Sampsa Holopainen (Helsinki & Wien)

On some problems of Ugric etymology: loans and substrate words

In this paper, the shared vocabulary of the Ugric languages (Hungarian and the Ob-Ugric languages Khanty and Mansi) is discussed. Words that have been considered loanwords from Iranian or Turkic languages into Proto-Ugric, the intermediary proto-language of Hungarian and Ob-Ugric in traditional models of Uralic taxonomy, are analyzed critically, and Proto-Ugric words that fulfill the criteria of substrate words are also analyzed. It is shown that a large part of the vocabulary traditionally reconstructed for Proto-Ugric in earlier etymological dictionaries like the UEW, consist of parallel loanwords (sometimes from unknown sources) rather than shared lexical innovations.

1. Introduction and research problems[1]

Problems with the reconstruction of the Ugric proto-language have been frequently mentioned in the literature discussing the history of the Uralic family (Kálmán 1988; Abondolo 1998; Salminen 2001: 387–391; 2002; J. Häkkinen 2007: 79–80; Janhunen 2009: 65; Zhivlov 2018; Holopainen 2019: 15–16; Grünthal et al. 2022: 492; Aikio 2022: 3–4; Saarikivi 2022: 31–32, 49–54), and the mutual relationship of Hungarian, Khanty, and Mansi remains one of the most debated and poorly understood parts of the classification of the Uralic languages. There are various problems connected with the phonological and morphological reconstruction of Proto-Ugric (see Salminen 2002 for a criticism of some of the suggested Ugric innovations). It has been noted that even Proto-Ob-Ugric, the alleged intermediary proto-language of Khanty and Mansi, seems to differ very little from Proto-Uralic, which further casts doubt on the unity of Ugric (Tálos 1984: 95–98; Aikio 2014: 29–30; 2018: 78–79). Both Ugric and Ob-Ugric entities have been also labeled as possible "areal-genetic units" or *Sprachbünde* by Helimski (2003), meaning that their shared innovations are the result of areal spread.

The problems with the Proto-Ugric lexicon have not received much attention in recent years, although individual etymologies have naturally been commented on in recent research (see below for a survey of recent studies). Some remarks on the Ugric etymologies have been recently presented by Holopainen (2022a) in the Appendix to Grünthal et al. (2022), but no proper etymological studies have been published on the topic recently. The shared vocabulary of Ugric is often invoked as one of the primary

---

arguments supporting the Ugric unity (see especially Honti 1997; 1998: 354–355; 2013: 102–104); K. Häkkinen (1983: 92–93) noted that the relatively small number of Ugric etymologies does not yield support for a Ugric unity, but she did not analyze the Proto-Ugric etymologies in detail. However, it is quite clear that the Proto-Ugric vocabulary presented in MSzFE and UEW involves many irregular comparisons, and even the individual etymological entries in these works frequently mention irregular developments when commenting on the Ugric etymologies. (Also, the vocabulary shared exclusively by Khanty and Mansi involves similar problems; this layer of vocabulary has been investigated by Sipos 2002; 2003, but further research is needed.)

Furthermore, even though MszFE and the UEW list Proto-Ugric etymologies, they do not present a phonological reconstruction of Proto-Ugric. Although a reconstruction of the Proto-Uralic and Proto-Finno-Ugric phoneme system is provided in the introduction to the UEW, a description of the Ugric proto-language is missing, and it remains unclear to the reader what the Proto-Ugric reconstructions are actually based on. In many cases the phonological reconstruction of the single etymologies is also only half done, with both MSzFE and UEW using the symbols ϑ (for any vowel) and $\ddot{\vartheta}$ (for any front vowel) instead of reconstructing the vocalism properly. One can note that the Ugric lexicon has been reconstructed based on intuition rather than on strict sound correspondences; this practice is prevalent in all the reconstructions of the UEW, as noted by Zhivlov (2014: 114), who calls these reconstructions "pre-reconstructions", but the Proto-Ugric reconstructions seem to be particularly problematic.

The situation of the Ugric lexicon can be compared with the Finno-Volgaic and Finno-Permic lexicon that has recently been discussed critically (Zhivlov 2015; Aikio 2015b: 43–47; Pystynen 2017; see also Itkonen 1997 for an earlier treatment of the "Finno-Volgaic" lexicon). It has become quite clear that a large part of the allegedly Finno-Volgaic and Finno-Permic etymologies are irregular words, pointing to possible substrate origin, or parallel loanwords from Indo-European languages. It is important to approach the vocabulary confined to Ugric from a similar point of view. However, the Ugric lexicon is more problematic in the sense that the historical phonology of the Ugric languages is less well known than that of "Finno-Permic" languages like Finnic, Mordvin, or Permic.

This issue is related to the larger problem of the historical phonology of the Ugric languages. There is no commonly accepted reconstruction of Proto-Ugric phonology, and the existing attempts by Sammallahti (1988) and WOT (1011–1069) include various problems (these will be commented on below in more detail). The lack of Proto-Ugric reconstructions stems from the problems of our understanding of the historical developments of both Ob-Ugric and Hungarian vocalism; see Zhivlov (2006, 2007) for a discussion of problems in the phonological developments of the Ob-Ugric languages. It is clear that as the historical phonology of Ugric is poorly understood, the reconstruction of the Ugric lexicon is far from settled.

In this paper, problems with some of the alleged Proto-Ugric etymologies are investigated. I will first comment on the Proto-Ugric etymologies that have been

considered possible loanwords from Iranian or Turkic languages, and then I will discuss irregular comparisons that could be classified as substrate words. Indo-Iranian loanwords in Ugric have been recently discussed by Holopainen (2019), and the evidence from Iranian loanwords will be used in the discussion of other loanword layers. New evidence can still be added, and the Iranian loans are approached in this paper especially through the question of whether they could have been borrowed already into Proto-Ugric or not. The research questions of this paper are whether there are common loanword layers shared by the three Ugric language groups (i.e. are the loanwords regular cognates?), and whether the correspondences in loanwords can be used to analyze other problematic Ugric etymologies.

Starting with known or possible loanwords is a good approach, as the history of the donor languages helps in working out the phonological developments and sieving out possible parallel loanwords. When the phonological correspondences of loanwords are clear, the sound correspondences in possible substrate words can be compared with them. The results can later be used when discussing other allegedly Proto-Ugric etymologies.

2. Background

2.1. The problems of Ugric unity and the role of lexicon in the classification of languages

The problematic taxonomy of the Ugric languages within the Uralic family cannot be discussed without commenting on the methodology for the genealogical classification of languages. It should be noted that most of the presentations of Ugric unity (such as Honti 1997; 2013: 102–104) do not comment on the relative importance of the innovations at the different levels of language (phonology, morphology, syntax, lexicon), although it is quite clear from the handbooks and literature that not all these levels have equal weight in discussions on taxonomy.

The role of lexicon in linguistic taxonomy is considered rather trivial by most handbooks of historical linguistics (Fox 1995: 306–307; Campbell 1999: 318–319; Clackson 2007: 10–12; 2022: 23). Words get borrowed more easily than grammatical markers or phonemes, and even though in principle one can distinguish between inherited and borrowed words, in practice it is not always very easy to sieve out borrowings and diffusion (Campbell 1999: 318–319). The use of lexical evidence alone can lead to a misleading picture of genealogical relationships between languages (Campbell & Poser 2008: 165). Especially when dealing with languages like Hungarian, Khanty, and Mansi, whose historical phonology is inadequately known, one should be especially careful with the lexical evidence. The problems of lexical loss and lexical innovation have also been highlighted by Salminen (2002) in his commentary on the Ugric taxonomy. Salminen notes that as shared words can often be retentions, the presence of a word as such does not tell us much (Salminen 2002).

Regarding subgrouping in general, Clackson (1994: 16–17; 2022: 26–27) high-lights the difficulties in distinguishing a real subgroup from a dialectal relationship. According to Clackson, different dialects can share innovations without belonging to the same subgroup. For example, Mycenean Greek shared some innovations with various Greek dialects, but all the later dialects also shared innovations that are missing from Mycenean. The situation with Ugric could well have been similar: Hungarian, Khanty, and Mansi clearly share some innovations, but this does mean that it is possible to reconstruct a common proto-language for these branches and that they indeed form a genealogical subgroup within Uralic.

It is also possible that even if the Ugric languages form a genealogical subgroup within Uralic, some of their common innovations result from the period after the Ugric proto-language had already split into different dialects. Olander (2022: 7–8) discusses the methodological issue of distinguishing a *proto-language* and a *common language* as has been done in Slavic and Indo-European studies: the latter term can be used to denote a group of closely related dialects that still share innovations even though the divergence from the proto-language has already taken place. For example, the Ugric lexical innovations can stem from *Common Ugric* rather than *Proto-Ugric* times. Such a distinction is rather new in Uralic studies, where the term *Common Uralic* has recently been used to describe a stage of Uralic dialects after the initial divergence of Proto-Uralic; see the discussion on Common Uralic by Grünthal et al. (2022: 497, 501).

The situation with Ugric can be compared with known examples of similar situations in different language families. For example, within the Indo-European family, there are known examples of problematic nodes sharing a large number of cultural vocabulary items. For example, Greek and Armenian exclusively share a number of vocabulary items, but it has been shown (Clackson 1994; Clackson 2022: 26) that these branches are not particularly closely related to each other. A similar situation can be noted with so-called North-West Indo-European (consisting of Germanic, Italo-Celtic, and Balto-Slavic). Also, Indo-Aryan and Iranian, which do form a real branch within Indo-European, share a large number of vocabulary items that are irregular and usually considered to have originated from a Central Asian substrate (Lubotsky 2001; see also Kuiper 1991): this shows that even in cases where we are dealing with real genealogical nodes, the lexical evidence does not have very much value.

2.2. Methodology of loanword studies

As mentioned above, Lubotsky (2001) has shown that among the shared vocabulary of Indic and Iranian there are a large number of loanwords, probably from the language spoken by the Central Asian culture known as the Bactria-Margiana Archaeological Complex. Lubotsky (2001: 301), referring to earlier works by Kuiper (1991; 1995), Beekes (1996), and Schrijver (1997), presents the following criteria to detect loan-words: the word has a limited geographical distribution; the word shows phonological

or morphological irregularity, unusual phonology, or unusual word formation; and the word belongs to a semantic category which is particularly liable to borrowing.

