ← Slavic, cf. Bulg *баща, бате, батьо, бачо* 'elder brother; uncle' (EWUng: 87)
2. Hung *gyerek, gyermek* 'child' ← WOT *\*ǰärmek, \*ǰärmik, \*ǰämrik* (EWUng: 495–496; WOT: 384–386)
3. Hung *mostoha* 'stepmother' ← Slavic, cf. Bulg *мащеха* 'stepmother' (EWUng: 998)
4. 125. Hung *unoka* 'grandchild' ← Slavic, cf. Serbo-Croatian *unuk* 'grandchild' (EWUng: 1578)
5. 126. Hung *sógor* 'brother-in-law' ← Austrian High German *swoger* 'brother-in-law' (EWUng: 1342)

## Samoyedic

### Forest Enets

1. EnF *ńabaku* 'elder sister' ← NenT *ńabako* 'elder sister; father's younger sister; father's brother's daughter (older than ego)' ← NenT *ńaba* 'stepmother; older brother's wife' < PSam *\*äpå* 'ältere Schwester' (Janhunen 1977: 21)

    The presence of prothetic initial nasal, which is typically not a feature of Forest Enets, points to a borrowing from Tundra Nenets where said nasal is an expected regular development. A loan origin of the Forest Enets word is further supported by the existence of EnF *abaa* 'elder sister; father's or mother's younger sister', the regular reflex of PSam *\*äpå*, whence also NenT *ńaba* and *ńabako*.

**Appendix 3: Kin categories with loanwords**

Kin categories with loanwords in Uralic languages ranked by the total number of borrowing events. (f) in the Kin category column indicates it is a category specifically when the speaker is female. Parameter column introduces the abbreviations of the kin categories. Plus sign indicates categories where the age distinction (e = elder, y = younger) has been merged into one category. The table has been sorted by the total number of borrowing events. The column with a maximum of one borrowing event per language has a § in cases when a term in the kin category in question has been borrowed both into the proto-language and into the individual languages in one of the subgroups; in the case of merged categories the § sign indicates that two categories exist in one language (e.g. Erzya has both WZ and WeZ) and these have been counted separately. # of languages with loanwords is the total number of languages where the kin term in question has a loanword. Asterisk in polysemy column indicates that at least in one of the languages the kin term is polysemous (i.e. linked to more than one kin category). The number in the coexistence column indicates in how many languages the borrowed kin term coexists with a term with no evidence of borrowing.

| Kin category | Parameter | # of borrowing events (total) | # of borrowing events (max. 1 / lang.) | # of languages with a loanword | Polysemy | Coex. |
|---|---|---|---|---|---|---|
| sister's husband | ZH+eZH | 11 | 9 § | 10 | * | 3 |
| husband's brother | HB+HeB+HyB | 10 | 10 | 12 | * | 0 |
| wife's brother | WB+WyB+WeB | 10 | 10 § | 10 | * | 1 |
| wife | W | 9 | 7 § | 8 | | 4 |
| grandchild | CC | 9 | 7 | 7 | | 2 |
| husband | H | 8 | 8 § | 10 | | 3 |
| wife's sister | WZ+WeZ | 8 | 8 § | 8 | * | 0 |
| wife's sister's husband | WZH+WeZH | 7 | 7 | 8 | * | 0 |
| child | C | 7 | 7 § | 7 | | 4 |

## Borrowability of kinship terms in Uralic languages

| Kin category | Parameter | # of borrowing events (total) | # of borrowing events (max. 1 / lang.) | # of languages with a loanword | Polysemy | Co-ex. |
|---|---|---|---|---|---|---|
| father's father | FF | 7 | 5 | 5 | * | 3 |
| father's mother | FM | 6 | 6 | 6 | * | 3 |
| elder sister | eZ | 6 | 5 | 6 | * | 1 |
| father | F | 6 | 5 | 6 | | 2 |
| mother | M | 5 | 5 | 6 | | 2 |
| daughter's husband | DH | 5 | 5 | 5 | * | 4 |
| mother's mother | MM | 5 | 5 | 5 | * | 2 |
| sister's son | ZS | 5 | 4 | 5 | * | 1 |
| sister's daughter | ZD | 5 | 4 | 5 | * | 2 |
| brother | B | 4 | 4 | 6 | | 1 |
| mother's brother | MB | 4 | 4 | 6 | * | 1 |
| younger brother | yB | 4 | 4 | 5 | * | 3 |
| mother's sister | MZ | 4 | 4 | 4 | * | 1 |
| brother's son | BS | 4 | 4 | 4 | * | 1 |
| father's brother's son | FBS | 4 | 4 | 4 | * | 2 |
| father's sister's son | FZS | 4 | 4 | 4 | * | 2 |
| mother's brother's son | MBS | 4 | 4 | 4 | * | 2 |
| mother's sister's son | MZS | 4 | 4 | 4 | * | 3 |
| brother's wife | BW+eBW | 4 | 4 | 4 | * | 0 |
| mother's father | MF | 4 | 3 | 3 | * | 2 |
| father's brother | FB | 4 | 3 | 3 | * | 0 |
| wife's mother | WM | 4 | 3 | 3 | | 2 |
| daughter | D | 3 | 3 | 7 | | 0 |
| younger sister | yZ | 3 | 3 | 5 | | 0 |
| son | S | 3 | 3 | 5 | | 1 |
| mother's younger sister | MyZ | 3 | 3 | 5 | * | 0 |

| Kin category | Parameter | # of borrowing events (total) | # of borrowing events (max. 1 / lang.) | # of languages with a loanword | Poly-semy | Co-ex. |
|---|---|---|---|---|---|---|
| wife's father | WF | 3 | 3 | 4 | * | 2 |
| elder brother | eB | 3 | 3 | 3 | * | 2 |
| father's sister | FZ | 3 | 3 | 3 | * | 1 |
| brother's daughter | BD | 3 | 3 | 3 | * | 1 |
| son's wife | SW | 3 | 3 | 3 | | 2 |
| father's sister's husband | FZH+FeZH | 3 | 3 | 3 | * | 0 |
| mother's sister's husband | MZH+ MeZH | 3 | 3 | 3 | * | 0 |
| father's younger sister | FyZ | 2 | 2 | 3 | * | 0 |
| son's wife's mother | SWM | 2 | 2 | 3 | * | 0 |
| son's wife's father | SWF | 2 | 2 | 3 | * | 0 |
| husband's father | HF | 2 | 2 | 3 | * | 2 |
| husband's sister | HeZ+HZ | 2 | 2 | 3 | * | 1 |
| parent | P | 2 | 2 | 2 | | 0 |
| ancestor | A | 2 | 2 | 2 | | 0 |
| father's sister's daughter | FZD | 2 | 2 | 2 | * | 1 |
| father's brother's daughter | FBD | 2 | 2 | 2 | * | 1 |
| mother's brother's daughter | MBD | 2 | 2 | 2 | * | 1 |
| mother's sister's daughter | MZD | 2 | 2 | 2 | * | 1 |
| father's wife (not mother) | FW(notM) | 2 | 2 | 2 | * | 1 |
| mother's husband (not father) | MH(notF) | 2 | 2 | 2 | * | 0 |

## Borrowability of kinship terms in Uralic languages

| Kin category | Parameter | # of borrowing events (total) | # of borrowing events (max. 1 / lang.) | # of languages with a loanword | Polysemy | Co-ex. |
|---|---|---|---|---|---|---|
| daughter's husband's mother | DHM | 2 | 2 | 3 | * | 0 |
| daughter's husband's father | DHF | 2 | 2 | 3 | * | 0 |
| sister | Z | 2 | 1 | 4 | | 1 |
| father's elder brother | FeB | 1 | 1 | 1 | * | 0 |
| mother's elder brother | MeB | 1 | 1 | 1 | * | 0 |
| spouse | E | 1 | 1 | 1 | | 1 |
| wife's brother's wife | WBW | 1 | 1 | 1 | * | 0 |
| sister's son (female speaking) | fZS | 1 | 1 | 1 | * | 0 |
| elder sister's son (female speaking) | feZS | 1 | 1 | 1 | * | 0 |
| elder sister's daughter (female speaking) | feZD | 1 | 1 | 1 | * | 0 |
| sister's husband (female speaking) | fZH | 1 | 1 | 1 | * | 0 |
| mother's brother's wife | MBW | 1 | 1 | 1 | * | 0 |
| father's brother's wife | FBW | 1 | 1 | 1 | * | 0 |

**Appendix 4: Kin categories without loans**

Relatively common kin categories with no borrowed kin terms in Uralic languages. The 20 categories listed here exist in three or more languages covering more than one subgroup of the Uralic tree (Samoyedic languages with no borrowed kin terms were not included under such criteria). This criterion was set to include only kin categories which are at least somewhat common in Uralic languages, as there were several categories which existed in less than three languages and/or covered only one subgroup.

| Kin category | Parameter |
|---|---|
| sibling | G |
| grandparent | PP |
| son's son | SS |
| son's daughter | SD |
| daughter's son | DS |
| daughter's daughter | DD |
| father's younger brother | FyB |
| father's elder sister | FeZ |
| mother's elder sister | MeZ |
| elder brother's son | eBS |
| younger brother's son | yBS |
| elder brother's daughter | eBD |
| younger brother's daughter | yBD |
| elder sister's son | eZS |
| younger sister's son | yZS |
| elder sister's daughter | eZD |
| younger sister's daughter | yZD |
| husband's mother | HM |
| husband's brother's wife | HBW |
| husband's sister's husband | HZH |

*Orçun Ünal*
Georg-August-Universität Göttingen

# Nova Turco-Samoiedica

The present paper focuses on the lexical contact between Turkic and Samoyedic and discusses nine new possible Turkic loanwords in Proto-Samoyedic and eight new possible Samoyedic loanwords in Turkic. The introduction offers a modest bibliography of the scattered studies on the subject. Two of the new Turkic loanwords in Proto-Samoyedic suggest that they reached the recipient language through the mediation of Yeniseian languages.

## 1. State of the art

The mutual contact between Samoyedic and Turkic has been the subject of many studies, and the aspect of lexical borrowing has been widely discussed. With no claim of exhaustiveness, the following sources concerning this subject may be cited: Ramstedt (1912), Paasonen (1912/1913–1916/1917), Donner (1924), Ligeti (1926–1932), Németh (1930: 118–119), Toivonen (1933: 102), Sinor (1947; 1980), Terent'ev (1976; 1982; 1983; 1989), Janhunen (1977), Filippova (1980b), Róna-Tas (1980; 1988: 743–746), Ligeti (1986: 495–497), Helimski (1995), Helimski & Stachowski (1995), Dybo & Normanskaja (2016: 50–51), and Piispanen (2018).

The most recent and most comprehensive work on this topic is Terent'ev (1999), albeit not without shortcomings. It contains many etymologies in both directions of borrowing, but they are not thought out well on the Turkic side. Furthermore, some comparisons are based on irregular sound correspondences. Likewise, Dybo (2007: 135–154) extensively dealt with the

words borrowed between Proto-Samoyedic (PS), individual Samoyedic languages, and Turkic. Dybo (2014: 10–11) reproduced data from earlier sources without adding anything new.

Joki (1952) examined the Turkic borrowings in the Sayan Samoyedic languages. Filippova (1974; 1980a; 1986a; 1986b), Tamás (1975), and Teníšev (1977) dealt with the Turkic loanwords in Selkup. Janhunen (1989) and Helimski (1991; 1998) focused on interaction between Mator and Turkic. Helimski (1986; 1992–1993; 1997), Blažek (2016), as well as Ünal (2022a: §8, §9, §11, §14) present some additional Turkic etymologies for Samoyedic words.

In the present study, I will investigate nine new possible Turkic loanwords in Proto-Samoyedic, three new possible Proto-Samoyedic loanwords in Proto-Turkic (PT), three new possible Old Selkup loanwords, and a new possible Old Mator loanword in Common Turkic. One of the earliest linguistic contacts of Turkic known to date was with Samoyedic. The Turkic loanwords in Proto-Samoyedic, which can be dated to the Proto-Bulgar and Proto-Common Turkic stages, are particularly important for better understanding the phonology of prehistoric Turkic. I hope that further research into Turco-Samoyedic linguistic contact will shed light on the history of the cultural interaction of these and other peoples.

## 2. Newly identified loanwords

### 2.1. Turkic → Proto-Samoyedic

PS *\*jojs* 'fat'

> PS *\*jojs* [jōs] 'Fett' [fat, grease] (SW 46) (> Enets *d′u″, jû′, jô,* Selkup тӧс /t′os/ [1879], *ćȫś*)[1] < Early Samoyedic *\*ōs* ← Yeniseian †*ōs* ← Common Turkic (CT) *ǖz* 'fat' (ED 278–279); cf. Kitāb al-idrāk li-lisān al-atrāk *öz* (read *üz*) 'yemek kaynarken üzerine çıkan yağ' [fat that rises while food is boiled] (Caferoğlu 1931: 66), Turkmen dial. ү̇змен (< *\*ǖz+män*?) 'naxar adï – nazvanie bljuda' [name of a dish] (TDGDS 179), Turkish dial. *üzlük* 'topraktan yapılmış, ufak, kulpsuz çömlek' [small, handleless earthen pot] (DS 4085), *özlük* 'topraktan yapılmış su kabı, bardak; küçük çömlek' [earthenware bowl, glass; small pot] (DS 3372), *özük* 'içi sırlı, küçük, yağ çömleği' [small, glazed oil pot] (DS 3373).

---

1. Bear in mind that these and other linguistic materials for the individual Samoyedic languages given below are not meant to be exhaustive but representative.

The Proto-Samoyedic form has no Uralic etymology. I think that it may have been borrowed indirectly from CT *ūz* 'fat'. The intermediary language apparently lacked front labial vowels and the voiced dental sibilant. This language may be identified as one of the Yeniseian languages. Yeniseian †*ōs* can be connected to Ket. *so* : *ṣⁱò* 'aus Fischdärmen gekochtes Fett' [fat cooked from fish intestines] (Donner 1955: 83) if a metathesis is assumed to have occurred. For the final consonant clusters in Proto-Samoyedic, Wagner-Nagy (2004: 93n.) rightly pointed out that "phonologisch sollen sie als die Sequenz Vokal + Halbkonsonant betrachtet werden, phonetisch könnten sie aber als Dipht[h]onge oder sogar Langvokale ausgesprochen werden".

The reason for the prothesis of the onset *\*j-* is convincingly explained by Helimski & Stachowski (1995: 42–43). The authors argued that the onset Samoyedic *\*j-* in closed syllables must be prothetic "as there are only very few stems with an initial *\*ü-* and *\*ö-* reconstructable for Proto-Sam[oyedic]." They suggested in conclusion the following change within Samoyedic: Early Samoyedic *\*ür* > Proto-Samoyedic *\*jür* 'fat (noun)'.

If PS *\*jojs* 'fat' is indeed a borrowing from Turkic *ūz* 'fat' as argued above, it would mean that *\*jür* 'Fett' [fat, grease] (SW 50) and *\*jojs* 'id' are hitherto unidentified doublets in Samoyedic. The former must have entered Samoyedic directly from Proto-Bulgar Turkic (PBT) *\*ūʀ(V̌)* whereas the latter came from Common Turkic *\*ūz* via Yeniseian.

PS *\*jujtâ-* 'dream'
  PS *\*jujtâ-* 'träumen' [dream] (SW 48) ← PT *\*ūdï-* 'sleep' > CT *ūdï-* (Khalaj *ūdi-*) 'sleep' (ED 42–43).[2]

The Proto-Samoyedic form has no Uralic etymology. Two Turkic words with long onset vowels entered Samoyedic with prothetic *\*j-*, see PS *\*jür* 'Fett' [fat, grease] (SW 50) ← PBT *\*ūʀ(V̌)* = CT *ūz* 'fat' (ED 278–279) (Helimski & Stachowski 1995: 42–43) and PS *\*jer* 'Mitte' [center, middle] (SW 43–44) ← PBT *\*ōʀă* 'center' > Chuvash *var* ~ *vară* 'seredina, serdcevina;

---

2. Against the consensus, Chuvash *śivăr-* 'sleep' can hardly be a cognate of CT *ūdï-*. The latter would be expected to yield *\*văr(ă)-* or a similar form in Chuvash. I am of the opinion that Chuvash *śivăr-* is related to CT *yavrï-* 'be(come) weak' (ED 879). The correspondence between these two forms is perfectly regular and the semantic change is understandable.

koren' (v zaprjažke)' [middle, core; root (in a harness)] (ÈSČJa I 99–100) = CT *ȫz* (Turkmen *ȫz*, Khalaj *ī͡ez*, Yakut *üös*) 'spirit, self, the interior part of an organism' (ED 278); cf. Common Mongolic (CM) *öre '(pit of the) stomach' (Nugteren 2011: 475) (see Piispanen 2018: 368 for further details). The second *j in the Proto-Samoyedic form may be seen as a substitution for the long vowel in the Proto-Turkic form.

A semantic change from 'sleep' to 'dream' or polysemy thereof is attested in many languages. To name just a few, Latin *somnus* 'sleep' > *somniāre* 'dream, daydream', Sanskrit *svápna-* 'sleep, dream', and Tamil *kaṉavu* 'dream, sleep' can be mentioned.