In principle, the methodology of loanword studies is similar to the study of substrate words: substrate words are also loanwords, although the sociolinguistic setting might differ from "normal" borrowing. The Ugric words might include both loanwords in the classical sense and also substrate words, but as the cultural and sociolinguistic context of the possible contacts between the Ugric languages and unknown languages of Eurasia is unknown, it makes little sense to try to separate *loanwords* from *substrate* in this study. The aim is to be able to distinguish possible borrowings in the lexicon common to the Ugric languages.

Aikio (2004: 8–9), in his study of substrate words in the Saami languages, repeats the criteria presented by Salmons (1992: 267) for detecting a substrate. The criteria by Lubotsky listed above can be used to detect individual borrowings, but the criteria of Aikio and Salmons rather concern the existence of a layer of substrate words. These criteria partly overlap, however: the amount of vocabulary of unknown origin is significant; these words concentrate on certain semantic fields; the words show non-native structural features (such as phonotactic combinations); some of the words show irregular sound correspondences. Similar features to detect a substrate are also listed by Salmons (2015: 106); he also mentions the criterion of checking whether the possible substrate word is a loan from (known) neighboring languages or a Wanderwort.

Regarding the Ugric vocabulary, it is obvious that the absolute number of possible substrate items is much smaller than in the Proto-Saami vocabulary that comprises hundreds of words. The small number of etymologies also makes the research more challenging. A notable difference compared to Saami is also that Saami historical phonology is much better known, which makes it easier to spot irregular words and distinguish them from real cognates.

2.3. The Ugric lexicon in the UEW

The UEW lists 126 certain or probable Ugric etymologies; this largely reflects the earlier presentation of MSzFE, although some etymologies still found in MSzFE are not included in the UEW's material. In addition, 52 "uncertain" etymologies are listed. However, it is not quite clear what the criteria for "certain" and "uncertain" etymologies are: while the 52 uncertain etymologies are indeed highly problematic comparisons, also the "certain" group of etymologies includes cognates that manifest various problems, and even the authors of the UEW sometimes note that the etymology involves irregular developments.

Although the Ugric cognates of the UEW include all kinds of lexemes, it has been repeatedly noted that this group includes several words related to horse nomadism and riding, and this has been used as key evidence for the idea that the speakers of Proto-Ugric practiced equestrian culture and nomadism. What has not received much attention, however, is that a significant part of these words display irregular sound

correspondences, which means that either the Proto-Ugric etymologies are incorrect (the suggested cognates in Khanty, Mansi, and Hungarian are not real cognates), or that even if the words are somehow related, they are not real cognates but perhaps parallel loanwords.

In addition to words related to horses, several other animal names are included in the Ugric vocabulary, such as the word for 'dog'. Some of these also involve irregularities and they are also discussed below.

The Ugric lexicon includes borrowings from Iranian (or already from Proto-Indo-Iranian). These have been discussed in detail by Korenchy (1972) and recently by Holopainen (2019). As noted already above, many of the loanwords involve irregularities, and also Korenchy considered some of these words parallel borrowings; in the UEW, it is usually assumed that the Iranian words were borrowed into Proto-Ugric.

In addition to Iranian loanwords, the UEW supports a Turkic origin for some of the Ugric words. The idea of Turkic loanwords in Proto-Ugric has been refuted by Róna-Tas (1988), who considers the opposite direction of borrowing possible. However, since some of these allegedly Turkic words also show irregularities within Ugric, it seems unlikely that they are really Proto-Ugric items, and a fresh analysis of these words is in order.

2.4. Problems with the reconstruction of Proto-Ugric historical phonology

Before dealing with the etymologies in detail, one should have a look at what can be said about the Ugric phonological developments. As noted above, there is no accepted reconstruction of Proto-Ugric historical phonology. Sammallahti (1988) presents developments based on outdated ideas of Proto-Finno-Ugric and Proto-Uralic: many of the sound changes suggested by Sammallahti (1988) are obsolete in light of the new Proto-Uralic reconstructions by Aikio (2012, 2015). The presentation of Ugric historical phonology in WOT is based mostly on the UEW material, making the conclusions unreliable. WOT also assumes several "sporadic" sound changes, which is methodologically problematic. Some problems have been highlighted by Bakró-Nagy (2013), who does not present a new reconstruction of Proto-Ugric, however.

Although the phonological reconstruction of Proto-Ugric vocalism is not settled and further research is needed, it seems clear that there are very few possible changes shared by Khanty, Mansi, and Hungarian. Some of the most problematic points include the development of Proto-Uralic *u (see Zhivlov 2014: 121). Sammallahti (1988) suggests that Proto-Uralic *u developed into Proto-Ugric *u and *$ŭ$, but many counter-examples for his criteria can be suggested, and the Khanty reconstructions of Zhivlov (2006, 2007) show that PU *u was retained in many cases where Sammallahti assumes a change PU *u > PUg *$ŭ$. Further research on the development of PU *u in Ugric is clearly needed, but it seems clear that the split assumed by Sammallahti is not a real Proto-Ugric innovation. The developments of *a, *o, and *$ẹ$ in Hungarian, Khanty and Mansi are presented in Zhivlov's (2014: 126) study on Uralic vocalism, and based on his remarks, it seems that there are no developments shared by Hungarian, Khanty,

and Mansi regarding these vowels. Sammallahti (1988: 500, 514) also suggests that Proto-Uralic *i developed into Proto-Ugric *ĭ and Proto-Uralic *e changed to Proto-Ugric *i, but these developments are also not convincing: in Hungarian, Proto-Uralic *e develops regularly to é, regardless of the stem vocalism, and there is no reason to assume that a change *e > *i > é would have happened in the history of Hungarian.

Most, if not all, of the changes Sammallahti assumes are incorrect, and the same can be said of the Proto-Ugric reconstruction presented in the Appendix to WOT. The latter reconstruction is in many cases more problematic than that of Sammallahti, as WOT presents "tendencies" and sporadic changes instead of regular sound laws. The account is based on the UEW's etymologies, which include many irregular and false comparisons. Also, the Proto-Ob-Ugric, Proto-Khanty, and Proto-Mansi reconstructions used in WOT involve problems. To mention some examples of problematic reconstructions in WOT: WOT assumes a sporadic change of PU *o to PUg *a in cases like PU *ńora > PUg *ńarə 'poplar' (WOT: 1043), but there is no reason to reconstruct an intermediary with *o here. A similar case is the idea that Proto-Uralic *u becomes Proto-Ugric *o in some cases, e.g. PU *kuŋi 'moon' > PUg *koŋə (WOT: 1045–1046, 1049–1051): it seems quite clear that the old *u was retained in Hungarian for a long time, and the Ob-Ugric reflexes do not warrant a reconstruction of Proto-Ugric *o. WOT also assumes that the Proto-Uralic stem vowels merged in Proto-Ugric, but this cannot be correct, as the old *-a and *-i-stems usually yield different initial-syllable vocalism in Hungarian, Khanty, and Mansi.

Some evidence for Proto-Ugric has recently been suggested by Zhivlov (2016), who argues that the emergence of retroflex *ṇ (> Khanty ṇ, Hungarian ny) is a Proto-Ugric innovation; this change shows no traces in Mansi, and the reflexes of *ṇ merge with PU *ń in Hungarian, meaning that it is not completely clear that we are indeed dealing with a Ugric innovation here. An old idea is that the merger of PU *s and *š as PUg *θ is a Proto-Ugric innovation, but this change is shared by Samoyedic, and exceptions in all Ugric languages point to an "incomplete" change (see Zhivlov 2018; Holopainen 2019: 24). Other suggested Proto-Ugric consonant changes are disputed: the reflexes of Proto-Uralic *ŋ in Proto-Ugric are not unitary (see Bakró-Nagy 2003; Zhivlov 2023: 139), and the allegedly Proto-Ugric change *w > *γ (listed as a Proto-Ugric change by Honti 1998: 353) is clearly reflected by Khanty only (Salminen 2002). These issues with consonantism require further research, and at present also the reconstruction of Proto-Ugric consonants is disputed.

The problem of Proto-Ugric historical phonology is related to the disputes involving the reconstruction of Proto-Ob-Ugric. As noted by Aikio (2014: 29–30), there is no commonly accepted reconstruction of the Ob-Ugric proto-language. Here the Proto-Khanty and Proto-Mansi reconstructions of Zhivlov (2006; 2007) are used, rather than the system of Honti (1982) that has been the standard system in many works, even though Honti's system has been heavily criticized (Tálos 1984; Helimski 1985). In the cases where the inter-Khanty and inter-Mansi comparisons show irregularities and the reconstruction of a Proto-Mansi or Proto-Khanty word is uncertain, the reconstructions are marked with a question mark.

2.5. Recent studies on Ugric etymology

The Ugric languages have played a relatively small role in Uralic etymological stud-ies: although Hungarian etymology has a long research tradition, the situation with the etymological study of Khanty and Mansi is less developed, and the reconstruc-tion of Proto-Uralic has traditionally been based mostly on the evidence from the so-called Finno-Permic and the Samoyedic languages (Janhunen 1981). Janhunen (1981: 2) notes, somewhat enigmatically, that the Ugric languages can be "deduc-tively" added to the reconstruction of Proto-Uralic that is derived from Finnic and Samoyedic evidence.

As noted above, the UEW lists 126 certain Ugric cognates and 52 uncertain ones. Some of these cognates have been criticized in subsequent research: it has either been shown that the etymology is not reconstructable for the Proto-Ugric stage due to phonological irregularities, or it has been shown that the word has cognates outside of Ugric, making the etymology correct as such but datable to earlier stages (Proto-Uralic or Proto-Finno-Ugric). Here I list those etymologies and briefly comment on the explanations:

Proto-Ugric *kajꝫ 'hair' (UEW: 854) is reflected in Mansi TJ χōj, P kōj, N xåj 'hair' (< Proto-Mansi *kāj) and Hungarian haj 'hair' according to the UEW, but Aikio has shown that phonologically regular cognates are found in Saami (North Saami guodja 'seed shell of a sedge', also in South, Pite, Lule, and Inari Saami < Proto-Saami *kuoję), Permic (Komi and Udmurt kị 'awn' < Proto-Permic *kị), and Samoyedic (North Selkup, Taz qu 'stalk, stem, oblong or slender object' < PSam *kåə). Aikio reconstructs the word as *kaji 'grass, stalk'. The word is thus clearly a Proto-Uralic item. However, it is interesting that only Hungarian and Mansi show the meaning 'hair' and this might, in fact, represent a shared semantic innovation.

Ugric *kaŋꝫ- (kaŋkꝫ-) 'climb, rise' (UEW: 127, 855) (> Hungarian hág 'mount a horse', Khanty (East) kaŋət- 'climb; go upwards a mountain, go against the stream', Mansi (South) kē̮ŋk- 'climb (klettern, aufklettern)') is a similar case. Aikio (2015b: 65) reconstructs *keŋka- 'crawl, climb' and assumes cognates in Finnic (Finnish kankea 'stiff, rigid') and Saami (North Saami guokkardit 'crawl', South Saami goegkerdidh 'crawl; climb'). The idea is plausible from the phonological point of view, and although the semantic developments leading to the present languages are not as straightfor-ward (the meaning 'stiff, rigid' in Finnish stands out, but the rest of the meanings probably can be derived from a Proto-Uralic meaning 'climb', with 'climb' > 'crawl' taking place in Saami), Aikio's etymology can be considered at least probable, even if not absolutely certain.

The Ugric background of Hungarian könyörög, Khanty (East) könəγ- 'growl omi-nously (of dog) (der Hund)' and Mansi (East, North) kēnγ- 'low, whimper' is doubted by Zhivlov (2016: 296). The UEW reconstructs the word as PUg *kenꝫ- 'growl, bel-low; scrape (vt), plane, dress', but the vocalism of the Ugric words does not point to earlier *e, and the etymology is also semantically dubious.

The Ugric etymology of Hungarian *far* 'back', *farok* 'tail' and Khanty (East) *pir̮* 'hinter etw. befindlich, Raum hinter etw.' (UEW: 407, 880) is convincingly refuted by Aikio (2018: 81–83), as the relationship is phonologically irregular and neither of the words points to Proto-Ugric **pur₃* that is reconstructed by the UEW. Instead, Aikio argues that Hungarian *far* reflects Proto-Uralic **ponči* 'tail'.

Hungarian *füst* 'smoke', Khanty (North) *posəŋ* 'smoke, dust', and Mansi (South) *pošėm*, (West) *pošəm*, (North) *posim* id. reflect Ugric **pič₃ (pić₃)* 'smoke' according to UEW (879). However, *füst* is an Iranian loanword according to Helimski (2002: 109), who also shows that the Ugric cognates are irregular. As the Iranian loan etymology is convincing (← Iranian **pazda-* 'smoke', cf. Sogdian *pzt-* 'smoke', **pazdaka-* > Ossetic *fæzdæg* id.; see Cheung 2007: 304, s.v. *pazd¹*), the Ugric etymology can clearly be rejected. The relationship between the Ob-Ugric cognates requires further research; the Khanty word is only found in the northern variety, which makes it dubious that the word is old.

The phonological problems with Proto-Ugric **päč₃-* 'part, separate, work itself loose' (UEW: 358–359) have been pointed out by Aikio (2014: 54, 65), who notes that the phonological reconstruction is uncertain (Aikio tentatively reconstructs Ugric **pi/ečV*). Hungarian *fesl-* 'separate oneself' points to Ugric **ä*, whereas the alleged Khanty and Mansi cognates rather point to **e*.

The Ugric etymology (UEW: **keć₃-* 'track, pursue') of Hungarian *kísér* 'accompany, pursue' and Khanty East (V) *kŏt́-*, South (DN) *kŏt́-* 'track, pursue' (< Proto-Khanty **küć-*) is refuted as irregular by Aikio (2015a: 10–11), who convincingly argues that the Khanty word reflects Proto-Uralic **küjə-* 'track; follow'.

The Ugric word for 'saddle', reconstructed as **närk₃* in the UEW (874) and allegedly reflected by Hungarian *nyerëg*, Khanty (East) *nöγər*, Mansi (South) *näwrā̇*, is briefly commented on by Zhivlov (2016: 300), who notes the irregular correspondences of the alleged cognates in Khanty, Mansi, and Hungarian and assumes a borrowing from some unattested language. As the Ugric words for 'saddle' are central for the assumption of Ugric horse culture, and since Iranian and Turkic etymologies have also been suggested for this word, it will be discussed in greater detail below.

UEW (832) reconstructs Proto-Ugric **ańć₃* 'backside, buttocks' to account for Hungarian *ágyék* 'loin' and Mansi N *ūńś* 'ass' (< Proto-Mansi **ūńć* (**ōńć* in Zhivlov's 2023 system)), but Sammallahti (1988: 542) considers these words reflexes of Proto-Uralic **ońća* 'flesh, hind' (> Fi *osa* 'part; luck'). This has also been supported by Aikio (2015b: 61). The etymology is phonologically convincing, and it is unlikely that the Hungarian and Mansi words form a separate Ugric item. However, it is interesting that the semantic difference is notable: as in the case of **kaji* above, also here we can assume that Hungarian and Mansi show a common semantic innovation, which technically could have happened at the Proto-Ugric level.

Hungarian *hamu* 'ash' and its Ob-Ugric cognates, Khanty (East) *kajəm* 'slag left in the fireplace from burning aspen', Mansi (East) *χūl'əm* 'ashes' (UEW: 194–195, 858: Proto-Ugric **kuδ̮m₃*) have been analyzed as derivatives of PU **kaďa-* 'leave' by Abondolo (1996: 93), and this same idea is supported by Aikio (2015a: 11) and Zhivlov

(2014: 120). The etymology seems plausible, even though not all the phonological details are quite clear (Hungarian word-final *-u* is of unclear origin, and it is unclear why the Uralic *-a* has a consonant stem in the derivative **kad'ma*; usually the Uralic **-i*-stems show consonant-stem allomorphs; Aikio 2022: 11).

The Ugric background of Hungarian *táltos* 'shaman' (allegedly related to Khanty (North) *tǫʌt* 'help, relief (during sickness or poverty)', Mansi (North) *tūlt: tūltėn* 'easily' according to UEW: 895, **tultɜ* 'magic, witchcraft') has been doubted by Abondolo (1996: 44) and is criticized also by WOT (841–846) and Róna-Tas (2017: 56–57); the latter also refutes Honti's (2017: 62–67) attempts to rehabilitate the Ugric etymology that is clearly irregular. WOT argues that the Hungarian word is a borrowing from West Old Turkic **taltučï* 'the one who exercises a loss of consciousness', and the origin of the Ob-Ugric words remains uncertain.

Some other Ugric etymologies have also been refuted by WOT and the subsequent article by Róna-Tas (2017). These include the uncertain Ugric etymology (UEW: 860: **kurV-*) for Hungarian *hord* 'carry' and Mansi *χart-, kart-* 'ziehen', criticized due to problems with the suffixal elements (WOT: 1319; Róna-Tas 2017: 64), the complicated issue of the Ugric ethnonym **mańćɜ* (UEW: 866–867 > Hungarian *magyar* 'Hungarian', Khanty (East) *mańť* 'name of a phratry', Mansi (East) *m͞ǟńś* 'Mansi'; WOT: 1320; Róna-Tas 2017: 66–67; these etymologies will be discussed below in more detail) as well as the allegedly Ugric metal name *wɣlmɜ* 'tin' (UEW: 899) that might be reflected by Hungarian *ólom* 'tin' and poorly attested South Mansi *wōlėm* id. (WOT: 1321; Róna-Tas 2017: 68–69).

This list shows that at least 14 etymologies listed in the UEW cannot be considered Proto-Ugric lexical innovations. This is already a notable part of the words assumed for Proto-Ugric. Additionally, Iranian loanwords allegedly borrowed into Proto-Ugric have been discussed by Holopainen (2019) and one etymology by Junttila & Holopainen (2022: 314–315) and Holopainen & Junttila (2022: 47, 51). Most of these seem to be parallel loans or implausible etymologies. These loanwords will be discussed in more detail below.

3. Commentary on Ugric etymologies

In this section, problematic Ugric etymologies will be critically examined. First, the Iranian loanwords in Ugric are analyzed, and conclusions based on their phonological irregularities are presented; then the Ugric etymologies with possible Turkic connections are analyzed in the same way. After that, possible substrate words are discussed.

3.1. Indo-Iranian loanwords

Indo-Iranian loanwords in Ugric have been discussed in detail by Munkácsi (1901), Joki (1973), and Korenchy (1972), and some new etymologies have been suggested by Harmatta (1997). The earlier etymologies are also commented on in the unpublished dissertation of Lushnikova (1990). Recently the evidence has been reviewed by

Holopainen (2019), and some etymologies are also commented on by Kümmel (2020). However, further phonological details can be added to these earlier works, and not all of the suggested etymologies were noted by Holopainen (2019). It is thus an important part of Ugric etymology to present an up-to-date review of these etymologies. Also, Harmatta's (1997) suggestions were not taken into account by Holopainen (2019), and it is important to discuss them here, as Harmatta suggests an Iranian etymology for many Proto-Ugric words, including important cultural terms such as 'horse' and 'saddle'. Harmatta assumes that the donor language of these loans is "East Iranian", a hypothetical proto-language that differs notably from the commonly accepted reconstructions of Proto-Iranian.

The etymologies that are deemed implausible by Holopainen (2019) are not discussed here, as they do not offer useful evidence for the reconstruction of the Ugric lexicon.

3.1.1. Plausible etymologies

Hungarian *ár* 'stream'

Khanty: East (V) *lar* 'lake formed during flood [während des Hochwassers an einem Wiesenufer entstandener See]', (Vj) 'flat area that is flooded during spring [tiefliegendes, baumloses Wiesenufer od. Wiesengelände, das im Frühling überschwemmt wird]'; South (DN) *tor* 'lake'; North (O) *lar* 'area filled with water during flood; lake' < Proto-Khanty (Zhivlov 2006) *ʌār* (Honti 1982: *ʌar*)

Mansi: East (KU) *tūr*, West (P) *tūr*, (LO) *tor*, North (So) *tūr* 'lake' < Proto-Mansi (Zhivlov 2006) *tūrə* (*tōrə* in Zhivlov's 2023 system) (Honti 1982: *tūrɜ*)

< Proto-Ugric (?) *ϑara* 'lake (?)' < PU *sara* or *šara* (Zhivlov 2014: 127)

← Proto-Indo-Iranian or Proto-Iranian *saras-* 'lake' < Proto-Indo-European *selos-*

(MSzFE: 90–91; UEW: 843–844; PUg *ϑarɜ* 'während des Hochwassers entstandener See'; Koivulehto 1999: 2152; Holopainen 2019: 217–218)

Although this is a possible case of accidental similarity, as noted by Holopainen (2019), formally the Indo-Iranian etymology suggested by Koivulehto (1999) is plausible on both semantic and phonological grounds. The Ugric cognates have a regular relationship and it is quite clear that they can reflect a Uralic *a–a* stem (Zhivlov 2014: 127). If this word is indeed borrowed from Iranian (or Proto-Indo-Iranian), it shows the most characteristic sound substitution of Indo-Iranian *a* by Uralic *a*, and the usual adoption of the Indo-Iranian thematic stem as a Uralic *a*-stem. Regarding the consonantism, it is notable that the word had to have been borrowed before the change

*s > *ϑ happened in Proto-Ugric or Common Ugric. This is important, as it means that the word was borrowed at a time when the predecessor of the Ugric languages still retained the Proto-Uralic system of sibilants.