PS *kät 'face, forehead, front'
    PS *kät [kat] 'Gesicht' [face] (Helimski 1997: §475; SW 65) (> MTK *kāt* ~ *kāˀn* 'Stirn, Gesicht' [forehead, face], Selkup гáтын 'before, facing, in the presence of' [1879], *ḳāt* 'Vorderseite, Stirn' [front, forehead]) ← PT *kat ~ *kït > CT (mainly Oghuz) *kat* ~ *kït* 'at, near, beside, in presence of' (ED 593; Boeschoten 2022: 241–242).

The Proto-Samoyedic form has no Uralic etymology. In my opinion, PS *kät was borrowed from PT *kat ~ *kït, which mostly occurs with possessive and locative suffixes in Oghuz languages and signifies 'in presence of (a superior)'. This word cannot be identified with CT *kat* 'layer' and must be treated as a different lexeme. In her index to an 14th-century Old Ottoman mathnawi, Şan (2022: 212) rightly distinguished both lemmas: *ḳat* (I) 'nezd, yan, huzur' [near, side, presence] and *ḳat* (II) 'tabakalar veya sıralar hâlinde bulunan şeylerin her bir tabakası' [each layer of things that are in layers or rows]. CT *kat* ~ *kït* lives on in Turkish dial. *kït* 'kat, yan, ön' [presence, side, front], *kït* 'yön, yakın' [direction, vicinity] (DS 2846). The primary meaning of the word can be assumed to be 'front (side)'. The donor form likely was *kït, because PS */ä/ [a] corresponds to Turkic */ï/ in some cases such as PS *pä 'Holz, Baum, Wald' [wood, tree, forest] (Helimski 1997: §267), which, in my view, was borrowed from PT *pī̠ ~ *pā > CT *hī̠* 'vegetation, bush, tree' (ED 1; Ünal 2022a: 53).

The base of PT *kat ~ *kït may be hidden in the hapax legomenon *qa-* (in *udu qa-*) 'advance' (?), which occurs in the *Dīwān Luγāt at-Turk*, folio 547 (Dankoff & Kelly 1982–1985/II: 270); cf. Middle Korean *ká-* 'go'.

PS *kupt- ~ *kopt- 'kind of metal'

> PS *kupt- ~ *kopt- (better *kumte ?) 'kakoj-to metall' [some kind of metal] (Terent'ev 1999: 192; see also Paasonen 1912/1913–1915/1916: §332; Toivonen 1933: 102; Joki 1952: 209–210; Janhunen 1977: 125; Helimski 1982: 121–122) ← Yeniseian †kumtV ← PT *k₁ümăt₂V̆ or *k₁ümüt₂V̆ 'silver' > CT kümüš (Yakut kömüs) 'silver' (ED 723–724), Chuvash kĕmĕl 'serebro' [silver] (ÈSČJa I 273) (see also Rybatzki 1994: 211–212; Antonov & Jacques 2011).

Some scholars have linked the Samoyedic words to CT kümüš 'silver' and argued for a change of š > t in Pre-Proto-Samoyedic. This can hardly be true since Proto-Samoyedic had only contact with Proto-Bulgar Turkic at this earliest stage. PS *jer 'Mitte' [center, middle], *jür 'hundert' [hundred], *jür 'Fett' [fat], *ki 'Zobel' [sable], mɘr 'Drüse' [gland] (SW 43–44, 50, 69, 87), *kę 'winter' [winter] (Helimski 1997: §522), and *puro '(wolf-)gray' (Róna-Tas 1980: 380; 1988: 744) are clear borrowings from Proto-Bulgar Turkic. Their respective cognates in Common Turkic are ȫz 'core, center', yǖz '100', ǖz 'fat', kīš 'sable', bāz 'gland, swelling', kiš 'winter', and bōz 'gray'.

The similarity between PS *kupt- ~ *kopt- and PT *k₁ümăt₂V̆ (or *k₁ümüt₂V̆) is conspicuous.[3] However, there are serious vowel and consonant problems which speak against a direct borrowing. The Proto-Turkic form would be expected to give **küpc- or **kəpc- in Proto-Samoyedic since PT */ü/ and */t₂/ entered Proto-Samoyedic as */ü/ and */c/, respectively, cf. PT *pit₂ă- ~ *püt₂ă- 'scrape, plane (wood), rub' (> CT *hiš- ~ *hüš-) → PS *picɘ- ~ *pücɘ- 'scheren' [shave] (Ünal 2022a: 25–28). Although, as indicated by Joki (1952: 209–210) and Rybatzki (1994: 211), back-vocalic variants of CT kümüš are attested, they are all late (see Fischer 1995: 79). Only the involvement of an intermediary language lacking front vowels and dental affricates could explain this borrowing. The Yeniseian languages, in general, fit this description. If this assumption is correct, PT *k₁ümăt₂ă[4] entered Yeniseian as *kumtɛ and this form, in turn, was borrowed into Proto-Samoyedic as *kumte. This scenario also supports the reconstruction of the Proto-Turkic phoneme */t₂/.

---

3.  See Ünal (2022a; 2022b) for the reconstruction of PT */t₂/ as the source of CT /š/ and Chuvash /l/.
4.  The Turkic word is usually assumed to be of Chinese or Sino-Tibetan origin. Another possible but tentative etymology would be Pre-Proto-Turkic *künpät₂ă

PS *pät- ~ *pätɜ- 'put into the pot'

  PS *pät- ~ *pätɜ- (~ *pätɜ̂-) 'in den Topf legen' [put into the pot] (SW 118)
  ← PT *pïdă- > PT *pïd-k₂ăt₂ă > CT *hïdī̃š ~ hidīš (DLT idiš ~ iðiš, KB
  iðiš ~ iðiš, Kazakh ïdïs, Tuvan idiš, Khalaj hidīš) 'cup, vessel' (ED 72;
  ÈSTJa I 328–329).

There are two important recent articles concerning CT *hïdī̃š*. Uçar (2017)
pointed out that CT *hïdī̃š* was originally back vocalic. Nugteren (2012: 78),
on the other hand, emphasized the irregular retention of intervocalic /d/ in
modern Turkic languages and reconstructed forms like *iddiš* and *iydiš*. In
my opinion, the regular change of /d/ > /y/ is obstructed by the following /k₂/.

   Dybo & Normanskaja (2016: 48) traced the Proto-Samoyedic verb to
Proto-Uralic (PU) *pVšä- 'žarit′, gotovit′′ [fry, cook] (UEW 385: *pišä-
'braten, kochen' [fry, cook]). If this etymology is correct, the direction of
borrowing must be from Proto-Samoyedic to Turkic. However, two facts
speak against this assumption. First, PS */p-/ would be expected to en-
ter Turkic as */b-/ (see below). Second, PS */ä/ is a new phoneme which
appeared through irregular changes and with new vocabulary (Janhunen
1981: 247). In summary, I consider PS *pät- ~ *pätɜ- to be a borrowing from
PT *pïdă-, which is the basis of CT *hïdī̃š ~ hidīš.[5]

PS *pəsi̭ 'cleft, female genitalia'

  PS *pɜ̂sâ 'Riss' [cleft] (SW 114), PS *pɜ̂si̭(-) 'cunnus' [vulva] (Helimski
  1987: 59), PS *pəsi̭ 'Spalte, cunnus' [cleft, vulva] (Helimski 1997: 219)
  (> Tundra Enets posi, MTK bisigä ~ biskä) ← PBT *pū̆sŭ < PT *pūtŭ >
  CT hūt (Old Uyghur üt ~ yüt, Khalaj hī̆ᵗt) 'hole, aperture' (ED 36;
  HWAU 830, 926); cf. Mongolic *pütügün 'vagina' (Krippes 1992: 153).

The Proto-Samoyedic form has no Uralic etymology. In my opinion, PS
*pəsi̭ may have been borrowed from a hypothetical Proto-Bulgar Turkic
form *pū̆sŭ of PT *pūtŭ. For a weakening of /t/ in Bulgar Turkic, the fol-
lowing examples can be listed:

---

'sun-like' < *kün 'sun; day' (> CT kün 'sun; day', Chuvash kun 'day') + pät₂ă
'identical, equal; similar, resembling' (> CT *(h)eš 'companion, comrade; one's
equal', Chuvash yïš 'quantity, amount; group, collective' ← Old Tatar); cf. CM
*adalï 'similar, same, as' (Nugteren 2011: 265) and CT adaš 'friend, comrade'
(ED 72) for the latter component.

5. This etymology has already been presented by Ünal (2022a: 68).

(1) CT *tïlta-*, Bulgar Turkic (BT) *\*sïlta-* 'make excuses, seek pretexts'; cf. CM *\*sïlta-* 'give an excuse, employ subterfuge', Even (Arman dialect) *sïlu̧k* 'jazyk' [tongue, language] (Doerfer & Knüppel 2013: 305).

(2) CT *tïš*, BT *\*sïʟ* 'tooth' (> Chuvash *šăl*); cf. CM *\*sidün* (< *\*sil.dün* or *\*sitün*[6]) 'tooth' (Nugteren 2011: 494).

(3) CT *tint(i)-*, BT *\*sint(i)-* 'examine'; cf. Written Mongol *sinǰi-* (< *\*sindi-*) 'look over carefully, examine, observe, inspect' (Lessing 1995: 714).

(4) CT *tün ~ tün*, BT *\*sünĭ* 'night' (> Chuvash *sĕm* 'dark'?); cf. CM *\*söni ~ \*süni* 'night' (Nugteren 2011: 504), Kitan 兮 杏 [244.059] ‹s.uñ› (< *\*söni*) 'night' (Shimunek 2017: 369).

(5) CT *tāl*, BT *\*sālă* 'branch'; cf. CM *\*salaa* (~ *salasun*) 'branch; space between fingers' (< *\*sala+*) (Nugteren 2011: 481).

(6) CT *tergi*, BT *\*sergi* 'table'; cf. CM *\*siree* (< *\*siregi*) 'table' (Nugteren 2011: 497).

Benzing (1959: 712) gives a series of comparisons in favor of a change CT /t/ to /s ś/ in Chuvash. Some of these comparisons have been rightly criticized by Ceylan (1997: 57–58).

PS *\*pə̂tä* 'bile'

PS *\*pə̂tä* 'Galle' [bile] (SW 115), PS *\*pätä* 'želč'' [bile] (Helimski 1993: 130), PS *\*pəte* 'Galle' [bile] (Helimski 1997: 106, 112, 242) ← PT *\*pŏt₁ă* (~ *\*pŏt₂ă̈* ?) 'bile' > CT *hŏt* (~ *hŏš* ?) (Tuvan *höt*, Western Yugur *höt*, Yakut *üös*) 'bile, gall; the spleen, the gallbladder' (ED 35–36; Ščerbak 1976: 245; Roos 2000: 404), Chuvash *vat* 'želč'' [bile] (ÈSČJa I 102).

Since Paasonen (1912/1913–1915/1916: §329), many authors have linked (albeit hesitantly) the Proto-Samoyedic form to Finno-Ugric (FU) *\*pešä* 'grün, Kupfer' [green, copper], which is only represented by Erzya *piže* and Moksha *piža* 'green; copper, brass; young, little; blue'.[7] Lehtisalo (1929: 123) etymologized the Mordvinic forms differently and considered them unrelated to the Samoyedic forms. Lehtisalo regarded the second syllable as a derivational suffix and traced the base back to *\*p̌i-* 'young, small', which may be related to Finnish *pieni* 'small, little'. Toivonen (1933: 94–95),

---

6. Ünal (2022a: 41, 68) argued that PM *\*sidün* < Pre-Proto-Mongolic (PPM) *\*sitün* was borrowed from Proto-Turkic *\*sīt₂ŭ ~ \*sīt₂ŭ̆*.

7. These include Sammallahti (1979: 34), UEW (384–385), and Mikola (2004: 27).

however, criticized the morphological analysis and the Mordvinic semantic change from 'young, small' to 'green' assumed by Lehtisalo.

I think that PS *pəte* may be a loanword from PT *pŏ̄t₁ă*. Unfortunately, the word is not attested in Khalaj. The Proto-Turkic onset *p-* is primarily based on the Tuvan dialectal and the Western Yugur data, which may seem unreliable on their own. However, Ottoman and Turkish dialectal *ödek ~ ödük* 'coward' (TS 3049; DS 3309–3310), a derivative of *öd* (< CT *hŏ̄t*) 'bile', occurs as *hödek* and *hödük* in several Turkish dialects (DS 2427–2428). My conviction is that this parallelism between Turkish, Tuvan, and Western Yugur establishes *h-* for Common Turkic.

The only remaining problem with this etymology is that the PT *\/ŏ̄/* is substituted by *\/ə/* in Proto-Samoyedic, although the latter is reconstructed as having an *\/ö/* (Helimski 2005: 37).

## PS *pə̂t³mä* 'sharp'

PS *pə̂t³mä ~ *pə̂t³mə̂* 'scharf' [sharp] (SW 115) ← PT *piti-* (?) 'be sharp' (derivative *piti-k₂V̆*) > CT *hiti± ~ yiti±* (Old Uyghur *iti ~ y(i)ti ~ yiti*, Ottoman Turkish *iti ~ yiti*, Turkmen *yiti*, Yakut *sïtï̄*, Khalaj *yittī, yitdi, hittīg*) 'be sharp; sharp' (ED 886, 889); cf. Proto-Yeniseian *etʰə* 'scharf, spitz' [sharp, pointed] (Werner 2002/I: 273).

The Proto-Samoyedic form has no Uralic etymology. Its base may have been borrowed from the Proto-Turkic verb *piti-*, if an onset *p-* can be safely reconstructed for it. The Common Turkic data is controversial. Many forms including Yakut suggest *yiti(g) ~ *yïti(g)*, whereas some Khalaj dialects exhibit *h-* in forms such as *hittī, hittīɢ*, and *hittīg* (Doerfer 1987: §497). We know that in some cases the onset *h-* in Khalaj is secondary: *hilān* for CT *yilan* 'snake', *hiɣ-* and *hiɣïš-* for CT *yig-* 'gather, collect', *hili-* for CT *yüli-* 'shave', *hitük* for CT *yit-ök* 'lost', *hoḷɣun* for CT *yilgun* 'tamarisk' (all Khalaj data from WCh and Doerfer 1987). If *hittī, hittīɢ*, and *hittīg* represent further cases of secondary *h-* in Khalaj, we must reconstruct *yiti± ~ yïti±* for Common Turkic and so the etymology above fails.

## PS *poə̂* 'year'

PS *poə̂j* 'Jahr' [year] (SW 127), PS *pŏə̂* 'Jahr' [year] (Katz 1987: 343), PS *poa* or *pua* 'Jahr' [year] (Helimski 1997: 239) (> Tundra Nenets *po ~ pō*, Forest Nenets *pou ~ pō, pōn* 'always', Selkup *po*) ← PT *pō* 'time' (> OT *ok ~ ook*) → PPM *po.n* > CM *hon* (plural *hod*) '(the duration

of a) year' (Nugteren 2011: 359), Kitan (屶, 頗) *po, 北 ‹po› *po 'time (時)' (Shimunek 2007: 88–89; Wu & Róna-Tas 2019: 72) → Jurchenic *pon 'time, season' > Jurchen *fon* 'time' (Kiyose 1977: 101), Manchu *fon* 'time, season' (Norman 2013: 118), Xibe foɴ [fɔn] '時, occasion' (Yamamoto 1969: §2649), *fon* 'Zeit' [time] (Kałużyński 1977: 168).

The Proto-Samoyedic form has no Uralic etymology. It strongly resembles PPM *po.n* 'year, time'. This resemblance has already been noted by Krippes (1992: 141). However, direct contact between Proto-Samoyedic and Mongolic cannot be proven. PT *pō* 'time' can be reconstructed as a bridge between Samoyedic and Mongolic. PT *pō* 'time' has yielded CT *hō+k* (perhaps originally) *'a short duration of time', (later) 'opportunity, interruption', which is attested in the Kül Tegin inscription (East 2–3), Bilge Qaghan inscription (East 4) *oks(u)z* 'with no interruption' (Erdal 2004: 345), and Old Uyghur *ok ~ ook* '(günstige) Gelegenheit, Zeit; Zwischenraum' [(favorable) occasion, time; interval] (HWAU 505).[8] West Old Turkic *üd+äk > *üd+äγ*, the base of which is cognate with CT *üd* 'times' and which itself entered Hungarian as *idő* 'time, weather' (TLH 437–439), is similar in formation.

I am in favor of the reconstruction PS *pŏ̂*. Evidently, Turkic long vowels entered Samoyedic as "V + schwa" in some cases: PS *t¹å̂j* (*t¹åj*) 'branch' (> Mator (Spasskij) *taj* 'branch', Karagas *taàschita* 'leaf') ← PT *tāl* 'branch' (Helimski 1992–1993: 103).[9]

## 2.2. Proto-Samoyedic → Proto-Turkic

PT *bïlčŭ-* 'ripen'
    PS *pi-* '(durch Kochen) reifen' [ripen (by cooking)] (SW 123–124) (< PU *peji-*) → PT *bï-* > PT *bï-lčŭ-* > CT *biš- ~ biš-* (*bïša ~ bïšu*) 'come to maturity, ripen' (ED 376–377), Chuvash *piś-* 'svarivat′sja; ispeč′sja; kalit′sja' [boil, bake, heat up], etc. (ÈSČJa I 435–436).