   The phonology of the Indo-Iranian donor form offers little help in the more precise dating of the loan, as Proto-Indo-Iranian *s was probably still retained in Proto-Iranian (Mayrhofer 1989), although not all details are clear. The old *s must have changed into *h in Common Iranian before Proto-Indo-Iranian *ć changed (via *ts) into *s, and because of this, some scholars (such as Cheung 2002, 2007) do reconstruct *h already for the Proto-Iranian stage. Formally the word could be a very early loan, but it should probably be older than the Ugric word for 'gold', discussed below, as the source of that word shows a stage of Iranian that already had a sibilant reflecting the Proto-Indo-European palatal stops.

   Hungarian *arany* 'gold'

   Khanty: East (V) *lorńə* 'copper', (Vj) *jorńi* 'brass, copper (from which samovars etc. are made)' < Proto-Khanty (?) *ʌorńV* (cf. Honti 1982: 139, No. 163)

   Mansi: East (KO) *tarəń* 'copper', West (P) *tarəń* 'copper', South (TJ) *tarə·ń* 'tin' < Proto-Mansi (?) *tarəń* (Honti 1982: 139, No. 163)

   < Proto-Ugric (?) *ϑVańa* < Pre-Ugric *sVrańa*

   ← Iranian *zaranya-* 'gold', Avestan *zaranya-* < Proto-Iranian *dzr̥Hanya-* (EWAia II: 816; AiWb: 1678, s.v. *zaranya-*)

   (MSzFE: 93–94; UEW: 843 PUg *ϑarańa*, *sarańa* 'gold; copper'; Abondolo 1996: 40)

This loan etymology is widely accepted (MSzFE: 93–94; Korenchy 1972: 77; UEW; Lushnikova 1990: 292; Abondolo 1996: 40; Holopainen 2019: 232–233), and it is clear that the words had to have been borrowed from the Iranian branch of the Indo-Iranian family: the *s which can be assumed for "Pre-Ugric", cannot reflect Proto-Indo-Iranian *ć, but it probably reflects later Iranian *z as is commonly assumed (Korenchy 1972: 77; Lushnikova 1990: 292); it cannot be ruled out that the donor language had Proto-Iranian *dz, but there are few examples of word-initial *dz in loanwords. Since we have no certain examples of *dz reflected as Ugric *s, it is more probable that the donor form already had a sibilant. This difference is rather crucial, as in this case the loan should be from a younger form of Iranian than the word for 'lake' discussed above.

   On the other hand, we have at least one loan that is confined to Khanty and Mansi but shows an affricate pointing to Proto-Iranian *dz. This is the word for 'reindeer calf' (Khanty (East) *pečəγ*, Mansi (North) *pāsiγ*) that is reconstructed as Proto-Ugric

päčäɣ by Zhivlov (2006), and according to Koivulehto (2007), it is a loanword from Proto-Iranian **patsu* 'cattle' (> Avestan *pasu*). It would be a problematic situation to assume that a word was borrowed into Proto-Ugric from an Old Iranian language that had sibilants, and a later loan in Proto-Ob-Ugric was borrowed from a more archaic, Proto-Iranian stage. This shows that the distribution of loans within Ugric is a rather problematic criterion that can hardly be used to distinguish different layers of loans, and it is possible and even probable that the Iranian loanwords in the Ugric languages, even apparently regular cases like the words for 'gold, copper', are parallel loanwords.

It is unclear how the vocalism of the Proto-Ugric word should be reconstructed, as the Khanty and Mansi dialects show aberrant vowel correspondences: Holopainen (2019) reconstructs **θ̢eraṅa* (and Pre-Ugric **seraṅa*) but this is erroneous and does not account for the Khanty and Mansi words regularly. Abondolo (1996: 40), who follows a different system of vowel reconstruction, reconstructs Proto-Ugric **u* (Proto-Ugric **surṅï*) for this word, but also this reconstruction is problematic.

It is notable that this word, like the word for 'lake' discussed above, was borrowed before the change **s > *θ*, so the recipient language was still phonologically identical to Proto-Uralic.

Hungarian *ostor* 'whip'

Mansi: South (TJ) *aśtə·r*, East (KU) *ōśtər*, North *ā̊śter* < Proto-Mansi **āćtər* 'whip' (Zhivlov 2013: 219)

< ? Proto-Ugric **oćtVrV*

← Proto-Indo-Iranian (Proto-Iranian?) **(H)ajtrā-* 'whip (?)' (EWAia I: 143; Wojtilla 2002)

(Munkácsi 1901: 494–495; MSzFE: 506–507; UEW: 333–334, 877; Korenchy 1972: 56–57; Holopainen 2019: 156–158)

This is a convincing etymology, but it seems probable that the Hungarian and Mansi words reflect parallel loanwords. Mansi **ā* can point to Pre-Mansi **o* (Zhivlov 2014: 124), but Hungarian *o* is usually not retained as such. However, PU **o* was regularly retained as Hungarian **o* before some clusters (Zhivlov 2014: 124). It is debatable whether Pre-Hungarian **a* can be reflected as Hungarian *o*: Proto-Uralic **paćka* 'shit' > Hungarian *fos* is a possible example of such a change, but no good parallels are known (furthermore, it is debatable whether **a* can be reconstructed for this Proto-Uralic word, as the vowel correspondences are problematic).

Regarding the consonantism, Mansi **ć* and Hungarian *s* do not correspond regularly to each other in inherited Uralic vocabulary, but some examples of such a correspondence are found in the shared Ugric vocabulary. Hungarian *s* can reflect the Proto-Uralic cluster **ćk*, and one could perhaps assume that *s < *ć* happened in some

other clusters as well, but there are no other examples of a *ćt cluster. The irregular correspondence in the consonantism points to parallel borrowing.

Mansi *ć is somewhat enigmatic, as one would not expect a palatal affricate from Iranian *š. Here one must note that although Proto-Mansi *ć is reconstructed as an affricate, most of the reflexes in modern Mansi languages are sibilants, and it is possible that the word was borrowed from Iranian at a time when the early varieties of Mansi had sibilants already.

As Proto-Iranian quite clearly had *št in this word, Zhivlov (2013) assumed that the Mansi word is a loan from the Andronovo Aryan, an unattested branch of Indo-Iranian that was first suggested by Helimski (1997). As *ćt assimilated to *št already in Proto-Indo-Iranian (Kümmel 2020: 242), it seems unlikely that a daughter language of Proto-Indo-Iranian had retained *ć in this cluster. We could assume that Mansi *ćt (? śt) is due to sound substitution, as the cluster *št was unknown in Mansi.

If the source form already had *št, it is difficult to say whether the donor language was Proto-Indo-Iranian, Proto-Iranian, or some later Iranian language, as the *št cluster was long retained in many languages, including Alanic. It is interesting that both Hungarian and Mansi words clearly denote 'whip': according to Wojtilla's (2002) philological investigation of the reflexes of Indo-Iranian *ajtrā-, it is probable that the original meaning was rather 'goad', but also the meaning 'whip' is attested in both Avestan and the Rig Veda. The sole meaning 'whip' is attested in some later Iranian languages like Middle Iranian. Wojtilla (2002: 589) mentions that this might point to a later Iranian origin of the Ugric words, but the situation is not quite clear.

Hungarian *hét* 'seven; week'

Khanty: East (V) *läwət*, South (DN) *tăpət*, North (O) *lăpət* 'seven; week' < Proto-Khanty *ʌ̄ăpət*

Mansi: South (TJ) *sāt*, East (KU) *sɔ̄̄t*, West (P) *sāt*, North (So) *sāt* 'seven; week' > Proto-Mansi *sā̄t*

< ? Proto-Ugric

← Proto-Indo-Iranian (Proto-Iranian?) *sapta-* 'seven'

(MSzFE: 283–284: PUg *säptɜ; UEW: PUg *ϑäptɜ, *säptɜ; Korenchy 1972: 70; Zhivlov 2018; Holopainen 2019: 239–240)

The Iranian etymology is accepted by all relevant sources on Ugric etymology, but the obvious irregularities are often ignored (MSzFE; UEW; EWUng; Abondolo 1996: 89). However, this etymology was considered a parallel loanword already by Sammallahti (1988). While the vowel relationships are regular, the word-initial consonants pose problems. Zhivlov (2018) notes that this word shows that the change *s > θ

happened earlier in Mansi than in Khanty: when the word was borrowed into Mansi, the change that had already happened was no longer productive, but in Khanty *s participated regularly in the change *s > θ. MSzFE and UEW argue that h in Hungarian *hét* is due to analogy from *hat* 'six', and this explanation is accepted by EWUng and Abondolo (1996: 88–89); if this is correct, then Hungarian *hét* with secondary h could formally be derived from the same Proto-Ugric form as the Khanty word, although here Khanty shows long *ā̈ that is not the usual reflex of Proto-Uralic *ä (this will be discussed in further detail below).

It is important to note here that the source form was quite archaic, retaining Proto-Indo-Iranian *s (the issue is similar to the Ugric word for 'lake' above). It is intriguing that here the Ugric languages show *ä as the substitution of Indo-Iranian *a, as the old "Common Ugric" loans usually show a back-vocalic substitution. It is unclear whether this points to a different layer of loans or not.