Although Terent′ev (1999: 190) considered the Proto-Samoyedic verb a loanword from Turkic, he also noted that the former has cognates in Finno-Ugric and the latter in Altaic without giving further specifics.

---

8. Old Turkic *üd+üš* 'a period of 24 hours' from *üd* 'time' is a similar derivation.
9. This etymology has been doubted by Janhunen (1989: 8).

The Proto-Samoyedic form indeed has clear Finno-Ugric cognates and a Uralic etymology in PU *peji- 'cook' (Aikio 2022: 24).[10]

If the Proto-Turkic *bïlčü- is a reciprocal/cooperative form as assumed, its base *bï- may have been borrowed from Proto-Samoyedic or an earlier stage. The reciprocal/cooperative suffix -(X)š- has /U/ as the aorist and converb vowel. We find bïšu in the *Daśakarmapathāvadānamālā* at 03525 (Wilkens 2016/I: 400). In that case, bïš-a and bïš-ar are secondary. It remains unclear why the Samoyedic front vowel entered Turkic as a back vowel.

**PT** *kälä- 'speak'

  PS *keə̂j 'Zunge' [tongue] (SW 66)[11] < Pre-Proto-Samoyedic (PPS) *keə̂ľ (< PU *käli 'tongue') → PT *käl 'tongue' > *käl+ä- > PT *kälä- 'speak' > CT käläči ~ käläčü 'talk, conversation' (ED 716), Chuvash kala- 'govorit′, skazat′, rasskazat′' [speak, say, tell], etc. (ÈSČJa I 214–215); cf. CM *kelen[12] 'tongue; speech, language', CM *kele- (< *kelele-) 'speak' (Nugteren 2011: 409–410).

PS *keə̂j has a perfect Finno-Ugric cognate in *kēli and has been traced back to PU *käli 'tongue' (Aikio 2012: 234). I think that PT *kälä- is a derivation of *käl 'tongue' and this, in turn, is borrowed from Pre-Proto-Samoyedic *keə̂ľ, which yielded *keə̂j in Proto-Samoyedic. This borrowing also shows that CT š does not go back to a palatal or palatalized *ľ (= *l₂), otherwise we would find *käšä- in Common Turkic. PPS *ľ was simply substituted by */l/ in Proto-Turkic.

**PT** *sus- 'scoop up'

  PS *so-js- 'schöpfen' [scoop] (SW 142) (< PS *so-) → PT *sus- > CT sus- (Turkmen dial. sus-) 'scoop up' (ED 856; TDGDS 159; HWAU 632), Chuvash ăs- 'čerpat′, cedit′' [scoop, strain] (ÈSČJa I 89–90).

---

10. Alternative or outdated reconstructions are as follows: FU *pejä- 'kochen (intr.)' [cook (intr.)] (Sammallahti 1979: §126), FU ?*pexi- < PU *pexi- 'kypsyä, kiehua' [cook, boil] (Janhunen 1981: 245), and PU *peje 'gotovit′, kipjatit′' [cook, boil] (Dybo & Normanskaja 2016: 48).
11. Aikio (2012: 245) alternatively reconstructed *kää(j) 'tongue'.
12. CM *kelen is a derivation of *kele- 'speak'. CM *köke- 'suck the breast' > *köken 'breast' (Nugteren 2011: 425) presents a similar formal and semantic formation.

Terent'ev (1999: 192) already compared Turkic *sus-* with PS *\*so-*, *\*sos-*, and *\*sot-*. I think that PT *\*sus-* can be a loanword from PS *\*so-js-*, a derivative of PS *\*so-* 'schöpfen' (SW 142). The simplification of /js/ through the elision of /j/ would be natural in Turkic since such a cluster is prohibited. It is also possible that PS *\*so-js-* was realized with a long vowel as *\*sōs-*.

## 2.3. Old Selkup → Common Turkic

CT *kančĭk* 'female dog'
PS *\*węn* 'Hund' [dog] (SW 173–174) > Old Selkup *\*kʷęn* (> Selkup /kęnaŋ/ k'ana·ɢ' etc. 'Hund' [dog], /kęnaľa/ 'ščenok, kleines Hündchen' [puppy], /kęnakka/ [Alatalo 2004: §2043]) → CT *\*kan* 'dog' > CT *kančĭk* 'bitch, dog' (ED 634–635); cf. Proto-Nivkh *\*ɢanŋ* 'dog' (Fortescue 2016: 65).

CT *kančĭk* 'female dog' is derived from *\*kan* 'dog', which in turn is borrowed from Old Selkup *\*kʷęn*. This also shows that the suffix *\*+čIk* formed feminine nouns. The latter can be compared with the Mongolic feminine suffix *\*+gčin* used for forming names of female animals.

CT *kāt* 'berry'
PS *\*wotɜ* 'Beere' [berry] (SW 177) > Old Selkup *\*kʷotə* (> Selkup /kotə/ kōᴅ, kǫ̌ḍ̍, ko·ttɨ̂ etc. 'Moorheidelbeere, *Vaccinium uliginosum*, golubika' [bog blueberry] [Alatalo 2004: §1917]) → CT *\*kātă* > *kāt* 'berry' (ED 593–594), Old Uyghur *kat* 'Beere' (HWAU 342); cf. Written Mongol *qad* 'an edible wild fruit resembling a small cherry' (Lessing 1995: 902), Written Mongol *qada(n)* 'currant, *Ribes altissimum*' (Lessing 1995: 902).

Terent'ev (1989: 276) traced CT *kāt* to PS *\*kęptu* 'black currant', reconstructed as *\*kęptâ* by Janhunen (SW 66). Helimski (1995: 80) rightly disputed this etymology.

In my opinion, the older Common Turkic form *\*kātă* may be a borrowing from Old Selkup *\*kʷotə* or *\*kōtə*, if /o/ was realized as [ʌ] or similar in some idioms. According to Sammallahti (1979: §212), the Samoyedic word is related to FU *\*vosɜ* 'Beere' [berry].

CT *karga* 'crow'
> PS *wə̂r-ə̂jə̂(-)* 'Krähe' [crow] (SW 170) > Old Selkup *kʷə̂rä* (> Selkup /kuə̂rä/ 'Krähe' [crow] [Alatalo 2004: §2248]) → CT *kara* > *kara+ga* > CT *karga* 'crow' (ED 653).

Erdal (1991: 83) derived CT *karga* from CT *kara* 'black'. Although this suggestion is possible, the base may have been borrowed from Old Selkup *kʷə̂rä* 'crow'. An analogy to *kara* 'black' cannot be excluded.

## 2.4. Old Mator → Common Turkic

CT *maŋ±* 'gait; walk'
> PS *weŋkǎ* 'Schritt' [step] (SW 174) > Old Mator *meŋə-* (> Mator (Spasskij) монгалты 'šag' [step], мангыля гайтыгымъ 'stupʹju edu' [I am going to step]) → CT *maŋ* 'gait' (ED 766), CT *maŋ-* 'walk' (ED 767).

Helimski (1997: §666) established the Mator word as *meŋǝl(-)* (?) 'Schritt, ? schreiten' [step, ? stride] and remarked that "der morphologische Bau der Formen bei Spasski kann verschiedentlich interpretiert werden". Künnap (1995: 117) instead identified *-ля* in мангыля with the *l*-gerund. If this analysis is correct, the base *meŋǝ-* may be considered as the source of the Common Turkic *maŋ±*.

## 2.5. Unidentified relationship

CT *avïčga ~ abïšga ~ abušga* 'old man'
> PS *wə̂ˡjs-* 'Greis, Ehemann' [old man, husband] (SW 169–170) (> Kamas *būźe, bùźe, bùˡźɛ*) → CT *a-vïč* or *a-vïš* > CT *avïčga ~ abïšga ~ abušga* 'old man, husband' (ED 6; Boeschoten 2022: 15).

The Samoyedic form has no Uralic etymology. The Turkic word is also of unknown origin (Röhrborn 2017: 54), but its last syllable *-ga* is generally considered a suffix. If this is true, the base *avïč* or *avïš* may be a loanword from an unattested Samoyedic form like *wəś*, which comes from PS *wə̂ˡjs-*. The Common Turkic onset *a-* turns out to be prothetic. In front of a foreign /w/, this would be quite expected.

## 3. Conclusion

I presented above nine new Turkic loanwords in Proto-Samoyedic, three new Proto-Samoyedic loanwords in Proto-Turkic, three new Old Selkup loanwords, and a new Old Mator loanword in Common Turkic. Regarding the lexical borrowing between Turkic and Samoyedic, the following conclusions can be drawn from the newly identified loanwords.

The Turkic loanwords in Proto-Samoyedic preserved the Turkic ancient onset consonant *p-, which was retained in Proto-Turkic as well as in Proto-Bulgar Turkic and Proto-Common Turkic but later yielded h- and finally ʔ-. However, these loanwords cannot be assigned to any of these stages or branches with certainty since they lack distinctive consonant features. Only PS *pôsi̭(-) may be considered of Bulgar Turkic origin, because the change of t > s occurred only in that branch of Turkic.

Two Turkic loanwords entered Samoyedic through Yeniseian. The first of these loanwords, namely PS *jojs [jōs] 'fat', ultimately goes back to CT ūz 'fat' and exhibits a clear Common Turkic feature. The second one, PS *kumte 'kind of metal', on the other hand, originates from Proto-Turkic *$k_1$ümằ$t_2$ằ 'silver' (> CT kümüš, Chuvash kěměl) and supports the reconstruction of the proto-phoneme */$t_2$/. Proto-Yeniseian *kətə 'winter' is similarly traced back to Proto-Turkic *$k_1$ï̆$t_2$ằ 'winter', which yielded CT kiš and BT *kïʟ (> Chuvash xël).

Apart from the Proto-Samoyedic loanwords in Turkic, which cannot be dated and located with certainty, the later borrowings entered Turkic from Southern Samoyedic languages, namely the older stages of Selkup, Mator, and possibly Kamas.

All the above data unequivocally demonstrates the significance of Samoyedic for reconstructing earlier periods of Turkic. The first contact between Samoyedic and Turkic must have occurred near the eastern border of South Western Siberia approximately in the third or second century BCE. This means that the oldest lexical borrowings took place between Pre-Proto-Samoyedic and Proto-Turkic.

## Abbreviations

| | | | |
|---|---|---|---|
| BT | Bulgar Turkic | PBT | Proto-Bulgar Turkic |
| CM | Common Mongolic | PPM | Pre-Proto-Mongolic |
| CT | Common Turkic | PPS | Pre-Proto-Samoyedic |
| dial. | dialectal | PS | Proto-Samoyedic |
| FU | Finno-Ugric | PT | Proto-Turkic |
| MTK | Mator–Taigi–Karagas | PU | Proto-Uralic |

## References

Aikio, Ante. 2012. On Finnic long vowels, Samoyed vowel sequences, and Proto-Uralic *x. In Hyytiäinen, Tiina & Jalava, Lotta & Saarikivi, Janne & Sandman, Erika (eds.), *Per Urales ad Orientem: Iter polyphonicum multilingue, Festskrift tillägnad Juha Janhunen på hans sextioårsdag den 12 februari 2012* (Mémoires de la Société Finno-Ougrienne 264), 227–250. Helsinki: Suomalais-Ugrilainen Seura.

Aikio, Ante. 2022. Proto-Uralic. In Bakró-Nagy, Marianne & Laakso, Johanna & Skribnik, Elena (eds.), *The Oxford guide to the Uralic languages*, 3–27. Oxford: Oxford University Press. https://doi.org/10.1093/oso/9780198767664.003.0001

Alatalo, Jarmo. 2004. *Sölkupisches Wörterbuch: Aus Aufzeichnungen von Kai Donner, U. T. Sirelius und Jarmo Alatalo* (Lexica Societatis Fenno-Ugricae 30). Helsinki: Société Finno-Ougrienne.

Antonov, Anton & Jacques, Guillaume. 2011. Turkic *kümüš* 'silver' and the lambdacism vs sigmatism debate. *Turkic Languages* 15. 151–170.

Benzing, Johannes. 1959. Das Tschuwaschische. In Deny, Jean & Groenbech, Kaare & Scheel, Helmut & Togan, Zeki Velidi (eds.), *Philologiae Turcicae Fundamenta* I, 695–751. Wiesbaden: Steiner.

Blažek, Václav. 2016. On the classification of the Samoyedic languages. *Finnisch-Ugrische Forschungen* 63. 79–125. https://doi.org/10.33339/fuf.86121

Boeschoten, Hendrik. 2022. *A dictionary of Early Middle Turkic*. With the editorial assistance of John O'Kane (Handbook of Oriental Studies, Section One, Ancient Near East 169). Leiden–Boston: Brill.

Caferoğlu, Ahmet. 1931. *Abû-Hayyân: Kitâb al-Idrâk li-lisân al-Atrâk*. İstanbul: Evkaf Matbaası.

Ceylan, Emine. 1997. *Çuvaşça çok zamanlı ses bilgisi*. Ankara: Türk Dil Kurumu.

Dankoff, Robert & Kelly, James. 1982–1985. *Maḥmūd al-Kāšyarī: Compendium of the Turkic dialects (Dīwān Luγāt at-Turk)*, Part I–III (Sources of Oriental Languages and Literatures 7, Turkish Sources 7). Duxbury, Mass.: Harvard University.

Doerfer, Gerhard. 1987. *Lexik und Sprachgeographie des Chaladsch: Textband und Kartenband*. Wiesbaden: Harrassowitz.

Doerfer, Gerhard & Knüppel, Michael. 2013. *Armanisches Wörterbuch*. Nordhausen: Bautz.

Donner, Kai. 1924. Zu den ältesten Berührungen zwischen Samojeden und Türken. *Journal de la Société Finno-Ougrienne* 40. 1–42.

Donner, Kai. 1955. *Ketica: Materialien aus dem Ketischen oder Jenisseiostjaki-schen*. Bearbeitet und herausgegeben von Aulis J. Joki (Mémoires de la Société Finno-Ougrienne 108). Helsinki: Suomalais-Ugrilainen Seura.

DS = *Türkiye'de Halk Ağzından Derleme Sözlüğü* I–XII, 1993. 2. Baskı. Ankara: Türk Dil Kurumu.

Dybo, Anna Vladimirovna. 2007. *Lingvističeskie kontakty rannix tjurkov: Leksičeskij fond, Pratjurkskij period*. Moskva: Izdatel'stvo Firma Vostočnaja Literatura.

Dybo, Anna. 2014. Early contacts of Turks and problems of Proto-Turkic reconstruction. *Tatarica: Language* 2. 7–17.

Dybo, Anna Vladimirovna & Normanskaja, Julia V. 2016. Prasamodijskaja leksika material'noj kul'tury. *Linguistica Uralica* 52(1). 44–53. http://dx.doi.org/10.3176/lu.2016.1.04

ED = Clauson, Sir Gerard. 1972. *An etymological dictionary of pre-thirteenth-century Turkish*. Oxford: Clarendon Press.

Erdal, Marcel. 1991. *Old Turkic word formation: A functional approach to the lexicon, I–II* (Turcologica 7). Wiesbaden: Harrassowitz.

Erdal, Marcel. 2004. *A grammar of Old Turkic* (Handbook of Oriental Studies, Section Eight, Central Asia 3). Leiden: Brill.

ÈSČJa = Fedotov, Mixail Romanovič 1996. *Ètimologičeskij slovar' čuvašskogo jazyka*, I–II. Čeboksary: Nauka.

ÈSTJa I = Sevortjan, Èrvand V. 1974. *Ètimologičeskij slovar' tjurkskix jazykov: Obščetjurkskie i mežtjurkskie osnovy na glasnye*. Moskva: Nauka.

Filippova, Tat'jana Mixajlovna. 1974. O nekotoryx rezul'tatax sopostavlenija sel'kupskoj leksiki s tjurkskoj. In Ubrjatova, E. I. (ed.), *Voprosy jazyka i literatury narodov sibiri (Sbornik naučnyx trudov)*, 81–86. Novosibirsk.

Filippova, Tat'jana Mixajlovna. 1980a. Otraženie kontaktov sel'kupov s tjurkskimi narodami v leksike sel'kupskogo jazyka. In Timofeev, K. A. (ed.), *Aktual'nye problemy leksikologii i slovoobrazovanija (Sbornik naučnyx trudov, vypusk 9)*, 137–142. Novosibirsk: Novosibirskij gosudarstvennyj universitet.

Filippova, Tat'jana Mixajlovna. 1980b. O drevnix samodijsko-tjurkskix kontaktax. In Ubrjatova, E. I. (ed.), *Sibirskij dialektologičeskij sbornik (na materiale jazykov korennyx narodov Sibiri)*, 36–47. Novosibirsk: Nauka.

Filippova, Tat'jana Mixajlovna. 1986a. Osobennosti fonetičeskoj adaptacii tjurkskix zaimstvovanij v sel'kupskom jazyke. In Nadeljaev, V. M. (ed.), *Fonetika jazykov Sibiri i sopredel'nyx regionov*, 165–168. Novosibirsk: Nauka.

Filippova, Tat'jana Mixajlovna. 1986b. Tjurkizmy v tazovskom dialekte sel'kupskogo jazyka. In *Dialektologija i areal'naja lingvistika tjurkskix jazykov Sibiri (Sbornik naučnyx statej)*, 40–51. Novosibirsk.

Fischer, Johann Eberhard. 1995. *Vocabularium sibiricum (1747): Der etymologisch-vergleichende Anteil*. Bearbeitet und herausgegeben von János Gulya (Opuscula Fenno-Ugrica Gottingensia 7). Frankfurt am Main: Lang.

Fortescue, Michael. 2016. *Comparative Nivkh dictionary*. München: Lincom.

Helimski [Xelimskij], Evgenij A. 1982. *Drevnejšie vengersko-samodijskie jazykovye paralleli*. Moskva: Nauka.

Helimski [Xelimskij], Evgenij. 1986. Etymologica 1–48: Materialy po ètimologii matorsko-tajgijsko-karagasskogo jazyka). *Nyelvtudományi Közlemények* 88. 119–143.

Helimski [Xelimskij], Evgenij A. 1987. Pravilo Xonti dlja veng. *fészek* i ego analog v matorsko-tajgijsko-karagasskom jazyke. *Sovetskoe Finno-Ugrovedenie* 23. 57–60.

Helimski, Eugene. 1991. On the interaction of Mator with Turkic, Mongolic, and Tungusic: A rejoinder. *Journal de la Société Finno-Ougrienne* 83. 257–267.

Helimski [Xelimskij], Evgenij. 1992–1993. Etymologica 49–79: Materialy po ètimologii matorsko-tajgijsko-karagasskogo jazyka. *Nyelvtudományi Közlemények* 93. 101–123.

Helimski [Xelimskij], Evgenij 1993. Prasamodijskie *ə̑ i *ə̈: Praural'skie istočniki i nganasanskie refleksy. In Bakró-Nagy, Marianne Sz. & Szíj, Enikő (eds.), *Hajdú Péter 70 éves* (Linguistica Series A: Studia et Dissertationes 15), 125–133. Budapest: MTA Nyelvtudományi Intézet.

Helimski, Eugene. 1995. Samoyedic loans in Turkic: Check-list of etymologies. In Kellner-Heinkele, Barbara & Stachowski, Marek (eds.), *Laut- und Wortgeschichte der Türksprachen: Beiträge des internationalen Symposiums, Berlin, 7. bis 10. Juli 1992*, 75–95. Wiesbaden: Harrassowitz.

Helimski, Eugen. 1997. *Die matorische Sprache: Wörterverzeichnis, Grundzüge der Grammatik, Sprachgeschichte*. Unter Mitarbeit von Beáta Nagy (Studia Uralo-Altaica 41). Szeged.

Helimski [Xelimskij], Evgenij. 1998. Tjurkskie leksičeskie raritety v matorskom jazyke. In Stachowski, Marek (ed.), *Languages and culture of Turkic peoples* (Studia Turkologica Cracoviensia 5), 125–134. Kraków: Księgarnia Akademicka.

Helimski, Eugen. 2005. The 13th Proto-Samoyedic vowel. In Wagner-Nagy, Beáta (ed.), *Mikola-Konferencia 2004*, 27–39. Szeged: SzTE Department of Finnougristics.

Helimski, Eugen & Stachowski, Marek. 1995. Turco-Samoiedica. *Folia Orientalia* 31. 39–52.

HWAU = Wilkens, Jens. 2021. *Handwörterbuch des Altuigurischen: Altuigurisch-Deutsch-Türkisch = Eski Uygurcanın El Sözlüğü*: Eski Uygurca-Almanca-Türkçe. Göttingen: Universitätsverlag Göttingen.

Janhunen, Juha. 1977. Samoyed-Altaic contacts: Present state of research. In *Altaica: Proceedings of the 19th Annual Meeting of the Permanent International Altaistic Conference, held in Helsinki 7–11 June 1976*, 123–129. Helsinki: Suomalais-Ugrilainen Seura.

Janhunen, Juha. 1981. Uralilaisen kantakielen sanastosta. *Journal de la Société Finno-Ougrienne* 77(9). 219–274.

Janhunen, Juha. 1989. On the interaction of Mator with Turkic, Mongolic, and Tungusic. *Journal de la Société Finno-Ougrienne* 82. 287–297.

Joki, Aulis J. 1952. *Die Lehnwörter des Sajansamojedischen* (Mémoires de la Société Finno-Ougrienne 103), Helsinki: Suomalais-Ugrilainen Seura.

Kałużyński, Stanisław. 1977. *Die Sprache des mandschurischen Stammes Sibe aus der Gegend von Kuldscha, I. Band: 1. F. Muromskis Sibenische Texte, 2. Wörterverzeichnis*. Warszawa: Państwowe Wydawnictwo Naukowe.

KATZ, HARTMUT. 1987. Zur Phonologie des Motorisch-Karagassisch-Taigischen. In Rédei, Károly (ed.), *Studien zur Phonologie und Morphologie der uralischen Sprachen: Akten der dritten Tagung für uralische Phonologie, Eisenstadt, 28. Juni – 1. Juli 1984* (Studia Uralica 4), 336–348. Wien: Verband der wissenschaftlichen Gesellschaften Österreichs (VWGÖ).

KIYOSE, GISABURO N. 1977. *A study of the Jurchen language and script: Reconstruction and decipherment*. Kyoto: Hōritsubunka-sha.

KRIPPES, KARL ANTHONY. 1992. *The reconstruction of Proto-Mongolian \*p-*. Bloomington: Indiana University. (Dissertation.)

KÜNNAP, AGO. 1995. The *l*-suffix of lative and gerund in Southern Samoyed. In Künnap, Ago (ed.), *Minor Uralic languages: Grammar and lexis*, 115–122. Tartu & Groningen: University of Tartu.

LEHTISALO, TOIVO. 1929. Ein paar worte zur vertretung des ururalischen *\*š* im samojedischen. *Finnisch-Ugrische Forschungen* 20. 121–125.

LESSING, FERDINAND D. (ed.). 1995. *Mongolian–English dictionary*. Compiled by Mattai Haltod, John Gombojab Hangin, Serge Kassatkin and Ferdinand D. Lessing. 3rd reprinting with minor type-corrections. Bloomington, Ind.: Mongolia Society.

LIGETI, LAJOS. 1926–1932. [Review] Kai Donner, Zu den ältesten Berührungen zwischen Samojeden und Türken, Helsingfors 1924, 42 S. (*JSFOu* LX, 1). *Kőrösi Csoma-Archivum* 2. 187–192.

LIGETI, LAJOS. 1986. *A magyar nyelv török kapcsolatai a honfoglalás előtt és az Árpád-korban*. Budapest: Akadémiai Kiadó.

MIKOLA, TIBOR. 2004. *Studien zur Geschichte der samojedischen Sprachen*. Aus dem Nachlass herausgegeben von Beáta Wagner-Nagy. Szeged: SzTE Finnisch-Ugrisches Institut.

NÉMETH, GYULA. 1930. *A honfoglaló magyarság kialakulása*. Budapest: Hornyánszky.

NORMAN, JERRY. 2013. *A comprehensive Manchu–English dictionary*. With the assistance of Keith Dede and David Prager Branner. Cambridge (Mass.) & London: Harvard University Asia Center, Harvard University Press.

NUGTEREN, HANS. 2011. *Mongolic phonology and Qinghai-Gansu languages*. Utrecht: LOT (Landelijke Onderzoekschool Taalwetenschap). (Doctoral dissertation.)

NUGTEREN, HANS. 2012. Diagnostic anomalies? Unusual reflexes of *d in South Siberian Turkic and Western Yugur. In Erdal, Marcel & Nevskaya, Irina & Menz, Astrid (eds.), *Areal, typological and historical aspects of South Siberian Turkic* (Turcologica 94), 75–89. Wiesbaden: Harrassowitz.

PAASONEN, HEIKKI. 1912/1913–1916/1917. Beiträge zur finnischugrisch-samojedischen Lautgeschichte. *Keleti Szemle* 13. 225–277. *Keleti Szemle* 14. 20–74, 249–281. *Keleti Szemle* 15. 78–134. *Keleti Szemle* 16. 1–66. *Keleti Szemle* 17. 1–111.

PIISPANEN, PETER SAULI. 2018. Turkic lexical borrowings in Samoyed. *Acta Linguistica Petropolitana* 14(3). 357–384.
http://dx.doi.org/10.30842/alp2306573714319

RAMSTEDT, GUSTAV JOHN. 1912. Zu den samojedisch-altaischen Berührungen. *Finnisch-Ugrische Forschungen* 12. 156–157.

Röhrborn, Klaus. 2017. *Uigurisches Wörterbuch: Sprachmaterial der vorislamischen türkischen Texte aus Zentralasien, Neubearbeitung, II. Nomina–Pronomina–Partikeln, Band 2: aš – äžük*. Stuttgart: Steiner.

Roos, Martina Erica 2000. *The Western Yugur (Yellow Uygur) Language: Grammar, Texts, Dictionary*. Rijksuniversiteit Leiden. (Unpublished PhD thesis.)

Róna-Tas, András. 1980. On the earliest Samoyed-Turkish contacts. In *Congressus Quintus Internationalis Fenno-Ugristarum, Turku 20.–27. VIII 1980, Pars III: Dissertationes Symposiorum Linguisticorum*, 377–385. Turku: Suomen Kielen Seura.

Róna-Tas, András. 1988. Turkic influence on the Uralic languages. In Sinor, Denis (ed.), *The Uralic languages: Description, history and foreign influences*, 742–780. Leiden: Brill.

Rybatzki, Volker. 1994. Bemerkungen zur türkischen und mongolischen Metallterminologie. *Studia Orientalia* 73. 193–251.

Sammallahti, Pekka. 1979. Über die Laut- und Morphemstruktur der uralischen Grundsprache. *Finnisch-Ugrische Forschungen* 43. 22–66.

Şan, Funda. 2022. *Hoca Mesud'un Dilinden Ferhengnâme-yi Sa'dî: Metin-Dil Özellikleri-Günümüz Türkçesine Aktarma-Sözlük, Kopenhag Nüshası*. Ankara: Nobel.

Ščerbak, Aleksandr M. 1976. O nekotoryx spornyx voprosax rekonstrukcii fonologičeskoj sistemy tjurkskogo prajazyka. In Tenišev, Èdxjam R. (ed.), *Tjurkologičeskie issledovanija*, 237–248. Moskva: Nauka.

Shimunek, Andrew E. 2007. *Towards a reconstruction of the Kitan language, with notes on Northern Late Middle Chinese phonology*. Indiana University. (Unpublished master's thesis.)

Shimunek, Andrew. 2017. *Languages of ancient Southern Mongolia and North China* (Tunguso-Sibirica 40). Wiesbaden: Harrassowitz.

Sinor, Denis. 1947. Autour d'une migration de peuples au V^E siècle. *Journal Asiatique* 235 (1946–1947). 1–77.

Sinor, Denis. 1980. Samoyed and Ugric elements in Old Turkic. *Harvard Ukrainian Studies* III/IV (1979–1980), Part 2. 768–773.

SW = Janhunen, Juha. 1977. *Samojedischer Wortschatz: Gemeinsamojedische Etymologien* (Castrenianumin toimitteita 17). Helsinki.

Tamás, Márk. 1975. Über die türkischen Lehnwörter im Sölkupischen. *Néprajz és nyelvtudomány* 19–20. 243–252.

TDGDS = Annanurov, Ata (ed.). 1977. *Türkmen diliniŋ ġïsgača dialektologik sözlügi = Kratkij Dialektologičeskij slovar' turkmenskogo jazyka*. Ašgabat: İlim.

Tenišev, Èdxjam Raximovič. 1977. O tjurkizmax v sel'kupskom jazyke. In *Altaica: Proceedings of the 19th Annual Meeting of the Permanent International Altaistic Conference, held in Helsinki 7–11 June 1976*, 235–239. Helsinki: Suomalais-Ugrilainen Seura.

Terent'ev, V. A. 1976. Neskol'ko samodijskix ètimologij. *Sovetskoe Finno-Ugrovedenie* 12. 290–292.

Terent'ev, V. A. 1982. K voprosu o rekonstrukcii prasamodijskogo jazyka. *Sovetskoe Finno-Ugrovedenie* 18. 189–193.

TERENT′EV, V. A. 1983. Drevnejšie zaimstvovanija iz tjurkskix jazykov v samodijskie. In Matjuščenko, V. I. & Tomilov, N. A. (eds.), *Problemy ètnogeneza i ètničeskoj istorii samodijskix narodov: Tezisy dokladov oblastnoj naučnoj konferencii po lingvistike*, 67–72. Omsk: OMGU.

TERENT′EV, V. A. 1989. Drevnejšie zaimstvovanija iz samodijskix jazykov v tjurkskie. *Sovetskoe Finno-Ugrovedenie* 25. 274–280.

TERENT′EV, V. A. 1999. Drevnejšie tjurksko-samodijskie jazykovye kontakty. *Journal de la Société Finno-Ougrienne* 88. 173–200.

TLH = RÓNA-TAS, ANDRÁS & BERTA, ÁRPÁD. 2011. *West Old Turkic: Turkic loanwords in Hungarian* I–II (Turcologica 84). Wiesbaden: Harrassowitz.

TOIVONEN, YRJÖ HENRIK. 1933. Ural. *š* > sam. *t*. *Finnisch-Ugrische Forschungen* 21. 94–102.

TS = *XIII. Yüzyıldan Beri Türkiye Türkçesiyle Yazılmış Kitaplardan Toplanan Tanıklarıyla Tarama Sözlüğü*, I–VIII (1996). (2. Baskı.) Ankara: Türk Dil Kurumu.

UÇAR, ERDEM. 2017. Kutadgu Bilig dizinindeki bir kelime hakkında III: İdiş~ıdış 'kap' Üzerine. *Hikmet* 3(6). 125–132.

UEW = RÉDEI, KÁROLY. 1986–1991. *Uralisches Etymologisches Wörterbuch*. Band I–II: Unter Mitarbeit von Marianne Bakró-Nagy, Sándor Csúcs, István Erdélyi †, László Honti, Éva Korenchy †, Éva K. Sal und Edit Vértes, Band III: Register Zusammengestellt von Attila Dobó und Éva Fancsaly. Wiesbaden: Harrassowitz.

ÜNAL, ORÇUN. 2022a. *On \*p- and other Proto-Turkic consonants* (Sino-Platonic Papers 325). Philadelphia: Department of East Asian Languages and Civilizations, University of Pennsylvania.
https://sino-platonic.org/complete/spp325_proto_Turkic_consonants.pdf

ÜNAL, ORÇUN. 2022b. Is the Tocharian mule an "Iranian horse" or a "Turkic donkey"? Further examples for Proto-Turkic \*/t₂/ [ts]. *International Journal of Old Uyghur Studies* 4(2). 126–154. https://doi.org/10.46614/ijous.1206344

WAGNER-NAGY, BEÁTA. 2004. Wort- und Silbenstruktur im Protosamojedischen. *Finnisch-Ugrische Mitteilungen* 26/27. 89–109.

WCh = DOERFER, GERHARD & TEZCAN, SEMIH. 1980. *Wörterbuch des Chaladsch (Dialekt von Xarrāb)*. Budapest: Akadémiai Kiadó.

WERNER, HEINRICH. 2002. *Vergleichendes Wörterbuch der Jenissej-Sprachen*, Bde 1–3. Wiesbaden: Harrassowitz.

WILKENS, JENS. 2016. *Buddhistischen Erzählungen aus dem alten Zentralasien: Edition der altuigurischen Daśakarmapathāvadānamālā*, Teil 1–3 (Berliner Turfantexte 37). Turnhout: Brepols.

WU YINGZHE & RÓNA-TAS, ANDRÁS. 2019. Khitan studies: I. The glyphs of the Khitan Small Script, 3. The consonants, 3.1 Labial stops. *Acta Orientalia Academiae Scientiarum Hungaricae* 72(1). 47–79. https://doi.org/10.1556/062.2019.72.1.4

YAMAMOTO, KENGO. 1969. *A classified dictionary of spoken Manchu: With Manchu, English and Japanese indexes*. Ed. by Institute for the Study of Languages and Cultures of Asia and Africa. Tokyo: Institute for the Study of Languages and Cultures of Asia and Africa. Tokyo: Gaikokugo Daigaku.

# Hunting for the Uralic accusative(s)

HONTI, László. 2022. *Az ősi uráli tárgyragok története és vesszőfutása: Accusatum et expulsum*. Budapest: Tinta Könyvkiadó. 283 pp.

Professor László Honti's career has spanned several decades and despite turning 80 this year, he has not slowed down: in the 2010s and 2020s Honti has – in addition to many articles – produced several monographs, the most recent of which, *Az ősi uráli tárgyragok története és vesszőfutása: Accusatum et expulsum*, is a scrutiny of the reconstruction of the Proto-Uralic accusative. The book is a thought-provoking monograph rich in detail, and in addition to the reconstruction of the accusative of Proto-Uralic, many other issues of historical-comparative Uralic morphophonology and morphosyntax are touched upon. Also, many questions of shallower time-depth, like the development and uses of various cases in individual Ob-Ugric and Samoyedic languages, are discussed. It is impossible to address all the points of the book in one review, so here I will give a brief overview of the book and comment on its general conclusions as well as some details and claims that leave some problems open. Honti's book has also been recently reviewed by Maticsák (2022), who gives a rather detailed overview of the contents of the book and agrees with Honti's conclusions. I shall concentrate on some more problematic details here. I want to stress that my criticism of most of these points here does not mean that Honti's conclusions cannot be correct or plausible; his main points are quite well-argued. Nevertheless, for some problems discussed, different solutions are available. As is well known to specialists in Uralic studies, the historical-comparative reconstruction of Uralic morphology and morphosyntax is complicated and cannot be said to be on the same advanced level as that of phonology (cf. Aikio 2022: 3), and regarding the reconstruction and history of the object cases in various Uralic languages and in Proto-Uralic, much work remains to be done.