> Hungarian *titok* 'secret (noun)', arch. *titok* 'secret (adjective), der. *titkos* 'secret (adjective)'

> Mansi: North *tūjt-*, West *tujt-*, *tujt-*, *tunt-* 'hide' < Proto-Mansi *tūjt- (*tōjt- in Zhivlov's 2023 system)

> < ? Proto-Ugric *taja-ttV-

> ← Proto-Indo-Iranian (Proto-Iranian?) *ta(H)ya- 'secret (?)' > Av *taya-* 'stolen; secret', *tāya-* 'theft' (AiWb: 647)

> (Korenchy 1972: 72; UEW: 892 PUg *tajɜ-ttɜ-; Holopainen 2019: 269–270)

The Mansi word points to an old *a–a-stem. It is disputed whether Hungarian *i* can be regularly derived from *aj; at least some Proto-Uralic *Vj sequences in closed syllables do develop into monophthongs, and EWUng (s.v. *titok*) mentions *here*, *hím*, and *tetű* as possible parallels, but none of these involve the sequence *aj. Proto-Uralic *Vj sequences in Hungarian require further research (see Pystynen 2014 for a recent discussion). This means that the Ugric etymology cannot be considered as completely certain. Semantically the Ugric etymology is plausible, and if we accept the Hungarian vowel development, then formally a Proto-Ugric etymology can be reconstructed.

If the Iranian loan etymology is correct, it is interesting that Hungarian and Mansi reflect a similar derivative with a suffix *-ttV-. This makes it plausible to assume that the Iranian word was borrowed into a common proto-language of Hungarian and Mansi. The underived stem, which would be Proto-Ugric *taja-, is not attested in either Hungarian or Mansi.

The UEW and other etymological dictionaries list several Hungarian words that are allegedly somehow connected with Hungarian *titok*, but no explanations for the nature of these connections are given. It is unclear if Hungarian *tilt* 'forbid', *tilos*

'forbidden' belong here somehow: from the point of view of word formation, it is difficult to connect these forms in any regular way to *titok*. WOT (1234) mentions the possibility that *tilt* and *tilos* might be borrowed from a West Old Turkic verb **tiγil-* that is assumed on the basis of East Old Turkic *tidil-* 'to be obstructed'. This is at least phonologically a much better solution than connecting these Hungarian words with *titok*.

See Holopainen (2019: 269–270) for comments on other words erroneously connected here, such as Hungarian *tolvaj* 'thief'.

3.1.2. Problematic etymologies

Hungarian, *mély* 'deep', dial. *mél*, *míl*

Khanty: North (O) *mă̆l* 'Tief'; East (V) *mĕl*; South (DN) *mĕt* 'deep' < Proto-Khanty **mil*

Mansi: North (N) *mil*, East (KU) *mil*, South (TJ) *mäl*, West (P) *ńiĺ* 'deep' < Proto-Mansi **mil*

< Proto-Ugric **melV* ← Iranian (Alanic) **mal*, Ossetic *mal* 'Tiefe im Fluss'

(MSzFE: 432; UEW: 870 **melʒ* 'Tief'; Abondolo 1996: 26 **milV*; Sköld 1925: 30; Rasmussen 1989: 236)

The Iranian origin of *mély* 'deep' has been suggested by Sköld (1925: 30) but it is not supported by the UEW or any of the etymological dictionaries of Hungarian. However, the etymology is supported by Rasmussen (1989: 236), who discusses the Indo-European background of the Alanic donor form (Rasmussen assumed that Ossetic *mal* reflects Proto-Indo-Iranian **māri*, a derivative of Proto-Indo-European **mori* 'sea'; the alleged Proto-Iranian formation is not reflected by any other Iranian language). As noted by Holopainen (2022b: 199, footnote 2), the etymology is plausible as such, but it is dubious that the donor form has gone through the Alanic change **ry > l*; the other convincing Iranian loanwords shared by Ugric point to a more archaic Iranian donor language, and there are no convincing examples of a clearly Alanic loan shared by Hungarian, Khanty, and Mansi (this is also noted by Sköld 1925: 88 as a possible obstacle to the etymology).

The Ugric cognates are regular, and clearly point to Proto-Ugric/Proto-Uralic **e* in the first syllable. The Iranian etymology remains unclear, but as this Ugric word is regular, it has little value in the discussion of parallel borrowings. Sköld (1925: 44–47) has shown that in Alanic loanwords into Hungarian, Alanic **a* (which usually derives from Proto-Iranian **ā*) is reflected by Hungarian *é*; deriving Hungarian *mély* (< *mél*) from Alanic **mal* fits this rule, as noted by Rasmussen (1989: 236–237). The reason behind this is possibly that Hungarian *é* in these words was a more open vowel

at the time of Alanic contacts. However, this issue must be discussed elsewhere in more detail. The important thing here is that there are no parallel examples of Alanic loans showing a similar sound substitution in Khanty and Mansi. This is a further argument against the etymology.

On the other hand, a borrowing from some form of Ugric into Alanic/Ossetic does not look very probable, as Ossetic/Alanic *a* (< Proto-Iranian *ā*) would be difficult to explain from Ugric *e*. At the present moment it seems impossible to say more about the possible connection between the Ugric and Ossetic words.

Hungarian *aszik* 'dry'

Khanty: North (O) *săs-* 'dry, become hard', East (V) *sos-* , South (DN) *săs-* id. < Proto-Khanty *sįsā- ~ *sas*

Mansi: North (So) *tōs-* 'dry', East (KU) *tōš-* 'trocknen', West (P) *tōš-*, South (TJ) *tāš-* id. < Proto-Mansi *tāšā- ~ *tāš*

< Proto-Ugric *θasV-* < ? *saśV-*

← Iranian *Hsawš- *Hsáwšati*, cf. Avestan *haoš-* 'austrocknen, verdorren' (EWAia II: 658; AiWb: 1738, s.v. *haoš-*)

(MSzFE: 98–99; UEW: 844 *θasɜ-*; Katz 2003: 221–222; Junttila & Holopainen 2022: 314–315)

Katz (2003: 221–222) has suggested an Iranian origin for this Ugric word. The loan etymology is possible, inasmuch as the semantic connection is plausible, and there are no phonological problems regarding the vocalism: the sequence *aw* would have been simplified in Ugric, as such a sequence was not found in inherited words, as noted by Junttila & Holopainen (2022: 314–315). However, the consonantism is somewhat more problematic, as Hungarian and Khanty show word-internal *s* that regularly reflects earlier *ś*; the Indo-Iranian word had a retroflex sibilant *š*, and we would expect *š* on the Uralic side, as this is what we find in most loans (such as Uralic *mekši* 'bee' ← Pre-Indo-Iranian *mekš-* 'fly, bee'). It is interesting that this word shows Proto-Mansi *š*, similarly as some other Indo-Iranian loans, such as Mansi *šętV* 'hundred' ? < Proto-Uralic *ćęta* ← Proto-Indo-Iranian *ćatám* id. (see Pystynen 2013; Zhivlov 2023: 143). This could point to parallel borrowing.

It has been argued by Junttila & Holopainen (2022: 314–315) as well as Holopainen & Junttila (2022: 47, 51) that the Ugric words represent parallel loanwords from Indo-Iranian, but this is not necessarily correct: the vocalism of the Khanty word can be connected to the other Ugric cognates, as Proto-Khanty *į* is the high ablaut grade of *a*, triggered by the suffixal *-ā-*; Proto-Khanty *sas* could be reconstructed as the underlying form, which would then regularly correspond to Hungarian *aszik* and

Mansi *tāšā- ~ *tāš. However, due to the Mansi sibilant *š, it is a plausible idea that the Ugric words reflect parallel borrowings from Indo-Iranian.

Regarding consonantism, here Mansi and Hungarian show the "Ugric" change of *s > *θ, whereas Khanty shows retained *s; this is a regular development in Khanty in words that had word-internal *ś. It is probable that Khanty never went through a phase of *θ in such cases (this is assumed by the UEW) but *s was retained under the influence of the word-internal sibilant. This is an argument against the Proto-Ugric change *s > *θ.

If the etymology is correct, the word points to a very archaic Iranian donor that still retained Proto-Indo-Iranian *s, similarly as *sara 'lake'.

3.1.3. Unconvincing etymologies

This subsection includes a commentary on etymologies suggested in earlier research but not discussed by Holopainen (2019).

Hungarian *kedv* 'mood; wish, desire'

Khanty: East (J) *kĕnt* 'anger', South (DN) *kĕnt* 'anger, malice', (Ni) *kănt* 'anger' < ? Proto-Khanty *kint*

Mansi: West (LM) *känt*, North (So) *kantəŋ* 'angry' < Proto-Mansi *känt*

< ? Proto-Ugric *kVntV*

← East Iranian *känti*

(MSzFE: 370; UEW: 861–862 *kȫntз* 'mood; Laune, Stimmung'; Abondolo 1996: 34)

Harmatta (1997: 74) has suggested that the Ugric words are borrowed from Iranian: he reconstructs East Iranian *känti* and argues that this is based on Old Indo-Aryan *kānti-* 'desire, wish' (this form and meaning is given by Harmatta without any references). However, that the Indo-Iranian source suggested by Harmatta is very poorly attested, and the East Iranian reconstruction is based on a Sanskrit etymology that has very uncertain Iranian cognates: it is unclear why Harmatta reconstructs an *i*-stem *kānti-*, as the Indo-Aryan word is attested only in Classical Sanskrit *káṇṭaka-* 'Dorn' (EWAia I: 292). A possible cognate exists in the Iranian Sangleči language, where a form *kandāg* 'Dorn' is found: this could formally reflect Proto-Iranian *kantaka-*, but EWAia considers this connection uncertain, as the Sangleči word might also be a loan from Indo-Aryan. Because of this, it seems far-fetched to assume that such a word would have existed in an Iranian variety of the steppe.

Regarding the Ugric etymology itself, it is also not certain that Hungarian *kedv* is a cognate of the Ob-Ugric words. Even though the semantic connection between 'hate' and 'mood' is possible, and Hungarian *e* and the Ob-Ugric vowels do not correspond regularly: Proto-Mansi short **ä* and Proto-Khanty **i* point to an earlier **e-ä* combination (Aikio 2015a: 6), whereas Hungarian *e* points to **ä*. Furthermore, the origin of the Hungarian word-final -*v* remains uncertain.

Abondolo (1996: 34) assumes that the Ugric words are related to Hungarian *köd* 'fog', Permic **kįd* 'mist', and the Samoyedic words for 'smoke', such as Tundra Nenets *syun* (it is usually assumed that these reflect Proto-Uralic **küntV* 'smoke'), but this idea is impossible due to the mismatch in vocalism.