The research problem of the monograph, the reconstruction of the accusative in nominal and

pronominal paradigms in Proto-Uralic, is a question that has been frequently discussed in Uralic linguistics, but a consensus is lacking. Although the reconstruction of the *m accusative is universally accepted, the accusative of the personal pronouns in Proto-Uralic is a disputed issue. The issue has been discussed by, among others, Salminen (1997: 131) and Janhunen (2020: 387). In his recent handbook chapter on Proto-Uralic, Aikio (2022: 13) notes suppletive oblique case forms of personal pronouns in Hungarian, Mansi, and Samoyedic with diverse backgrounds, but he does not comment on the reconstruction in more detail.

The book opens with an introduction that involves an overview of research history and discussion of Proto-Uralic case system and numeral inflection (here not all the reconstructed cases are dealt with, however, despite the Proto-Uralic case system receiving a subchapter of its own). This is followed by the main part of the book that consists of the presentation and discussion of accusatives and other cases, and their use and origins in different branches of Uralic (Finnic, Saami, Mordvin, Mari, Permic, Ob-Ugric, Samoyedic). The inflection of personal pronouns is also presented in detail. The history of various Uralic cases as morphemes receives much attention; the problems in the reconstruction of the function of the plural suffix *j in Proto-Uralic is discussed at length. Also, the origin and development of the objective conjugation in the Uralic languages is addressed. Many of these questions have been discussed by Honti in previous works and he refers copiously to his earlier research results, in some case also modifying some of his previous views.

Honti also discusses the uncertainties in the reconstruction of the accusative and other cases in the plural and dual in Proto-Uralic. The Proto-Uralic suffix *j receives a detailed treatment. This *j is reflected in Saami as the marker of the plural genitive and also as part of the plural suffix of other cases, and in Samoyedic as the plural accusative, and these are commonly assumed to represent Proto-Uralic inheritance, also reflected as the *i plural marker of the oblique cases in Finnic. (The Hungarian *i* in possessive plural forms like *fiai* 'his/her sons' is often assumed to reflect the same Proto-Uralic suffix – see Abondolo 1998: 21; Aikio 2022: 15 – but Honti does not discuss this possible Hungarian reflex.) The reconstruction of the exact function of this *j in Proto-Uralic has a long history; Salminen (1996: 27) assumed simply that *j was the marker of the oblique plural case form in Proto-Uralic. Honti concludes

here that *j was originally probably a collective suffix (*gyűjtőnévképző*).

Regarding Honti's detailed presentation of the accusatives and object cases in the various Uralic languages, one can note that in what detail the different "languages" or "dialects" are discussed varies. For example, South Estonian is not discussed separately in detail, despite its taxonomic importance, and the Saami languages are called "dialects" (*lapp nyelvjárások*) contrary to modern views and conventions in Uralic linguistics. Many Saami languages are discussed, but sometimes the naming of the languages is misleading and also varies. We find "Eastern Saami/Lapp" (*keleti lapp*), and North Saami is sometimes "northern Saami" (*északi lapp*), sometimes "Norwegian Saami" (*norvég lapp*) – this reflects the naming in different sources that are cited, but a non-specialist reader interested in, for example, the origin of the Hungarian accusative suffix might be easily led astray. The presentation of the Samoyedic languages is fraught with similar issues: Nenets, Enets, and Selkup are dealt with as single languages, and the forms in different "dialects" (such as Tundra Nenets and Forest Nenets) are not presented but simply "Nenets" paradigms are given (some Forest Nenets phenomena are briefly discussed

later in the book but mostly Honti speaks of simply *jurák* (Nenets), *jenyiszei* (Enets), etc.).

The higher levels of taxonomy, such as Proto-Ugric, are commented on to some extent, but there is no overview of the reconstruction of the accusative or other cases for these intermediary proto-languages. However, Honti seems to operate according to the traditional model of Uralic taxonomy, and also North and South Samoyedic exist as taxonomic units in Honti's treatment, despite the fact that these taxonomic nodes are disputed, cf. Janhunen (1998: 458–459).

The monograph ends in a rather brief chapter presenting conclusions, followed by a summary in German and the bibliography. The German summary presents the main contents and conclusions of the book in very informative way, but the concluding chapter does not really give a proper overview and one needs to browse through the book if one wants to know Honti's main arguments. The debated issues on the reconstruction of the Proto-Uralic accusative as well as the numeral markers are discussed in more detail in the first chapter of the book.

The basic conclusion that Honti reaches is that in Proto-Uralic, nouns and personal pronouns had different endings for accusatives, and

languages like Finnish and Khanty that show a similar *t accusative in personal pronouns reflect an archaic feature retained from Proto-Uralic. Honti also concludes that traces of the *t accusative are found in Mansi, Forest Nenets, and Selkup. The Hungarian *t*-accusative, for which several different origins have been suggested, is according to Honti generalized from the pronominal accusative. It is interesting that Zhivlov (2023: 153–154) has suggested a similar view very recently and this possibility is also mentioned by Abondolo & Valijärvi (2023: 208–209) in the same handbook (this book was published around the same time as Honti's book, so mutual references in either direction would have been impossible). Honti also assumes that in addition to the personal pronouns, *t might have been used as the accusative of animate nouns in Proto-Uralic (according to Honti, the *t* accusatives used of persons like *Kallet*, *Kertut* attested in some Finnish dialects might be relicts of such use).

Although the structure of the book is clear and it is easy to find information, there is unfortunately one major typesetting error: pages 61–62 are printed twice. Moreover, there are a couple of spelling mistakes. Here I will list points of criticism concerning the contents of the book, in addition to the issue of

glottonyms and taxonomical units mentioned above.

Although Honti describes the research history in detail and also quotes the views of many earlier works (including also the most important recent resources), some relevant recent references related to the reconstruction of the Uralic case system are missing (Salminen 1996; Ylikoski 2016; Janhunen 2020). This means that some counterarguments to Honti's claims are not taken into account.

A notable issue concerning the conclusions drawn by Honti, the reconstruction of nominal *m and pronominal *t accusatives for Proto-Uralic, is the origin of the inflection of personal pronouns in Hungarian and Mansi, namely the use of possessive suffixes in the accusative forms (*engëm* 'me (1SG.ACC)', *tégëd* 'you (2SG.ACC)'). Honti discusses the system in Hungarian and Mansi, but he assumes that the system might go back to Proto-Ugric, not considering the possibility that this might have been the Uralic system. As Honti refers to Helimski's (1982) discussion of the background of the Hungarian and Mansi pronouns, it would have been good to analyze the possibility of a Proto-Uralic origin of this system, especially as this idea has been supported by Salminen (1996: 26), a publication which is missing from Honti's references.

Honti does not comment on the idea expressed by Helimski that the suppletive inflection of personal pronouns in Samoyedic with possessive suffixes reflects a similar Proto-Uralic system. I am not claiming here that the idea of Helimski and Salminen is correct, but this issue should be addressed in more detail before the conclusions on the pronominal *t accusative can be accepted.

The issue of the origin of the *g* in Hungarian *engëm*, *tégëd* is a major, debated issue (see for example Abondolo 1998; Abondolo & Valijärvi 2023) and it is also discussed by Honti. Helimski (1982: 95–97) observes a similar system in Samoyedic and assumes that Proto-Uralic *ke ~ *ki is reflected by Hungarian *g* and the form *ket (in the modern Proto-Samoyedic reconstruction) appearing in the suppletive paradigm of the Northern Samoyedic personal pronouns. It remains unclear why Honti does not accept the common origin of Hu *g* and the Samoyed element with *k. Honti refers to Aikio's (2006) idea that the element *ket found in the Samoyed forms is a reflex of Proto-Uralic *keti 'skin' that has cognates in Finnic, Saami, and Mordvin. Honti discusses possible Permic and Ob-Ugric connections of this Proto-Uralic noun, but these are not very convincing in the light of historical

phonology; it is also difficult to understand Honti's comment that he has established a "new" Proto-Uralic word family when his etymology includes the same cognates already mentioned by Aikio.

In his 2021 paper Honti discussed the issue of Proto-Uralic *keti in more detail and did not categorically rule out the possibility that Hungarian *g* might be in some way related to this Proto-Uralic word. There are no major phonological problems in Helimski's explanation of *engëm* from *ämVŋkVm or the like, but if the *kV element is related to *keti, the loss of *t would be admittedly problematic and would require more phonological research. This issue has not been, to my knowledge, properly addressed anywhere.

However, regardless of the origin of the *g and its relation to Samoyed *ke(t), it is true that Hungarian, Mansi, and Samoyedic do show a similar pattern of inflection of the personal pronouns. This distribution is obviously limited, but not more limited than the spread of the *t accusative, and it should be considered possible that this system reflects the most archaic Proto-Uralic situation, as has been argued by, for example, Salminen (1996: 26). It is true that in this case the origin of the *t accusative in Finnic and Khanty (as well as Hungarian)

would remain obscure. Obviously the situation with Hungarian first and second personal plural accusatives *minket, titeket* with the compulsory -*t* accusative (contrary to *engëm, tégëd* with only possessive suffixes) is a problem in this explanation – the issue is not properly addressed by Helimski (1982).

A related issue concerns the reconstruction of Proto-Uralic 2SG possessive suffixes, where Hungarian shows a voiced *d*, similarly as with the ending of the objective conjugation. The Proto-Uralic accusative 2SG ending is reconstructed as *mti (see, for example, Janhunen 1982; Salminen 1996), which would regularly yield Hu *d*, but Honti prefers to explain the Hungarian *d* through sporadic voicing from *-*t*. The origin of the Hungarian *d* is a debated issue, but the possibility to derive it from *mti has gained some support (see Abondolo 1998). Honti's reluctance to accept *mti > *d* is due to the morpheme order in Ugric: as the Ugric languages show predominantly a morpheme order of case suffix (Cx) + possessive suffix (Px), Honti argues that the accusative suffix *m could not have preceded the possessive suffix in Pre-Hungarian. The "Finnish-type" morpheme order Px+Cx for Proto-Uralic has been supported by several scholars (for example, Nichols 1973: 234–235;

Janhunen 1982: 33; 2020: 388; Raun 1988: 561; Aikio 2022: 16) but Honti has argued also earlier (see e.g. Honti 2009) that Proto-Uralic had both orders, like Mari and Permic do today, and the unitary system in branches like Finnic is an innovation. His arguments are not entirely clear, for example Honti (2009: 174) notes that some Hungarian postpositions showing the "old" morpheme order show only "veraltete Kasussuffixe", but it is difficult to understand why this would be a valid counterargument to assume the Px+Cx order for Proto-Ugric and Proto-Uralic. It would be difficult to explain the lack of traces of the "Ugric" type of order in Finnic, Saami, etc. if Proto-Uralic had both systems. Even if Proto-Uralic would have had both systems, as Honti claims, in the system of possessive declination the accusative marker would have in all probability come first, as an ending like 2SG *mti can be reliably reconstructed based on comparative evidence, so the idea that *d* in the Hungarian 2SG endings reflects a cluster *mt is totally acceptable.

Also, Honti addresses the origin of the Proto-Samoyed coaffix *kə in his discussion of the history of the Samoyedic case system. It is true that there seems to be no commonly accepted origin of the Samoyedic coaffix (for example, Janhunen's

1998 presentation of Samoyedic does not comment on the issue). However, Mikola's old proposal has been criticized in more recent research and some alternatives have been suggested. Honti assumes that *kə could be derived from Proto-Samoyedic *kəj 'side', but he does not provide phonological arguments to back up this idea.

It should be noted that Ylikoski (2016: 47–48, 61) has criticized the lative origin and suggests that the possible connection of the Samoyedic coaffix with the Mari postposition *gə̂č̌* should be investigated. It is also good to note here that Gusev (2018) has also criticized the lative explanation in a conference presentation that is available online. He assumes a derivational origin of both Samoyedic coaffixes *-kə and *-ntə.

Although the lative explanation is indeed problematic, Honti's new explanation lacks phonological details; the difference in vocalism cannot be easily explained and the solution must be considered rather hypothetical. Of course, also Ylikoski's recent remark on the possible connection of the Samoyedic coaffix and the Mari postposition is an idea that also needs further phonological scrutiny. The origin of the Samoyedic coaffix remains unclear for the time being.

Despite these points of criticism, Honti's monograph is a major contribution to historical-comparative Uralic morphology and his ideas and conclusions will certainly inspire and provoke discussion on many details of Uralic case system – and yielding further discussion is the main task of research. Honti presents interesting conclusions and future work on Uralic case systems will show whether he was on the right track.

*Sampsa Holopainen*

## References

Abondolo, Daniel. 1998. In Abondolo, Daniel (ed.), *The Uralic languages*, 1–42. London & New York: Routledge.

Abondolo, Daniel & Valijärvi, Riitta-Liisa. 2023. Introduction to the Uralic languages, with specific reference to Finnish and Hungarian. In Abondolo, Daniel & Valijärvi, Riitta-Liisa (eds.), *The Uralic languages*, 1–80. 2nd edn. London & New York: Routledge.

Aikio, Ante. 2006. New and old Samoyed etymologies (Part 2). *Finnisch-Ugrische Forschungen* 59. 9–34.

Aikio, Ante (Luobbal Sámmol Sámmol Ánte). 2022. Proto-Uralic. In Bakró-Nagy, Marianne & Laakso, Johanna & Skribnik, Elena (eds.), *The Oxford guide to the Uralic languages*, 3–27. Oxford: Oxford University Press. https://doi.org/10.1093/oso/9780198767664.003.0001

Gusev, Valentin. 2018. *Co-affixes of local cases and locative adjectives in Samoyed: Presentation at the 7th International Conference on Samoyed Studies, University of Tartu, 26–27 October 2018.* https://sisu.ut.ee/sites/default/files/samoyedology2018/files/gusev_tartu-2018_presentation.pdf

Helimski 1982 = Хелимский, Е. А. 1982. *Древнейшие венгерско-самодийские языковые параллели (Лингвистическая и этногенетическая интерпретация).* Москва: Наука.

Honti, László. 2009. Die Morphemreihenfolge im Uralischen. *Plurilinguismo* 16. 165–176.

Honti, László. 2021. Az *engëm(et)* és a *tégëd(et) g*-jének rejtélyéről. *Folia Uralica Debreceniensia* 28. 131–152.

Janhunen, Juha. 1982. On the structure of Proto-Uralic. *Finnisch-Ugrische Forschungen* 44. 23–42.

Janhunen, Juha. 1998. Samoyedic. In Abondolo, Daniel (ed.), *The Uralic languages*, 457–489. London & New York: Routledge.

Janhunen, Juha. 2020. Grammaticalization in Uralic as viewed from a general Eurasian perspective. In Bislang, Walter & Malchukov, Andrej (eds.), *Cross-linguistic variation and universal tendencies*, 361–398. Berlin & Boston: de Gruyter.

Maticsák, Sándor. 2022. [Review of] Honti, László 2022: *Az ősi uráli tárgyragok története és vesszőfutása: Accusatum et expulsum. Folia Uralica Debreceniensia* 29. 227–233. https://finnugor.arts.unideb.hu/fud/fud29/15-Ismerteteseko1.pdf

Nichols, Johanna. 1973. Suffix ordering in Proto-Uralic. *Lingua* 32. 227–238.

Raun, Alo. 1988. Proto-Uralic comparative historical morphosyntax. In Sinor, Denis (ed.), *The Uralic languages: Description, history and foreign influences* (Handbook of Oriental Studies: Section 8 Uralic & Central Asian Studies, Volume 1), 555–571. Leiden & Boston: Brill.

Salminen, Tapani. 1996. Comments on Lázló Honti's paper "Zur Morphotaktik und Morphosyntax der uralischen/finnisch-ugrischen Grundsprache". In Leskinen, Heikki & Seilenthal, Tõnu (eds.), *Congressus Octavus Internationalis Fenno-Ugristarum, Pars 8: Oratio plenaria et commentarii & symposia & orationes officiales*, 25–27. Jyväskylä.

Salminen, Tapani. 1997. *Tundra Nenets inflection* (Mémoires de la Société Finno-Ougrienne 227). Helsinki: Société Finno-Ougrienne.

Ylikoski, Jussi. 2016. The origins of the western Uralic *s*-cases revisited: historiographical, functional-typological and Samoyedic perspectives. *Finnisch-Ugrische Forschungen* 63. 6–78.

Zhivlov, Mikhail. 2023. Reconstruction of Proto-Uralic. In Abondolo, Daniel & Valijärvi, Riitta-Liisa (eds.), *The Uralic languages*, 117–175. 2nd edn. London & New York: Routledge.

# The largest bidirectional dictionary of North Saami and Swedish to date

Svonni, Mikael. 2023. *Davvisáme-giela-ruotagiela, ruotagiela-davvi-sámegiela sátnegirji = Nordsamisk-svensk, svensk-nordsamisk ordbok.* Kiruna: Ravda Lágádus. 503 pp.