Hungarian *ló* (: *lovat*) 'horse'

Khanty: East (V) *loγ*, South (DN) *law*, North (O) *law* 'horse'

Mansi: South (TJ) *low*, East (KU) *lo*, West (P) *luw*, North (So) *luw* 'horse'

< ? Pug **luwV* ← East Iranian **loγə* < **vlaγə* < Proto-Iranian **bāraka-* (~ Middle Persian *bārag* 'horse') (Harmatta 1997: 72)

(MSzFE: 405–406; UEW: 863–864 PUg **luwɜ (luγə)* 'horse; Pferd')

The word for 'horse', a prime example of the Ugric equestrian vocabulary, involves various problems, some of which have been recently noted by Holopainen (2022c: 107–108). Harmatta (1997: 72) assumes that the Ugric word was borrowed from Iranian, but this etymology is highly problematic. The vowel correspondences are irregular. Hungarian *ó* alternating with *o* could probably reflect earlier **u*, but no exact parallels to this development are known. The vowel correspondences within Ob-Ugric, too, are irregular, and even Honti (1982) reconstructs simply **lV̄γ* as the Proto-Ob-Ugric form. Harmatta actually assumes that the Iranian loanword would have been borrowed into some kind of "Common Ugric", thus after Proto-Ugric had already diverged, but he does not comment on the phonological problems in detail.

Harmatta reconstructs East Iranian **loγə* as the source form, assuming the following developments: **loγə* < **vlaγə* < Proto-Iranian **bāraka-* 'horse'. Proto-Iranian **bāraka-* can be reconstructed on the basis of Middle Persian *bārag* 'horse' (a similar formation is also reflected in some other Iranian languages, see Bailey 1979: 270–271), but it is not the normal Iranian word for horse; most of the Iranian languages reflect Proto-Indo-Iranian **(H)aćwa-* 'horse', from Proto-Indo-European **h₁ekwos*.

Harmatta's explanation is problematic in that it is based on unattested East Iranian developments: the form **loγə* is assumed only based on the Ugric words and Proto-Turkic **ulag*, which Harmatta likewise derives from this Iranian word.[2] No

---

2. The problems in the Iranian etymology for Proto-Turkic **ulag* have been highlighted by Holopainen & Czentnár (2022).

inner-Iranian evidence exists. While all the Ugric words could probably reflect a form *loɣə in the source language, because of the hypothetical form, this etymology cannot be accepted.

Regarding the relationship within Ugric, one could argue here that *ɣ has influenced the vocalism somehow, but no parallels can be found. Abondolo (1996: 45) assumes that the Proto-Ugric reconstruction was *lïɣ, and the labial vowel in Hungarian is due to *ɣ. While such an explanation is not *a priori* impossible, reconstructing Proto-Ugric *i (*ï in Abondolo's notation) does not help with the irregular cognates in Ob-Ugric; neither the Khanty nor the Mansi cognates reflect regularly Proto-Ugric/Proto-Uralic *ẹ.

Due to the irregular vocalism, it seems most likely that the words for 'horse' in Khanty, Mansi, and Hungarian are parallel loanwords. It is difficult to argue for a loan when no source can be presented. However, a loan from an unknown source is a better explanation than an irregular cognate. Various loan explanations have been suggested, but none of them has received much support.

An old idea is that the Ugric word is a loan from Turkic *ulag 'horse'. This explanation has been refuted by Róna-Tas (1988: 749–750), who lists both phonological difficulties and semantic problems: the loss of word-initial *u- would be unwarranted, and the original meaning of the Turkic word seems to have been 'post horse, transport horse'. The semantic discrepancy is not necessarily an obstacle, but it is true that other, more fitting words for 'horse' would have been found in Proto-Turkic. However, as Proto-Ugric must have been a much earlier proto-language than Proto-Turkic (as will be discussed in more detail below), it is *a priori* implausible to assume contacts between these two stages. As the Hungarian, Khanty, and Mansi words for 'horse' are not regular cognates, it would be possible to assume that these are later loans from Turkic, but the aforementioned problems with phonology and semantics would remain in this case too, so this also seems unlikely.

Hungarian *magyar* 'Hungarian', *megyër* 'name of a Hungarian tribe'

Khanty: East (V) *mańt́* 'name of a Khanty phratry; name of the mythical ancestor of the members of this phratry', South (I) *mońt́*, North (O) *maś* id. < ? Proto-Khanty *mā̆nć

Mansi: North *mańśi*, South *mäńćī*, East *möäńś*, West *māńś* 'Mansi; child that has not been baptized' < ? Proto-Mansi *mā̆ńćV

← ? Proto-Indo-Iranian or Proto-Iranian *manuš* > Avestan *manuš-*, *manuš.čiθra-* personal name, 'descendant of Manuš', Old Indo-Aryan (Vedic) *mánu-* 'Mensch, Mann, Menschheit' (EWAia II: 309)

(MSzFE: 415–417; UEW: 866–867; Korenchy 1972: 60; Holopainen 2019: 133–135)

The idea that the shared ethnonym of the Ugric peoples is borrowed from Iranian is old, but this etymology is not very plausible. Contrary to Holopainen (2019: 133–135), who considered the etymology possible, I would like to suggest that the loan etymology be completely rejected. As ethnonyms often have a complicated background and the etymologies of many ethnonyms are not well understood, it is not at all certain that the ethnonym *magyar*, *mańśi* (etc.) has to be a loan from some known source. The alleged Indo-Iranian source denotes 'man, human', whereas this meaning is not attested as such in any of the Ugric languages. This is not the biggest problem of the etymology, however (even though this idea has been criticized by Zimmer 1990: 15, footnote 43). The cluster *ńć is a bigger hindrance, as it is not at all clear how this should have arisen from the Indo-Iranian form (this problem has been recently pointed out by Kümmel (2017)). Korenchy argues that the word would have been borrowed into Ugric as disyllabic, with loss of the second-syllable vowel, and as the resulting *nš cluster would have been impossible in Ugric, *ńć would have been the fitting substitution. This is a complicated explanation, and because the etymology is semantically not quite obvious, it is far-fetched to assume this chain of development to save the Indo-Iranian etymology.

Regardless of the Indo-Iranian etymology, it is clear that the alleged cognates in Khanty, Mansi, and Hungarian cannot be regularly derived from the same Proto-Ugric reconstruction, as noted by Holopainen (2019). The vowel correspondences are irregular, and also the reconstruction of a Proto-Ob-Ugric word faces difficulties, as is clear from Honti (1982: 164–165) who reconstructs both *mā̄ńć and *mānć as the possible predecessor of the Ob-Ugric forms.

The Hungarian ethnonym *magyar* also involves additional problems: it is usually assumed that the word is a compound, and while this is a plausible idea, the origin of the latter part (-*ar/ër*) is unclear. As the earliest attested forms of the ethnonym seem to reflect a form *mogyër*, the front vocalic variant -*ër* has to be the original form. Regarding the origins of -*ër*, different solutions have been suggested, but the traditional idea of deriving it from Uralic *ürkä or *irkä is impossible, as this etymology is completely implausible (as has been noted by Holopainen et al. 2019). Harmatta (1997: 78) suggested an Alanic etymology for -*ër*, assuming it is a loan from Alanic *ir, a reflex of Proto-Iranian *wiHra 'man'; this same word is found in the Ossetic ethnonym *Iron*. While this etymology seems promising (Hungarian *ë* usually reflects Pre-Hungarian *i), it cannot be accepted as such, as there is no evidence of word-initial *w- being lost in Alanic early enough, cf. the Hungarian word *özvegy* 'widow' that has to reflect Alanic *widæʒi (> Ossetic *idæʒ*). The possible relationship of Hungarian -*ër* to Turkic *ēr* 'man' (enigmatically mentioned by the UEW in connection to *ürkä) requires further research.

The possible relationship of this allegedly Ugric ethnonym to another alleged Proto-Ugric etymology, *mańćɜ, *maćɜ 'Märchen; erzählen; tale; tell (tale, story)' (UEW) is mentioned by the UEW, and this has been recently defended by Honti (2017: 129–147). However, as even the ethnonym *mVńćV cannot be reconstructed for Proto-Ugric, it is clear that the word does not stand in any derivational relationship with

another Proto-Ugric word. Also, the allegedly Proto-Ugric *mańćɜ, *maćɜ includes various problems: according to the UEW this is reflected by Hungarian *mese* 'tale', *mesél* 'tell', and Khanty (East) *mańí* 'tale', *mańí-* 'tell tales', South (DN) *mońí* 'tell', *mońí-* 'tell tales', and North (O) *maś* 'tale', *maś-* 'tell tales'. Hungarian *mese* cannot be derived from any back-vocalic reconstruction, the sibilant *s* reflects neither *ńć nor *ć regularly, and the correspondence Hungarian *s* ~ Khanty *ńć is completely irregular. It is clear that the Hungarian and Khanty words are not reflexes of the same Proto-Ugric word, but as the semantic connection is quite close, it is possible that the words are parallel loanwords from some unknown source.

Hungarian *nyerëg* 'saddle'

Khanty: East (VK, Vakr) *nöγər* 'saddle' < Proto-Khanty *nə̄̄γər*

Mansi: South (TJ) *näwrā̄* 'saddle', East (KU), West (P) *naγər*, North (N) *naʾir* 'saddle' < Proto-Mansi *nā̄γrā̄*

< ? PUg *närkV

← East Iranian (?) *nəγer* 'saddle (?)' < ? Proto-Iranian *niwara-, > Khotanese Saka *nyūrr* (Harmatta 1997; Ponaryadov 2022; cf. Bailey 1979)

(MSzFE: 481–482; UEW: 874 PUg *närkɜ 'saddle; Sattel')

Harmatta (1997: 72–73) assumes that this is an Iranian loanword and belongs to the same "East Iranian" layer of loans as the Ugric word for 'horse' discussed above. The same Indo-Iranian etymology has recently been suggested also by Ponaryadov (2022). As is the case with the words for 'horse', also this "East Iranian" etymology is highly problematic and should be rejected. First of all, the donor form *nəγer reconstructed by Harmatta is completely speculative. The reconstruction of *γ is not based on any Iranian evidence. The assumed meaning 'saddle' is also not based on real Iranian evidence: the Khotanese Saka word *nyūrr* that Harmatta mentions means 'harness' (Bailey reconstructs the Proto-Iranian predecessor of this word as *niwarn-). A meaning somewhat close to 'saddle' is only attested in the compound *aśā-nyūrrāna* 'with horse harness', but it is also quite clear that this does not really denote 'saddle', and it is dubious to assume that a meaning 'saddle' would have existed in an East Iranian language of the steppe.