For decades, Mikael Svonni has played an instrumental role in the promotion and development of the North Saami language in Sweden. As both a native speaker and an emeritus professor of Saami linguistics, he has published not only scholarly papers but also editions of older North Saami texts, as well as translations of fiction for children and adults. Recently, he has also published grammars of the language, written in North Saami (2015, second edition 2019) and Swedish (2018); for reviews of these works, see Julien (2016), Kejonen (2019), and Siegl (2020).

For the general Saami public, however, Svonni is without a doubt best known for his pedagogical North Saami–Swedish, Swedish–North Saami dictionaries (1990, 2013, app version 2017), preceded by a North Saami–Swedish word list (1984). Found in the homes, schools, and workplaces of North Saami learners and speakers throughout the country, the products of Svonni's lexicographic work are used daily by teachers, students, journalists, translators, and writers. The book under review is thus a welcome update of the author's 2013 dictionary.

With hard covers featuring a scenic mountain landscape, the book consists of three parts: front matter (pp. v–xi, in both North Saami and Swedish), a North Saami–Swedish section (pp. 13–267), and a Swedish–North Saami section (pp. 269–503). There is no list of abbreviations. The most apparent difference between the book at hand and its predecessors is its volume: the number of North Saami lemmas is approximately 18,000, and the number of Swedish lemmas tally 14,000 (to be compared to the 15,000 North Saami and 10,000 Swedish lemmas of the 2013 version). While this makes the present edition by far the most comprehensive dictionary of North Saami aimed at an audience in Sweden, it does not compare in scope to the North Saami–Norwegian dictionary by Kåven et al. (1995) with 35,000 lemmas, nor to the North Saami–Finnish dictionary by Sammallahti (2021) with its 53,000 lemmas. That said, the book is likely to cover the needs of most students and teachers, the North Saami section ranging from

*abessiva* 'abessive' to *Álánda* 'Åland' and the Swedish section ranging from *abborre* 'perch' to *övrig* 'other'.

In adding new words to the dictionary, the author has been guided by the Swedish word list SAOL (*Svenska Akademiens ordlista*). An explicit purpose has also been to include "modern" vocabulary, in addition to words pertaining to traditional Saami culture. In practice, this means that apart from words like *nulla* 'reindeer doe that has shed its antlers' and *guldohallat* 'be covered in snow (due to a snowstorm)', the book also includes lemmas like *kiive* 'kiwi' and *mánnávávdno* 'stroller'.

The variety of North Saami featured in the dictionary is the standard literary language, which in several respects differs from the varieties spoken in Sweden, i.e. the eastern Čohkkiras (Jukkasjärvi) and Gárasavvon (Karesuando) dialects. While the book contains some words characteristic of one or both of these varieties (e.g. *ruomas* : *ruopmas-* 'wolf', *ubbá* 'kiss'), the selection of lemmas could be described as normative. In particular, many local variants of words with counterparts in the standard language have not been included, such as *akta* 'one', *čoaivi* 'stomach, belly', and *skovlo* 'school' (standard *okta*, *čoavji*, *skuvla*), all featured in the North Saami–Swedish section of the 2013 edition. Occasionally, non-standard forms have

nevertheless found their way into the dictionary, e.g. *čoargi* 'goldeneye' (standard *čoađgi*).

Several loanwords included in the 2013 book have also been left out from the present edition, e.g. *boatkat* 'kick', *noját* 'lean' (< Finnish/Meänkieli *potkia*, *nojata*), and *vearál* : *vearáld-* 'world' (cf. Swedish *värld*, Old Norse *verǫld* etc.). An effect of this normative stance is, ironically, that Sammallahti's (2021) North Saami–Finnish dictionary in some cases represents the varieties of North Saami spoken in Sweden better than the work under review, for instance including the common borrowings *eidde* 'mom' and *isá* 'dad' (< Finnish/Meänkieli *äiti*, *isä*) with the remark that they are used in Sweden. It should however be pointed out that Svonni's dictionary does also contain loanwords not found in Sammallahti (2021), such as *uŋgá* 'bear cub (yearling)' (cf. Swedish/Norwegian *unge* 'offspring, kid'), *lávddis* : *lávdás-* 'plate', and *mustet* : *musteg-* 'blueberry' (< Finnish/Meänkieli *lautanen* : *lautas-*, *mustikka*).

A number of lemmas in the book under review are to my knowledge not found in other North Saami dictionaries, with the exception of Svonni's previous works. Such words, many of which have equivalents in the neighboring Lule Saami language, make the dictionary valuable for scholars. The adjective

*ovddis* : *ovdás-* 'whole (not broken)', for instance, has cognates throughout the western branch of Saamic (e.g. Eastern Čohkkiras North Saami *ogddis* : *ogdás-*, Lule Saami *åbdes* : *åbddås-*, South Saami *emties*), allowing for the reconstruction *\*omtēs*. A form *ovddis* must also be the source of the derived verb <*ovdasmam*> 1SG.PRS 'become whole, after having been apart' in the North Saami dictionary of Leem (1768–81), thus providing an etymology for this previously obscure word.

An important task of any bilingual dictionary is the presentation of lemmas, grammatical information, and translations. For a pedagogical dictionary, such as the one under review, there is a difficult balancing act between comprehensiveness, on the one hand, and accessibility, on the other. Svonni succeeds well in keeping lemmas succinct and translations pithy, while also providing necessary information on inflection. Both in the North Saami–Swedish and in the Swedish–North Saami section, each lemma is followed by morphological data and an indication of word class, as well as one or several translations into the other language, similarly accompanied by clear and relevant data on inflection.

Data on morphosyntax, such as the transitivity of verbs or the governing of specific cases by certain lemmas, is not made explicit in the dictionary, although for some headwords short example clauses make it possible to infer this information. For instance, the clause *mii ballat sus* 'we are afraid of him' shows that the verb *ballat* 'be afraid' governs the locative case. Such examples are much appreciated by learners of a language, and in my view, more of them could have been included. In fact, a number of example clauses given in Svonni (2013) are excluded in the present edition, e.g. under the headword *liikot* 'like', leaving it to the reader to figure out that this word governs the illative case in the standard language.

A sometimes overlooked possibility of dictionaries is the opportunity to indicate phonological information that is not apparent in the practical orthography of a given language. Svonni makes use of this possibility, albeit in a somewhat unconventional way. In North Saami, consonants make a phonological three-way length contrast, distinguishing short, long, and overlong consonants. In the official orthography, however, this contrast is not indicated, conflating long and overlong consonants. In many dictionaries and other pedagogical materials, this ternary contrast is specified by means of marking overlong consonants with the so-called *deattamearka* (or *deaddomearka*, lit. 'stress mark') <'>, e.g. *beas'si* 'birch bark' vs. *beassi* 'nest; oven'.

In Svonni's dictionary, on the other hand, overlong consonants are not overtly indicated as such. Rather, their length can be deduced from the presentation of their gradation pattern, e.g. *beassi, ss* 'birch bark' vs. *beassi, s* 'nest; oven'. While this clever way of marking the difference between long and overlong consonants has the benefit of staying close to the standard orthography, there is still a need to indicate the length of consonants also in non-gradating words. For such lexemes, overlong consonants are indicated by an apostrophe in square brackets after the lemma in question, e.g. *orru* [-r'r-] 'inhabitant'. I cannot help but wonder whether it would not have been easier to consistently indicate overlength in the lemmas themselves, provided of course with a note that this sign is left out in most North Saami texts. Notably, there are also instances where an overlong consonant is not indicated where it should be, e.g. in the words *oarri* 'squirrel' and *girječálli* 'author'.

Another drawback of the official North Saami orthography is that it does not distinguish between short /i, e, u, o/ and long /iː, eː, uː, oː/. In some dictionaries this contrast is indicated by a macron over the long vowels <ī, ē, ū, ō>, but as in the case of overlong consonants, Svonni indicates this phonemic distinction by way of square brackets after the lemma, e.g. *giron* [giː-] 'rock ptarmigan'.

Here too, however, I often miss an indication of length, e.g. for the first vowels of *firon* 'whisk', *rebeš* 'fox', *vulos* 'down, downwards', and *doron* 'fighter, thug, troublemaker'. Also in Swedish entries, the indication of vowel length is somewhat arbitrary, e.g. specifying a long first vowel in *förstuga* [föːr] 'hall' but not in *förskola* 'preschool'.

As in any work of this scope, there are bound to be some minor flaws. The headword <*muottál*> 'sister's daughter, sister's son' should, for instance, read *muottal*, the derivational suffix of both <*heaboheabmi*> (s.v. *skamlös*) 'shameless' and <*fuolaheape*> (s.v. *sorglös*) 'careless' should read *-heapme* (or *-heapmi*) '-less', the preposition *ilmmá* 'without' (< Finnish/Meänkieli *ilman*) is lacking among the word entries despite being used in the preface, and the verb *rávásmuvvat* 'age, mature (about people)' is inexplicably provided as a translation of Swedish *övrig* 'other'. Learners of North Saami would likely have appreciated the inclusion of the Swedish lemmas *ja* 'yes' and *nej* 'no' (to which the interjection *juo* and a negative auxiliary verb, respectively, correspond). It can also be noted that the title of the book is given as *Davvisámegiela-ruotagiela…* 'North Saami–Swedish…' on the front cover and on the second title page, but as *Davvisámegiela-ruotagiel…* on the first title page. Such inconsistencies do

not, however, detract significantly from the usefulness of the dictionary, and it is only to be hoped that this print edition will soon be followed by a digital version, much as the author's 2013 dictionary was followed by a widely used app in 2017.

Perhaps more than any other branch of linguistics, lexicography is a cumulative discipline. With the publication of the book under review, Mikael Svonni expands on his previous work, creating a user-friendly resource for teachers, students, and others who come in contact with North Saami in Sweden. Scholars will also be happy to find a number of words not attested in other dictionaries of the language. Together with the grammars by the same author, Mikael Svonni's latest dictionary crowns a life's work dedicated to the maintenance of the North Saami language.

*Olle Kejonen*

### References

JULIEN, MARIT. 2016. [Review of] Mikael Svonni, *Davvisámegiella – sánit ja cealgagat* [sic]. *Láidehus sámi lingvistihkkii*, Giron/Kiruna: Ravda lágádus 2015. *Journal of Northern Studies* 10(2). 217–229.

KÅVEN, BRITA & JERNSLETTEN, JOHAN & NORDAL, INGRID & EIRA, JOHN HENRIK & SOLBAKK, AAGE. 1995. *Sámi-dáru sátnegirji: Samisk-norsk ordbok*. Karasjok: Davvi Girji.

KEJONEN, OLLE. 2019. [Review of] Svonni, Mikael. 2018. *Modern nordsamisk grammatik*. Kiruna: Ravda lágádus. *Multiethnica* 39. 68–70.

LEEM, KNUD. 1768–81. *Lexicon Lapponicum bipartitum, Lapponico-Danico-Latinum & Danico-Latino-Lapponicum, cum Indice latino*.

SAMMALLAHTI, PEKKA. 2021. *Sámi-suoma sátnegirji: Pohjoissaame – suomi -sanakirja*. Karasjok: Davvi Girji.

SIEGL, FLORIAN. 2020. What makes a grammar a modern grammar? Review of Mikael Svonni, *Modern nordsamisk grammatik*, Kiruna:

Ravda Lágádus, 2018. *Journal of Northern Studies* 14(1). 83–94.

SVONNI, MIKAEL. 1984. *Sámegiel-ruotagiel skuvlasátnelistu: Samisk-svensk skolordlista*. Jokkmokk: Sámiskuvlastivra.

SVONNI, MIKAEL. 1990. *Sámi-ruota, ruota-sámi sátnegirji: Samisk-svensk, svensk-samisk ordbok*. Jokkmokk: Sámi Girjjit.

SVONNI, MIKAEL. 2013. *Davvisámegiela-ruotagiela, ruotagiela-davvisámegiela sátnegirji: Nordsamisk-svensk, svensk-nordsamisk ordbok*. Karasjok: ČálliidLágádus.

SVONNI, MIKAEL. 2015. *Davvisámegiella – sánit ja cealkagat: Láidehus sámi lingvistihkkii*. Kiruna: Ravda Lágádus.

SVONNI, MIKAEL. 2017. *Sátnegirji*. Ravda Lágádus. (Apple/Android app.)

SVONNI, MIKAEL. 2018. *Modern nordsamisk grammatik*. Kiruna: Ravda Lágádus.

SVONNI, MIKAEL. 2019. *Davvisámi giellaoahppa*. Kiruna: Ravda Lágádus.

# A North Saami dialect dictionary in a new format

Skåden, Asbjørg (ed.). 2022. *Márkku sánit*. Ravda Lágádus. (Android app.)

North Saami is spoken in a vast area covering northernmost Norway, Sweden, and Finland. This area, comparable to Hungary in size, is home to considerable dialectal variation. While it is understandable that the North Saami literary standard language is based on the most populous of these varieties – the so-called Finnmárku (Finnmark) or Inland dialects – this also has the side effect that other varieties are seldom represented in grammars, dictionaries, or teaching materials.

Among these marginalized dialects is the western variety of Čohkkiras (Jukkasjärvi) North Saami, spoken in the border area of Nordlánda (Nordland) and the former county of Romsa (Troms) in Norway. From the 1980s onwards, a small group of language activists have worked on the revitalization of this critically endangered dialect, sometimes referred to as *márkosámegiella* (or *markasamisk* in Norwegian). A significant milestone in this effort is the publication of the thematic North Saami–Norwegian dialect dictionary *Márkku sánit* in

2010, edited by teacher and publisher Asbjørg Skåden (1946–2020) and based mainly on recordings made in the 1980s and 1990s. Unfortunately, this primary data was lost in a house fire in 2017 (p.c. Asbjørg Skåden, 2018), leaving the edited dictionary as a monument of this important work.

This dictionary, which has been out of print for some time, has now been republished as a free app for Apple and Android devices as a result of a cooperation between the local Saami language center Várdobáiki and the publisher Ravda Lágádus. While this digital version largely corresponds to the printed dictionary, the app has several advantages compared to its paper predecessor. Most importantly, it provides North Saami lemmas in alphabetical order in addition to the thematic grouping of entries, as well as the possibility to search for words both in North Saami and Norwegian.

Unlike some other digital dictionaries of North Saami, the app does not require – or even allow – the use of special characters in search queries. Instead, a search for the letter *c*, for instance, yields results for both *c* and *č*. It is also possible to search for parts of words

using a percent sign as a wildcard, although this information is not conveyed in the app itself. A minor drawback of the app is that it lacks the background information found in the front and back matter of the 2010 book, including names of the contributors to the project. Also omitted are the photographs and illustrations of the print edition.

The number of words and expressions in the dictionary is hard to estimate, in part due to some lemmas being listed under several semantic fields. However, the total number of entries, including idioms, is well above 5,000. The app covers many different domains, ranging from body parts and kinship terms to names of insects and units of measurement. The editor deserves praise for the inclusion of not only single lexemes but also greetings, proverbs, and nursery rhymes, as well as curse words and obscenities. Parents-to-be and scholars of onomastics will be equally delighted to find many proper names included, although first names are only found in the thematic section of the app and not in the alphabetical listing of lemmas. Also very welcome is a section on discourse markers ("småord i en samtale"). The use of many words is further shown through example sentences, albeit often without accompanying translations. All in all, this makes the app potentially useful for a wide audience, ranging from local language learners to seasoned scholars. Members of the latter group will in particular appreciate the inclusion of a number of words that appear to be unattested in other North Saami dictionaries, e.g. *suovvu* 'widening in the mouth of a river', *veavgut* 'burr (only about ptarmigan)', and *čeassa* 'rennet (for making cheese)' (< Norwegian *kjæsa*, *kjæse* etc.).

For linguists, however, use of the dictionary is made difficult by a lack of basic grammatical information, such as the consonant gradation of lemmas. Furthermore, its value for scholars is marred by opaque transcription principles. In the preface to the print edition, it is explicitly stated that the dictionary is not normalized, and while it is clear that no systematic normalization of lemmas has taken place, it nevertheless appears that many words have been rendered closer to the standard language than they are in the local dialect. For instance, the dictionary features words with the dental fricatives *đ* and *ŧ*, which to my knowledge are not found in this variety, e.g. <*ođđamánnu*> 'new moon' and <*muotta*> 'aunt, mother's sister' (dialectal *odda-*, *muohtá*, standard *ođđa-*, *muottá*). In some cases, forms with *đ* and *d* are given

as separate headwords, implying a nonexistent opposition between e.g. <gođđit> 'weave' and <goddit> 'knit' (dialectal *goddit*, standard *gođđit* 'weave, knit'). Conversely, the counterparts of Standard North Saami *láddi* '(Finnish-speaking) settler' and *láđđi* 'broadcloth' are both rendered as <láđđi>.

Other entries mix features of the standard language and the local dialect, e.g. <jotŋa> 'lingonberry' (dialectal *jotnja*, standard *jokŋa*). Also particularly common is the rendering of etymologically long *á* as short *a* (as in the above-mentioned <muotta>) and of short *a* as long *á*. While such spellings may sometimes reflect the actual pronunciation of the dialect, this is not always the case, as in e.g. <háddi> 'price' (pro *haddi*). Notably, more apparent typos from the printed volume have also been carried over to the digital edition, e.g. <gaŋgggahit> 'lie stretched out', with three *g*'s.