Also, a Turkic etymology for the Ugric word has been suggested, but this is refuted by WOT (1210–1213), as the assumed Proto-Turkic source *ńeger is based on outdated reconstructions (such Turkic words for 'saddle' as Middle Turkic *egär* cannot reflect a Proto-Turkic word with a word-initial nasal). Because of this, it is likely that the "Ugric" and Turkic words for 'saddle' have nothing to do with one another.

The Ugric cognates are completely irregular, as has been noted by Zhivlov (2016: 300): neither the vocalism nor the consonantism shows regular correspondences. Zhivlov assumes that the words for 'saddle' in Ugric are loans from some unknown source. This is clearly the most plausible option. As a word connected to horses and riding, it is plausible to assume that the Ugric words for 'saddle' belong to the same layer of substrate or superstrate vocabulary as words like *ló* 'horse' (see also Holopainen 2022c: 108–109).

Already MSzFE, although cautiously supportive of the Ugric etymology, noted that the modern saddle could not have been known in Proto-Ugric times. This is an important argument: if the saddle is a recent cultural innovation, the borrowing of this term into Khanty, Mansi, and Hungarian cannot have happened very early. It is possible that even though the split of the predecessors of the Ugric languages must have happened early, these languages could have formed some kind of "areal-genetic unit" at a much later point, and loanwords like 'saddle' could have been borrowed during this time. MSzFE suggests the possibility that the word might have been borrowed between the Ugric languages. This is certainly possible, but it is very difficult to show the directions of borrowing, due to the difficulties in phonological reconstruction.

Hungarian *szekér* 'cart'

Khanty: East *likər*, *ikər* 'sled' < ? Proto-Khanty *Λīkər*

< ? PUg *säkVrV*

← ? Proto-Indo-Iranian *ćakarta-*, cf. Old Indo-Aryan (Vedic) *śakaṭa-*, *śakaṭī́-* 'cart, wagon' < ? *ćakarta-*, *ćakartī́* (EWAia II: 601–602)

(MSzFE: 576–577; UEW: 886 PUg (?) *säkɜrɜ*; Korenchy 1972: 73–74; Holopainen 2019: 256, 269)

Problems with the Ugric etymology are mentioned by the MSzFE and the UEW (886), and it is obvious that no Proto-Ugric word can be reconstructed, as the word-initial consonant correspondence between Hungarian and Khanty is irregular. It can be added that the vowel correspondence Hungarian *e* ~ Khanty *$\bar{\iota}$* is also irregular. The Indo-Iranian loan etymology is considered uncertain by Holopainen (2019), but it seems that even this is too optimistic and the etymology should be completely rejected (the Indo-Iranian etymology is mentioned as very uncertain already by UEW: 886 and is also doubted by Ligeti 1986: 151). Regarding the possible Proto-Ugric etymology, it is impossible to derive the alleged cognates from a Proto-Ugric reconstruction. Already Korenchy (1972: 73–74) notes that the Hungarian and Khanty words are not regular cognates and have to be parallel loanwords, as the word-initial consonants do not correspond regularly; Korenchy considered the words parallel loanwords.

However, according to Parpola (2020: 185–186), the Sanskrit word is a loanword from Dravidian *caṭṭakam. This makes the Indo-Iranian origin of the Ugric words very unlikely, as there is no evidence whatsoever of the word ever existing in any early Indo-Iranian variety of the Eurasian steppe.

Harmatta (1997: 76) suggests an alternative Iranian etymology, again assuming an "East Iranian" source but this also involves its problems. Harmatta assumes an Indo-Iranian root *sak- and reconstructs a noun *saka-tra- [sic]) to account for the Indic form (note that Harmatta reconstructs Proto-Indo-Iranian *s, which obviously cannot account for Sanskrit ś-), and as the origin of the Hungarian and Khanty words Harmatta reconstructs East Iranian *saka-kara-. However, this form is unattested in Iranian, so the etymology is implausible.

If the words are not borrowed from Iranian, they could perhaps belong to the same layer as the Ugric equestrian vocabulary. However, the Khanty word denotes 'sledge', and it is not clear that it was originally connected with equestrian nomadic culture.

Hungarian *szó* 'word'

Khanty: East (Trj) *săɣə*: *på̊ŋkəʌsăɣə* 'song that is sung by the soothsayer after consuming fly agaric', South (Ko) *săw* 'melody', North (O) *săw* 'voice' < Proto-Khanty (?) *sịɣ* (Honti 1982: 182, No. 574: Proto-Khanty *soɣa*)

Mansi: ? East (K) *sauŋ*, North (So) *sow* 'melody', (N) *så̊w* 'word; sound, voice' < Proto-Mansi *saw* (Honti 1982: 182, No. 574)

< ? Proto-Ugric *sawV* 'word, speech'

← ? "East Iranian" *savä* < Proto-Indo-Iranian *srawas* (Harmatta 1997: 74)

(MSzFE: 591–592; UEW: 885 PUg *saw*)

The Indo-Iranian etymology suggested by Harmatta (1997) suffers from similar problems as the other allegedly "East Iranian" words. The reconstructed form is entirely hypothetical and not based on any attested East Iranian evidence. A loan from some other Iranian form, such as Proto-Iranian *srawah/srawas*, also seems plausible. Although formally it might be possible to derive Hungarian *szó* from a reconstructed Indo-Iranian word that had a word-initial cluster *sr-, with a simplification of the cluster as *s-, such a development lacks parallels. Moreover, the Ugric words have competing Uralic etymologies.

In addition to the Indo-Iranian etymology, also a Turkic etymology for the Proto-Ugric word has been suggested: it has been assumed that the Ugric word is a loan from a Proto-Turkic word akin to Old Turkic *sav* 'Wort, Rede'. Both MSzFE and UEW still list the Turkic etymology, but this idea is refuted by Róna-Tas (1988) in

connection with other possible Turkic borrowings in Proto-Ugric. Recently however it has been argued (WOT: 808–810) that the words in Hungarian and Ob-Ugric are separate borrowings from the same Turkic source.

3.1.4. Conclusions on Iranian loanwords

It seems that some Iranian loans formally go back to Proto-Ugric, but there are also many words that clearly show irregularities. It is plausible to assume that the Iranian loanwords were not acquired into Proto-Ugric but "Common Ugric", that is, into already divergent dialects or languages (early forms of Hungarian, Khanty, and Mansi).

It is not at all clear that the Iranian loanwords reflect one layer. Since the number of loans is quite small, it is theoretically possible to assume that there was not even any period of separate Ugric–Iranian contacts, but rather all the loans confined into Ugric simply reflect retentions from the earlier period of contacts between Indo-Iranian and "Common Uralic". However, since some borrowings show characteristic cultural influence ('whip' being related to horses, and 'seven' also showing strong cultural influence), it is probable that there was a separate period of Ugric–Iranian contacts. However, it is not clear whether this was a single contact episode or consisted of separate periods. The loans probably do not reflect only one layer, as the word for for 'gold' shows quite clear "Old Iranian" phonology, whereas the other loans point to a very archaic donor.

3.2. Shared vocabulary between Ugric and Turkic

The idea of Turkic loanwords in Proto-Ugric has a long research history, but in more recent research, the idea of Turkic borrowings in Proto-Ugric times has been criticized by Róna-Tas in several publications (1988; WOT). His reasons relate both to chronology (Proto-Turkic clearly was quite a more recent proto-language, showing more shallow time depth than Proto-Ugric), and to the various problems with the suggested etymologies. These problems are discussed below in more detail. Instead of assuming borrowing from Turkic into Ugric, Róna-Tas argues that at least some of the words in question are northern Eurasian cultural words (or Wanderwörter). He thus assumes some kind of connection between the Ugric and Turkic words. For some words, like Hungarian *szó* 'word' (already discussed above), Róna-Tas assumes an accidental resemblance with the Turkic words.

Although Róna-Tas's arguments seem quite convincing, it is important to revisit these etymologies and analyze them in the light of an up-to-date view of Uralic historical phonology. It is important to stress that even if the words are Wanderwörter in Ugric and Turkic (as well as the other so-called Altaic languages), it has not been convincingly demonstrated that the words in question go back to Proto-Ugric, and it is plausible to assume that like the Iranian loans, also these words are independent borrowings from some source into the already diverged Ugric languages.

The problems with the possible Turkic origin of the Ugric words for 'horse' (Hungarian *ló* etc.) and 'saddle' (Hungarian *nyerёg* etc.) were already discussed above, as these problems are related to the Indo-Iranian etymologies suggested for these words.

3.2.1. The etymologies

Hungarian *hattyú* 'swan'

Khanty North (O) *χŏtəŋ*, East (V) *kŏtəŋ*, South (DN) *χŏtəŋ* 'swan' < Proto-Khanty **kutəŋ*

Mansi: North (So) *χotaŋ*, East (KU) *χotəŋ*, West (P) *kotəŋ*, South (TJ) *kotā·ŋ* 'swan' < Proto-Mansi **kutaŋ*

< ? Proto-Ugric **kottɜŋɜ* 'swan'

← ? Proto-Turkic **kutan* 'swan' (see WOT: 353–355 with references)

(MSzFE: 278; UEW: 857 **kottɜŋɜ*; Abondolo 1996: 49: **kutёŋV*; Róna-Tas 1988: 749; Katz 2003: 313)

The Turkic origin of the alleged Proto-Ugric word is still listed as a possibility in the UEW. This has been criticized by Róna-Tas (1988: 749), who argues that Finno-Ugric **tt* could be substituted by Turkic **t* but not vice versa, meaning this is a loan from Ugric into Turkic. A similar criticism is also expressed by Katz (2003: 313).

However, the situation is not so simple. In order to assume a loanword from Ugric into Turkic, we have to be able to show that the word can indeed be reconstructed into Proto-Ugric. It seems that there are phonological problems that make this assumption unlikely.