These shortcomings in transcription have practical implications – for instance, a form such as <oabba> 'sister' (pro *oabbá*) implies an inaccurate illative singular *\*obbii* (pro *oabbái*). Additionally, the imprecise rendering of lemmas

makes comparison with other varieties challenging. To name but one example, the word <guobčá> 'large, big-boned person' should likely read *guobža* instead, corresponding to the form *guobˈǯa* 'exceptionally large animal, object' in Just Knud Qvigstad's (1853–1957) word list from the same area (n.d., available online at the National Library of Norway; also published as Skåden et al. 2004, 2008, cf. Aikio 2006). Judging from Qvigstad's form, this word is a regular reflex of Proto-Saamic *\*kuomče* (cf. Standard North Saami *guovža* 'bear'), showing the same consonantism in this variety as *lábži* 'rein' (< *\*lāmčē*, cf. Standard North Saami *lávži*). Were the form <guobčá> to be taken at face value, this cognate relationship would perhaps have gone unnoticed.

Such criticism notwithstanding, it must be recognized that *Márkku sánit* is the result of an enormous effort by dedicated language activists. Although scholars should use it only with great care, the dictionary contains interesting data that is now more accessible than ever for linguists and local language users alike.

*Olle Kejonen*

# References

Aikio, Ante. 2006. [Review of] Asbjørg Skåden, Ardis Ronte Eriksen & Marit Einejord (eds.), *Just Qvigstads lappiske ordbok fra Ibestad, Lenvik og Ofoten – bearbeidet versjon*. Skániid girjie, 2004. *Finnisch-Ugrische Forschungen* 59. 233–235.

Qvigstad, Just Knud. n.d. *Lappisk ordliste: Lenvik. Ibestad. Ofoten.* https://urn.nb.no/URN:NBN:no-nb_digimanus_317718 (Accessed 2023-06-13.)

Skåden, Asbjørg (ed.). 2010. *Márkku sánit*. Evenskjer: Skániid girjie.

Skåden, Asbjørg & Eriksen, Ardis Ronte & Einejord, Marit (eds.). 2004. *Just Qvigstads lappiske ordbok fra Ibestad, Lenvik og Ofoten: Bearbeidet versjon*. Evenskjer: Skániid girjie.

Skåden, Emma Margret & Einejord, Marit & Eriksen, Ardis Ronte (eds.). 2008. *Just Qvigstads lappiske ordbok fra Ibestad, Lenvik og Ofoten: Tillegg fra Gratangen, Bearbeidet versjon*. Evenskjer: Skániid girjie.

# Ein Opus magnum der historischen Phraseologie

Forgács, Tamás. 2021. *Történeti frazeológia: A történeti szólás- és közmondáskutatás kézikönyve* [Historische Phraseologie: Handbuch der historischen phraseologischen Forschung] (Segédkönyvek a nyelvészeti tanulmányozásához 218). Budapest: Tinta Könyvkiadó. 476 S.

Der Leiter des Instituts für Ungarische Sprache und Literatur an der Universität Szeged, Prof. Dr. Tamás Forgács, hat seit Mitte der 1990er Jahre Dutzende Untersuchungen im Bereich der Phraseologie veröffentlicht. Er hat ferner ein mehr als 800 Seiten umfassendes Wörterbuch der ungarischen Redewendungen und Sprichwörter (*Magyar szólások és közmondások szótára*, 2003) herausgegeben und eine Monografie über die Grundprinzipien der phraseologischen Forschung verfasst (*Bevezetés a frazeológiába: A szólás- és közmondáskutatás alapjai*, 2007). In seinem neuen Handbuch (*Történeti frazeológia: A történeti szólás- és közmondáskutatás kézikönyve*, 2021) befasst Forgács sich mit der historischen Phraseologie und speziell mit ihrer Methodologie.

Wie Forgács konstatiert, kann die historische Sprachwissenschaft bereits auf eine zweihundertjährige Geschichte zurückblicken, während phraseologische Forschung erst seit etwa hundert Jahren aktiv betrieben wird und sich hauptsächlich auf eine deskriptive oder vergleichende Perspektive konzentriert hat. Zwar wurde in der historischen Sprachwissenschaft die Entwicklung von Wörtern untersucht, doch sie interessierte sich nicht für die Herkunft und die Veränderungen von Redewendungen. Die Phraseologen wiederum konzentrierten sich auf die Kulturgeschichte der Entstehung von Phraseologismen oder auf ihre volkskundlichen Besonderheiten und ließen eventuelle Veränderungen in deren Struktur oder Bedeutung ebenso unbeachtet wie die Ambiguitätsentwicklung von Phrasemen oder ihre Ungebräuchlichwerdung.

Wie lassen sich verfestigte Phrasen in historischen Texten erkennen? Mögliche Fixpunkte für die Identifizierung sind u.a. eine seltsame, scheinbar nicht zum Kontext passende Wortwahl, eine ungewöhnliche Wortfolge, Wiederholung. Es gibt Themenbereiche wie das Sterben, die typischerweise mit Umschreibungen ausgedrückt werden, indem man beispielsweise Kinegramme verwendet, z.B. *den Löffel abgeben, ins Gras beißen*, oder kommunikative Situationen wie etwa die Begegnung mit einem anderen

Menschen. In alten Texten wird oft auch explizit erwähnt oder durch abweichende Schrifttypen signalisiert, dass man eine feste Redewendung verwendet. Die sog. Zwillingsformeln erweisen sich häufig als verfestigte Phrasen, z.B. *Lug und Trug* (< Lüge und Betrug). In den Redewendungen können einzigartige Lexeme auftreten, die in anderen Kontexten nicht verwendet werden, z.B. ung. *felveszi a nyúlcipőt* 'nimmt die Beine in die Hand' („zieht Hasenschuhe an"). Mögliche Kennzeichen sind auch morphologisch abweichende Elemente, fremdsprachige Ausdrücke oder semantisch scheinbar unvereinbare Wörter.

Im dritten Kapitel untersucht Forgács die mit der Phraseologisierung verbundenen Prozesse, die den Elementen der Wortschatzentwicklung entsprechen, also Wortbildung, Entlehnung und Bedeutungswandel. Sein Ausgangspunkt ist die von Irmhild Barz (1985) postulierte Zweiteilung in primäre Phraseme, d.h. in Phrasen, die auf der Grundlage freier Wortverbindungen entstehen, und in sekundäre Phrasen, die auf verfestigten Phrasen aufbauen. Die Bildung eines Phrasems kann u.a. auf Figurativität, Metaphorik, Metonymie, Ellipse basieren. Die Komponenten von Phrasen können ausgetauscht werden, eine Redewendung kann also neue Phrasen mit gleicher Struktur generieren. Andererseits können verschiedene Phrasen miteinander kontaminieren.

Das vierte Kapitel ist der inneren Struktur von Phrasen und Redewendungen gewidmet, also der Frage, was einen normalen Ausdruck oder Satz von seiner bildlichen Entsprechung unterscheidet und welche Veränderungen die Phraseologisierung zum Beispiel in der Syntax, den Rektionen und der Wortfolge auslöst. Forgács schildert die Metaphorisierungs- und Abstraktionsprozesse, durch die Redewendungen sich festigen und formen. Separat behandelt werden auch unlogische oder absurde Wendungen wie zum Beispiel *bolhából elefántot csinál* 'aus einer Mücke einen Elefanten machen' („aus einem Floh einen Elefanten machen") sowie einzigartige Wörter, die nur in Phrasen begegnen, wie *kutyaszorítóba kerül* 'in die Klemme geraten' („in den Hundeschraubstock geraten"), *vargabetű* 'Umweg' („Schusterbuchstabe"). Sowohl *kutyaszorító* als auch *vargabetű* finden sich bereits als Lemmata in Wörterbüchern, obwohl sie außerhalb der etablierten Phrasen nicht vorkommen.

Das letzte Kapitel behandelt den Wandel phraseologischer Einheiten. Im Lauf der Zeit kann sich sowohl die Form als auch die Bedeutung von Phrasen verändern. Die neue und die alte Variante können eine Zeitlang parallel existieren.

Natürlich können einzelne Phrasen auch völlig aus dem Sprachgebrauch verschwinden, was häufig auf Veränderungen des Weltbildes oder des Lebenskreises zurückgeht. Den Anstoß zum phraseologischen Wandel geben auf der Ebene des Formenbaus zum Beispiel der lexikalische und morphosyntaktische Wechsel, auf der Ebene der Semantik der denotative und besonders der konnotative Wechsel sowie auf der Ebene der Wort- und Kulturgeschichte die Motivierung und Remotivierung der Äußerung. Diese von Marcel Dräger (2012) postulierten Veränderungen in der Entwicklung von Redewendungen werden von Forgács anhand eines umfangreichen ungarischen Belegmaterials präzise dargestellt und analysiert.

Den Abschluss des umfassenden Buches bilden ein Schlagwortverzeichnis (432–436) und eine Liste aller im Buch präsentierten ungarischen (437–454) und anderssprachigen (465–476) Phrasen.

Ich konnte hier nur einige der zentralsten oder generellsten Elemente des Monumentalwerks von Forgács hervorheben. Der Autor fundiert alle dargestellten Phänomene gründlich und stützt sich dabei auf einen reichhaltigen – sowohl ungarischen als auch internationalen – Fundus an Belegmaterial und theoretischer Literatur. Hauptziel des Werks ist die Darstellung und Weiterentwicklung der Methodologie der historischen Phraseologieforschung, doch als Nebenprodukt bietet es eine Tiefenlotung der Geschichte der ungarischen Redewendungen. Die kompakte Darstellungsweise stellt hohe Anforderungen an die Aufmerksamkeit des Lesers, belohnt ihn aber mit vielen Erkenntnissen und Aha-Erlebnissen. Das Werk ist ein Opus magnum sowohl des Autors als auch der historischen Phraseologie, in seinem Fachbereich eine „Theorie von allem".

*Sirkka Saarinen*

## Literatur

Barz, Irmhild. 1985. Primäre und sekundäre Phraseologisierung. In Fleischer, Wolfgang (Hrsg.), *Textbezogene Nominationsforschung* (Studien zur deutschen Gegenwartssprache: Linguistische Studien, Reihe A, Arbeitsberichte 123), 119–140. Berlin: Akademie der Wissenschaften der DDR, Zentralinstitut für Sprachwissenschaft.

Dräger, Marcel. 2012. Plädoyer für eine diachrone Perspektive in der Phraseographie. In Filatkina, Natalia & Kleine-Engel, Ane & Dräger, Marcel & Burger, Harald (Hrsg.), *Aspekte der historischen Phraseologie und Phraseographie*, 193–226. Heidelberg: Universitätsverlag Winter.

# Mihail Mosin 1940–2022

Professor Mihail Mosin verstarb am 2. Dezember 2022 in Saransk. Er war am 2. November 1940 im Dorf Žabino im Gebiet Ardatov der Autonomen Sozialistischen Sowjetrepublik Mordwinien geboren. Nach Abschluss seiner Ausbildung an der Lehranstalt in Ardatov im Jahr 1959 wurde Mosin zum Leiter des Kulturclubs in seinem Heimatdorf gewählt. Zwei Jahre später, 1961, begann er sein Studium an der nationalen Abteilung der Philologischen Fakultät der Staatlichen Universität von Mordwinien. Mosin schloss seine Universitätsausbildung zum Lehrer für mordwinische und russische Sprache und Literatur 1966 mit hervorragenden Noten ab. Thema seiner Diplomarbeit waren die mordwinisch-marischen lexikalischen Parallelen. Im selben Jahr ging Mosin zur sog. Aspirantur, d.h. zum Aufbaustudium an das Institut für Finnougristik der Universität Tartu unter Leitung des legendären Akademikers Paul Ariste. An Ariste und an die bedeutsame Zeit in Tartu dachte Mosin oft und mit großer Wärme zurück. Seine philologische Kandidatendissertation (wie sie im Sowjetsystem genannt wurde) legte er 1971 vor; sie behandelte die estnisch-mordwinischen lexikologischen Beziehungen.

Aus Tartu kehrte der frischgebackene Kandidat an die Staatliche Universität von Mordwinien zurück, wo er als Assistent und später als Lektor am Institut für die mordwinischen Sprachen tätig war. 1973 wurde Mosin zum Dozenten ernannt. In den Jahren 1984–1986 war er wissenschaftlicher Mitarbeiter des Instituts und stellte seine Dissertation über die Entwicklung der Struktur des finnisch-ugrischen Wortstammes in den mordwinischen Sprachen fertig. Er promovierte 1987 in Tartu. 1989 wurde Mosin zum Professor ernannt.

Anfang der 1980er Jahre arbeitete Mosin zwei Studienjahre lang als Russischlehrer an der Universität Tampere und hielt auch einen Mordwinisch-Kurs an der Universität Helsinki. Er hatte 1979 zusammen mit N. S. Bajuškin ein Lehrbuch des Ersanischen für Russen veröffentlicht. Auf Finnisch erschien es, übersetzt und herausgegeben von Eino Koponen, 1983 in der Reihe Hilfsmittel für das Studium der finnisch-ugrischen Sprachen der Finnisch-Ugrischen Gesellschaft. Da er sich sowohl in Estland als auch

in Finnland aufgehalten hatte, eignete sich Mosin einen eigenen, Estnisch und Finnisch verbindenden „gesamtostseefinnischen" Dialekt an, mit dem er sich in beiden Ländern verständlich machen konnte. Vor allem zu Beginn seiner Laufbahn verfasste Mosin sowohl für den schulischen als auch für den universitären Gebrauch mehrere Lehrbücher, die das Ersanische unter verschiedenen Aspekten behandelten.

Mosins ureigenste Forschungsgebiete waren die Lexikologie, sowohl aus synchronischer wie aus diachronischer Perspektive, sowie der Vergleich des mordwinischen Wortschatzes mit anderen finnisch-ugrischen Sprachen, aber er veröffentlichte auch zahlreiche Artikel und Monografien zu anderen Bereichen der Sprachwissenschaft. So umfasst die Liste seiner Publikationen nahezu 300 Titel. Mosin verfasste populärwissenschaftliche Artikel für ersanisch- und russischsprachige Zeitschriften und trat häufig im Radio und im Fernsehen als Sachverständiger auf, wenn es um mordwinische oder allgemeiner um finnisch-ugrische Themen ging.

Mosin war bereits 1978 erstmals zum Dekan der philologischen Fakultät der Staatlichen Universität von Mordwinien gewählt worden und übte das Amt bis zu seinem Umzug nach Finnland 1980 aus. 1987 wurde er erneut zum Dekan gewählt und blieb bis 2013 im Amt. Insgesamt war Mosin also 28 Jahre als Dekan tätig. Gleichzeitig leitete er auch das Institut für Allgemeine und historisch-vergleichende Sprachwissenschaft 1988–1991 und dessen Nachfolger, das Institut für Finnisch-ugrische und vergleichende Sprachwissenschaft 1991–2018, insgesamt also 30 Jahre lang. Mosins Machtstellung sowohl am Institut als auch in der Fakultät war unerschütterlich. Er war für vieles verantwortlich, besaß aber auch reichlich Energie für die Wahrnehmung gemeinschaftlicher Angelegenheiten. Er war ein Mann der Tat, der daran gewöhnt war, Hindernisse zu überwinden und Probleme zu lösen.

Neben der Anhäufung akademischer Positionen vermehrten sich auch die gesellschaftlichen Aufgaben, deren Liste beachtlich lang ist. Mosin war u.a. Vorsitzender des interregionalen Komitees der gesellschaftlichen Bewegung des mordwinischen Volkes 1994–2014 und Vorsitzender des Bundes der finnisch-ugrischen Völker Russlands 2005–2009. Mosins Wertschätzung veranschaulichen die zahlreichen offiziellen Auszeichnungen, die er im Lauf seines Lebens erhielt. Er war u.a. verdienter Wissenschaftler der Mordwinischen Republik (1995) und der Russischen Föderation (2002) und erhielt sogar drei Mal den Staatspreis für Wissenschaft und Technik der Mordwinischen Republik (2000, 2006, 2014).

Mosin hatte intensive Beziehungen vor allem zur Universität Turku: Er arbeitete aktiv mit der 1993 an der Universität gegründeten Forschungseinheit für die Sprachen des Wolgagebiets zusammen. Mosin gehörte der Leitungsgruppe der Forschungseinheit an und beteiligte sich an den Projekten der Einheit. Als deren Ergebnis entstanden u.a. Wörterbücher: *Ersäläis-suomalainen sanakirja* (1995, Jaana Niemi und Mihail Mosin), *Suomalais-ersäläinen sanakirja* (1999, Alho Alhoniemi, Nina Agafonova und Mihail Mosin) und *Reverse dictionary of Mordvin* (2004, Jorma Luutonen, Mikhail Mosin und Valentina Shchankina). Von ihrem Gründungsjahr an organisierte die Forschungseinheit Symposien zu den Sprachen des Wolgagebiets. Das dritte dieser Symposien fand 1998 unter Mosins Leitung in Saransk statt, ebenso das neunte 2013. Über das finnische Programm der verwandten Völker entsandte er jährlich mordwinische Aufbaustudierende als Stipendiaten an die Universitäten Turku und Helsinki. Mosin war assoziiertes Mitglied der Finnisch-Ugrischen Gesellschaft.

Mosin war in vielen Bereichen aktiv. Bei der Verwirklichung von Zielen, die er als wichtig ansah, ließ er sich durch Hindernisse nicht aufhalten, sondern trieb sie entschlossen voran. Er war ein großzügiger Gastgeber. Es war ein Höhepunkt vieler Feste, wenn Mosin mit seiner schönen Tenorstimme wehmütige ersanische Lieder sang.

*Sirkka Saarinen*

# Tiit-Rein Viitso 1938–2022

Tiit-Rein Viitso, the University of Tartu professor of Finnic languages, died on 2 December 2022 at the age of 84. Viitso was one of the most active researchers of Estonian and related languages in the generation after World War II, he systematically documented the Livonian language from its last native speakers, and he was a linguist with a thorough knowledge of multiple varieties of the Finnic languages.

Viitso's interest in the Livonian language, and with it the other Finnic languages, was sparked when he was still a schoolboy. In the spring of 1948, Paul Ariste, who later became a legendary Finno-Ugrist in Soviet-occupied Estonia and a leading figure in the field, paid a visit to Viitso's aunt, who lived in nearby Tartu and worked at the university. As they chatted, it emerged that Ariste was going to visit the Livonians for two weeks, the first visit since the war. Viitso had learned a Livonian phrase from a book that he had borrowed from his school library. He tested the professor's knowledge, and Ariste recognized the source. When Ariste returned from his trip, he invited this schoolboy to his home and made him a present of two Livonian books: a collection of poems by Karl Stalte and the Livonian translation of the New Testament (*Ūž Testament*, 1942). Viitso later assumed that the latter one had been owned by Professor Alo Raun, but left behind in Estonia when that scholar fled to the USA during the war. This remarkable acquaintance continued through Viitso's studies and affected his choice of career. It was with Ariste that Viitso, then sixteen years old, got to visit the Veps people in 1954 instead of having to participate in a Spartakiad for all schoolchildren in the Soviet Union.

Due to this interest that had arisen in his school years, young Viitso applied to the University of Tartu to study Estonian and Finno-Ugrian languages and was accepted. Upon graduation, he worked as a junior researcher at the calculating center of the University of Tartu (1965–1973), until he found a firm position for twenty years at the Institute of Estonian Language and Literature in Tallinn (1973–1993), first as a senior researcher and then as a leading researcher. He served as the first visiting professor of Estonian at the University of Helsinki in 1989–1991, and then moved back to the University of Tartu, where from 1993 until his retirement in 2003 he was Professor of Finnic Languages.

Viitso's published output, which began in 1961, spanned nearly sixty years. Within it, the central themes are the history, phonetics, and phonology of the Finnic languages, as well as their inflectional typology, with first Veps and Estonian serving as focuses of research, later expanding to the Finnic languages in general and his long-cherished Livonian. Viitso's bachelor's thesis *Äänisvepsa murde väljendustasandi kirjeldus* (1968) gave a phonological description of North Veps and shows the influence of Western generative linguistics.

Viitso's background in phonetics was strongly reflected in how he viewed a language's development, but morphophonology also gradually came to play a larger role in his work. A deep connection between morphophonological changes and phonetic developments can be seen in his interest in the nature of consonant gradation and quantity alternation in the southern Finnic languages and the history of how the quantity alternation arose in general. He had already examined this in his article "Finnic gradation: types and genesis" (*SFU* 1981) and he expanded this treatment in later publications.

Viitso was fascinated by the obvious phonological differences between the Finnic languages. The southern Finnic languages had a more complex quantity system, and he felt it important to show three distinctive quantities in the description of Estonian. Through phonology he sought to explain the taxonomy of these closely related languages, which could also be interpreted historically.

In terms of describing the Estonian language, his most interesting and important articles include his explanation of the erosion of Estonian noun morphology "Eesti keele kujunemine flekteerivaks keeleks" (*Keel ja Kirjandus* 1990). Viitso also offered a summary of his views on the history of the Finnic languages and Estonian in particular in the chapters which he contributed to the Routledge volume *The Uralic languages* (1998).

Viitso got the idea of classifying the Finnic languages based on phonetic criteria from an unusual source. After he managed in 1960 to get access to some issues of *Language*, the journal of the Linguistic Society of America, he became interested in the connection between phonemes and the reduced mathematical notation then in use. Viitso gained a wider knowledge of the languages of the northern hemisphere and the literature on Native American languages, and he began to apply a model of development based on three obstruent series. This is where his work on the subgrouping of the

Finnic languages began; this was a matter that other linguists and different generations working with the Finnic languages had tried to solve.

It was phonological details that served as the basis for his inductive conclusions, e.g. in his articles "The History of Finnic õ in the first syllable" (*SFU* 1978), "Kriterien zur Klassifizierung der Dialekte der ostseefinnischen Sprachen" (*Dialectologia Uralica* 1985), and in the latter's sister publication "Läänemeresoome murdeliigenduse põhijooned" (*Keel ja Kirjandus* 1985). He presented an even more elaborated matrix map at the Congressus Internationalis Fenno-Ugristarum in Tartu in 2000 in his article "Finnic affinity" (CIFU 2000), and finally a synthesis with broader context in his collection *Liivi keel ja läänemeresoome keelemaastikud* (2008), which was published in his mature years as a researcher and can be seen as encapsulating Viitso's whole career.

Estonia's regaining of its independence dramatically changed the environment for research. International networks arose and began to significantly affect projects and funding. Estonia had a strong tradition of research into phonetics, and the University of Tartu launched new projects to study the prosody of the Finno-Ugrian languages (Erzya, Meadow Mari, Livonian, Inari Saami, 2001–2007) under Ilse Lehiste, who had escaped during World War II and became a respected phonetician in the USA. Viitso, too, took part in these projects as a specialist of several languages.

However, the most important theme of the second half of Viitso's career were the insights gathered from the Livonian language and the practical skills he had acquired from native informants. Viitso made his first visit to the Livonian coast already in 1961. More focused efforts on Livonian began in connection with the third Congressus Internationalis Fenno-Ugristarum held in Tallinn in 1970, and the summer of 1972 saw the beginning of regular fieldwork and summer trips to the Livonian coast. Livonian became Viitso's main research focus for four decades, and his classification of the inflectional typology of Livonian nouns and verbs offers a comprehensive insight into Livonian's intricate morphophonological variation at the synchronic level. Unlike in the northern Finnic languages, where a morphological form can generally be predicted from its category, Livonian represents the other extreme. Lexical inflection is divided into numerous types, where the distinctive morphological changes that occur in a word demand a lexical perspective and a morphological taxonomy based on lexical types.

Although the number of Livonian speakers continued to decline, with fewer than ten left by the 1990s, researchers' interest in Livonian gained a new impetus and took various forms. Scholarly articles and collections were followed by dictionaries, and finally in 2012 the trilingual *Līvõkīel-Ēstkīel-Leţkīel sõnārõntõz* (Livonian-Estonian-Latvian dictionary) jointly compiled by Viitso and Valts Ernštreits. As a preliminary work, Ernštreits had produced a more narrowly focused Livonian–Latvian–Livonian dictionary (1999), to which Viitso contributed appendices and a preliminary version of his Livonian declension and conjugation types. Viitso worked for twenty years on the inflectional typology of Livonian and ultimately came up with an overall picture of the different inflectional types. Guided by his long-honed expertise in morphophonology, he first presented a classification of 127 noun and 48 verb inflectional types in *Liivi keel ja läänemeresoome keelemaastikud* (2008), which brought together his numerous research focuses, but in the 2012 dictionary this was further developed so that the table in the Appendix contains 242 noun and 61 verb inflectional types. For his inflectional typology, the central criteria were paradigmatic differences shown in various word types, the geometry of syncretism, and especially the stem allomorphy caused by morphophonological variation.

A long-term collaborative effort with Finns and colleagues in the Republic of Karelia produced the three-part Finnic language atlas, for which he was the editor responsible for the second part (2007). The atlas was published in three volumes (*Atlas linguarum fennicarum* 1–3, 2004–2010) and Viitso prepared a total of over fifty maps with commentary and etymological summaries.

Although Viitso had already retired from his university duties in 2003, he continued to serve as a senior research until 2015. His long years of experience as a phonetician and expert on Livonian and the development of the Finnic languages benefited the entire scholarly community.

*Riho Grünthal*

## Sources

Klaus, Väino. 1998. Kiriusutelu Tiit-Rein Viitsoga 1998. *Keel ja Kirjandus* 41. 207–212.

Tuisk, Tuuli & Pajusalu, Karl. 2018. A journey of discovery through the Livonian world: The scholarly work of Tiit-Rein Viitso. *ESUKA – JEFUL* 9(2). 13–34.

# Raimo Anttila 1935–2023

The Indo-Europeanist and general linguist Raimo Aulis Anttila passed away on 27 January 2023 in Turku after a long illness. Combined with the 2020 abolishment of the Indo-European program of the University of Helsinki, the loss of the greatest Finnish Indo-Europeanist of all time truly marks the end of the 150-year legacy of Indo-European studies in Finland.

Indeed, largely non-Indo-European-speaking Finland has an exceptionally long tradition in Indo-European studies. At the (Imperial Alexander) University of Helsinki, Otto Donner (1835–1909) was appointed the first docent in 1870 and the first professor in 1875. He was succeeded by two more professors of (Sanskrit and) comparative (Indo-European) linguistics, Julio Nathanael Reuter (1863–1937) and Pentti Aalto (1917–1998). In addition to these professors, there have been three docents of Indo-European linguistics at the University of Helsinki: Yrjö Moses Biese (1903–1983), Martti Nyman (b. 1944), and Petri Kallio (b. 1969).

Ironically, the most brilliant Finnish Indo-Europeanist never studied Indo-European linguistics in Helsinki or even in Finland. Anttila was born on 21 April 1935 in the Vintala village of Lieto. His roots went deep in the Southwest Finnish soil, his ancestors having been farmers for generations. His parents Tyyne Raakel *née* Alikirri (1915–1994) and Lauri Nikolai Anttila (1904–1964) became the first to break the tradition when they moved to Turku in order to work as a pressing worker and a metal worker, respectively. Due to his working-class background, Anttila was always incredibly handy for a scholar, being able to make anything from wooden swords to cockle stoves.

Anttila became interested in languages already as a wartime child evacuee in Swedish Lapland, and from 1956 onwards he studied English, German, Latin, and Greek at the University of Turku. Incidentally, it was English philology that he majored in, and nothing in his 1961 master's thesis *Stilostatistinen tutkielma Towneleyn mysteerinäytelmien kahden eri kerrostuman vierasperäisistä sanoista sekä kotimaisista substantiiveista* predicted his future as an eminent Indo-Europeanist. Yet his English professor was none other than Biese mentioned above, who had made a career change from Indo-Europeanist to Anglicist.

Anttila's opposite career change from Anglicist to Indo-Europeanist ultimately took place at Yale University, where from 1962 he studied structural

linguistics under Bernard Bloch (1907–1965), comparative linguistics under Isidore Dyen (1913–2008), and, last but not least, Indo-European linguistics under Warren Cowgill (1929–1985). It was Cowgill who first suggested that Anttila look into *Schwebeablaut*, something that Anttila back then "had no idea what it was", until he was put straight by his fellow student Alfred Bammesberger (b. 1938), another Anglicist turned Indo-Europeanist.

Yet Anttila was a quick learner, and his 1966 Yale dissertation *Proto-Indo-European Schwebeablaut* (published in 1969) remains the greatest Finnish achievement in Indo-European studies. Granted, Anttila himself would have been the first to disagree, since he instead preferred *Uralische Evidenz für die Laryngaltheorie* by his friend Jorma Koivulehto (1934–2014). While there is no accounting for taste, Anttila's dissertation has stood the test of time better than Koivulehto's 1991 monograph, because it continues to be recommended as further reading on *Schwebeablaut* in almost every twenty-first-century introduction to Indo-European linguistics.

In 1965 Anttila moved from Yale to UCLA, where he remained until his 2006 retirement despite his short stint as a professor of general linguistics at the University of Helsinki in 1972–1974. One can only wonder how Indo-European linguistics would have developed in Finland if he had declined to become a professor of Indo-European and general linguistics at UCLA in 1974, not least because, among other things, Anttila was also a versatile teacher whose students stretched from the general linguist Lyle Campbell (b. 1942) to the linguistic archaeologist James P. Mallory (b. 1945), neither needing any further introduction.

Early on Anttila moved from Indo-European to more general historical and comparative linguistics, as best exemplified by his 1972 classic, *An Introduction to Historical and Comparative Linguistics* (whose 1989 reprint *Historical and Comparative Linguistics* contains a new introduction as well as one additional chapter). At that time the American linguistic scene was dominated by the so-called Chomskyan revolution which Anttila's 1977 monograph *Analogy* was among the first to correctly identify as "a most curious linguistic bubble", having "no scientific merit" and representing "a big step backwards". Unfortunately for him, what is self-evident today was long considered heretical and even blasphemous.

Yet Anttila never abandoned Indo-European linguistics, and he returned to it in his last monograph *Greek and Indo-European Etymology in Action: Proto-Indo-European \*aǵ-*, published in 2000 but based on decades of mental exercise. As usual this swan song of his was much more than its

title might suggest, as it did not settle for only one Proto-Indo-European root but delved deep into different fields of the humanities. Stylistically Anttila was always entertaining and never boring, fully confirming the Saussurean wisecrack that "la linguistique évolutive est amusante". Despite living over half a century in the USA, he remained a true epitome of Finnishness, preferring honesty over politeness.

Anttila's very last publication was the most revealing: the 2018 English translation of *The Gundestrup Cauldron* by his senior archaeologist friend Unto Salo (1928–2019). Not anyone would bother to translate a 210-page book for free, to say nothing of a world-famous professor emeritus in his eighties. Anttila did so, "because he found the text exceptionally valuable for cultural history", just to give an illustrative example of how altruistic he was as a person. Despite his achievements, he was always modest and down-to-earth, and great as he was as a linguist, he was even greater as a friend.

*Petri Kallio*

*Sananjalka* on vuodesta 1959 lähtien ilmestynyt laadukasta humanistista tutkimusta julkaiseva tieteellinen vuosikirja.

*Sananjalassa* julkaistaan vuosittain noin kaksitoista vertaisarvioitua artikkelia suomen ja sen sukukielten, kotimaisen kirjallisuuden, folkloristiikan, kansatieteen, uskontotieteen ja arkeologian alalta. Lisäksi vuosikirja tarjoaa luettavaksi kiinnostavia katsausartikkeleita ja arvosteluja tuoreista humanistista alaa koskettavista kirjoista.

*Sananjalkaa* julkaisee vuonna 1929 perustettu Suomen Kielen Seura.

# *Sananjalka*

Numerossa 65 muun muassa :

MARIA SARHEMAA: *Jonnet ei muista* -konstruktion muoto ja merkitys Suomi24:ssä ja Twitterissä

KAARINA HIPPI: Murteet kielellisessä elämäkerrassa: monikulttuuristen suomenpuhujien käsityksiä itsestään suomen murteiden käyttäjinä

HANNA LAPPALAINEN JA MAIJA SAVINIEMI: Puhuttelu ja sen metakielinen kommentointi Kalle Päätalon Iijoki-sarjassa

ELINA VITIKKA: Referointi rajatyönä: virheellisten terveysväitteiden kumoaminen tietokirjallisuudessa

MILLA JUHONEN: Keppulointia ja rehjustusta Rööperin kulmilla: Paikannimistö sosiaalisen maantieteen rakentajana Outi Pakkasen ja Matti Yrjänä Joensuun rikosromaaneissa

Lisäksi uusimmassakin numerossa on runsaasti muita artikkeleita, katsauksia ja kirja-arvosteluja.

# VIRITTÄJÄ

**V**anhin ja suurin suomenkielinen kielitieteellinen lehti

**A**jankohtaisia ja monipuolisia kirjoituksia suomen kielen ja sen sukukielten tutkimuksesta ja opetuksesta

**N**eljä numeroa vuodessa, yhteensä noin 630 sivua

**K**otisivuilla laaja hakemistotietokanta julkaistuista kirjoituksista

**K**irjaesittelyjä, katsauksia ja keskustelua

**A**ikaisempia vuosikertoja ja tuoreita lektioita verkkolehdessä

*https://journal.fi/virittaja*

*virittajablogi.kotikielenseura.fi*

**127. vuosikerran sisältöä:**

- Selkokielen tutkimuksen teemanumero (1/2023)
- HARRY LÖNNROTH  "Lyödään ne hepnaadilla!" Alvar Aallon kielellinen elämäkerta (2/2023)
- HARRI MANTILA & MATTI LEIVISKÄ  *Särkemys*, *Satimus* ja *Ylimys* -tyyppisistä paikannimistä (3/2023)
- RIKU ERKKILÄ  Mistä uralistiikan sijatermit tulevat? Suomen ja muiden uralilaisten kielten sijatermistön kehitys (4/2023)

**Tilaushinnat 2024**

- Suomeen 82 € (Kotikielen Seuran jäsenille 79 €, opiskelijoille 33 €)
- Ulkomaille 82 € (Kotikielen Seuran jäsenille 79 €, opiskelijoille 39 €)
- Verkkolehden tilaus Kotikielen Seuran jäsenille 42 €, opiskelijoille 22 €