To start with the consonantism, Khanty and Mansi *t* can point to either **t* or **tt* in Proto-Ugric or Proto-Uralic, as the single and geminate **t* merged in these languages. The Hungarian palatal geminate *tty* does not reflect regularly either **t* or **tt*, and in fact this geminate sound does not reflect any other Proto-Ugric or Proto-Uralic sound either. Also the word-final consonantism presents problems: Khanty and Mansi have *ŋ*, whereas in Hungarian the assumed word-final consonant has been vocalized. Although the development of Proto-Uralic **ŋ* in Ugric is not sufficiently well understood (cf. Bakró-Nagy 2003), Khanty and Mansi do not usually have *ŋ* reflecting Proto-Uralic **ŋ*. This makes the idea of reconstructing Proto-Ugric **ŋ* dubious. There are also no known parallels to the development **-ŋ > ú* in Hungarian. It is possible that Hungarian *ú* and Ob-Ugric *ŋ* reflect later additions to a Proto-Ugric stem, but it is difficult to prove this, as these elements are not known derivational suffixes.

Turning to the vocalism, Hungarian *a* usually reflects Proto-Uralic **o* in -*i*-stems or Proto-Uralic **a* (Zhivlov 2014: 124). Some cases of **u > a* are found, but the details are uncertain and many alleged examples of this development are false etymologies or based on false reconstructions. Mansi shows *o* in all varieties, which reflects Proto-Mansi **ɔ*, while Khanty *ŏ* points to Proto-Khanty **u* in Zhivlov's system. Words showing similar vocalism include the reflexes of Proto-Uralic **tulka* 'feather' and **tukti* 'Querholz'. A Proto-Ugric stem **kuttV* or **kutV* could be reconstructed to account for the Ob-Ugric forms, but the origin of **ŋ* remains unclear.

In WOT it is assumed that the Ugric and Turkic words are northern "wandering words". This seems a more plausible explanation than to assume a loan from Proto-Ugric into Proto-Turkic. Even though the Ob-Ugric words show regular vowel correspondences, the problems with the Hungarian word mentioned above mean that it is not clear that a Proto-Ugric word can be reconstructed. It is possible that the word for 'swan' was borrowed separately into the Turkic languages and into the Ugric languages, and there is reason to assume that at least Hungarian borrowed the word separately from Ob-Ugric.

A Wanderwort is the most plausible explanation here; it is possible that the Ugric and Turkic words are connected, but this is far from certain.

Hungarian *hód*

Mansi: North (So) *χuntəľ* 'Maulwurf', East (KM) *k̥ontəľ*, West (P) *kuńťəľ*, South (TJ) *końťəl* 'beaver' < (? Proto-Mansi **kuńťə-l*)

< ? Proto-Ugric **kumtɜ* (**kuntɜ*) 'beaver'

← ? Proto-Turkic **qumtuz* 'beaver'

(MSzFE: 289; UEW: 858–859; Róna-Tas 1988: 750; Katz 2003: 313)

Contrary to the word for 'swan' that Róna-Tas considered a Ugric loanword into Turkic, the word for 'beaver' he considers "dubious" (Róna-Tas 1988: 750). His reason for this is not very well founded, however: Róna-Tas notes that it is unlikely that the Turkic word would have been borrowed into Ugric without the suffix -*uz* (> *ur*). This is not a very good argument, as there were no words ending in a sibilant in Proto-Uralic/Proto-Ugric. Many Indo-Iranian loanwords also have been borrowed as vocalic stems, even if they ended in a consonant in the donor language. Also Katz (2003: 313) doubts the etymology.

However, there is some unclarity in the reconstruction that indeed makes the idea of a loan from Turkic into Proto-Ugric unlikely. Proto-Uralic/Proto-Ugric **u* is usually not reflected by long *ó* in Hungarian (note that in this word, *ó* does not alternate with short *o*, as in words like *ló*: *lovat* where long *ó* is due to later contraction). The Mansi word shows Proto-Mansi **u*, which can reflect Proto-Uralic **u*. Mansi

shows an unclear suffix -*l*, while no traces of such a suffix are found in Hungarian. The cluster *nt* in Mansi corresponds regularly to Hungarian *d* that can reflect earlier **nt* – the UEW reconstructs **mt* or **nt*, the Turkic loan would presuppose **mt* but as these clusters merged in both Hungarian and Mansi, formally either option is possible.

As is the case with the word 'swan' above, the phonological difficulties involved in the reconstruction of the Proto-Ugric word are rather small, but it is still true that the correspondence Hungarian *ó* ~ Mansi **u* is not what we would expect in inherited words. Because of this, it is better to treat these words as not cognate.

A possible solution to explain the relationship of these words is to assume that Hungarian *hód* and Mansi **kuńťə-l* are parallel borrowings from Turkic. There is no phonological obstacle in explaining the Hungarian and Mansi words as parallel loans from Turkic, but as the etymology of the Turkic word is unclear, it is also possible that the word is a parallel loan into Proto-Hungarian, Proto-Mansi, and Turkic from some unknown source (North Khanty *χundịl* 'mole' is a loan from Mansi, as is stated by the UEW).

Hungarian *ír* 'write'

Khanty East (V) *jeri-* 'signal', (Trj) *jărip-* 'make, carve a sign (einen Strich ziehen, kratzen (Nagel, Holzsplitter usw.), riefeln, linieren (z.B. Papier, nur einen Strich; schreiben (Geheimwort))' < ? Proto-Khanty **jeri-* ~ **järi-*

< ? PUg **jarɜ-* 'write, draw, paint; schreiben, zeichnen, malen'

(MSzFE: 321–322; UEW: 850 **jarɜ-*)

The Ugric etymology involves various problems, and different competing etymologies for the Hungarian and Khanty words have been presented, in addition to the idea that the Proto-Ugric word could have been borrowed from Turkic. Already DEWOS (404) notes the irregularities between the Hungarian and Khanty vowels and is critical of the etymology. The vowel correspondences are indeed problematic, and the UEW's reconstruction **jarɜ-* is ad hoc and it does not account for either the Hungarian or Khanty: Hungarian *í* cannot reflect Proto-Ugric **a*, and the Khanty vocalism does not reflect any back vowel regularly.

The Hungarian word has competing Turkic etymologies, of which the latest suggestion by Róna-Tas (1992; see also WOT: 464) seems the most convincing. An alternative Turkic etymology is supported by WOT, and this seems to be the most convincing option: WOT reconstructs West Old Turkic **ïr-*, cognate to the attested East Old Turkic *ïr-* 'to make a notch, or breach, in (something)' (see also Károly 2001 for a discussion of this Turkic word). This etymology is more convincing both semantically and phonologically. WOT notes the Hungarian derivative *irdal-* 'to slit in, to mark out, to make cuts into (mostly into the skin of the fish before roasting)' which has retained the original meaning.

The UEW lists the traditional etymology (initially suggested by Gombocz and discussed in detail by Sz. Kispál 1951: 49–52) which derives the alleged Ugric word from Proto-Turkic *yar- 'write', reflected by Chuvash śïr id. This etymology was doubted by TESz (II: 228). Also WOT (464) rejects the etymology due to phonological difficulties, as the word-initial consonant causes problems. Noting the discrepancies in phonology, Kispál (1951) assumed that the Turkic word was first borrowed into Khanty and then from Khanty to Hungarian, but this explanation remains speculative. Sinor (1977) offers an alternative solution, suggesting that Hungarian ír was borrowed from an early Turkic form *śïr akin to Chuvash: the word would have been borrowed into Pre-Hungarian after *ś > s had already happened, and *ś would have been substituted by a Pre-Hungarian sibilant that later disappeared. This explanation is interesting but lacks parallels.

The origin of the Khanty word is more problematic. According to WOT, the Hungarian and Khanty words might be explained as parallel loanwords from Turkic, but the Khanty word is not discussed in detail by WOT. A Turkic origin for the Khanty word does not seem very probable. The word-initial j- is difficult to explain from Turkic *ïr-, and also there are no certain examples of very early Turkic loanwords in Khanty, so a loan from early Bulgar Turkic *yar- would be unlikely. Futáky (1975: 74) notes that also a Tungusic origin has been suggested for the Khanty word;, he considers the etymology uncertain, noting that since the Khanty word should be a very early Tungusic loan and Pre-Khanty–Tungusic contacts have not been properly investigated, the etymology cannot be supported as such. The assumed Tungusic origin is reflected by, for example, Nanai ńiruči- 'write'. It seems dubious that Khanty *j- could result from Tungusic *ń-, and the Tungusic origin of the Khanty word remains problematic.

3.2.2. Conclusions on Ugric–Turkic lexical connections

The conclusion of Róna-Tas (1988) that no Turkic word was borrowed into Proto-Ugric is confirmed by this study. However, his idea that some words were borrowed in the opposite direction, from Ugric to Turkic, is not supported. The Ugric words for 'swan' and 'beaver' involve irregularities that show they are not real cognates in Ugric. It is probable that they are parallel loanwords into Ugric and Turkic from some unknown source, although it also cannot be ruled out that at least 'beaver' was borrowed separately into Mansi and Hungarian from Turkic.

3.3. The Ugric horse vocabulary: substrate words or loanwords?

The major dictionaries of Hungarian, Ob-Ugric, and Uralic etymology consider the words related to horses to be Proto-Ugric items. This view is also reflected by Honti (1997, 2013) in his studies on Ugric unity, and Parpola (2012) repeats the arguments in his discussion of the early spread of Uralic. However, it has been noted already earlier (Hajdú 1962: 85) that these words might be loans somewhere, and Harmatta has specifically argued for the Iranian origin of some of these words. As noted above,

the Ugric words for 'saddle' (Hungarian *nyerëg* etc.) have been considered loanwords from an unknown source by Zhivlov (2016: 300). We also saw above that the words for 'saddle' and 'horse' cannot be real Proto-Ugric cognates but they look like loanwords from some unknown source. It is important to give similar scrutiny to other horse-related words as well.

A more thorough look at these words shows that the equestrian vocabulary of Ugric fulfills Salmons' criteria for substrate vocabulary quite well. Below all the relevant etymologies are analyzed in detail.

Hungarian *fék* 'halter; brake'

Khanty East (V) *päk* 'rein (of reindeer); bridle (of horse) (des Pferdes)' < Proto-Khanty *$p\bar{a}k$*

Mansi (Verchoturje) *nexъ* 'dozen bridles (дюжина уз(д)а)'

< ? PUg *päkkV*

(MSzFE: 191; UEW: 878 PUg *päkkз*)

The comparison presented in the UEW involves problems with vocalism: Hungarian *é* points regularly to Proto-Ugric/Proto-Uralic *e* (the only examples of Hungarian unalternating *é* as the reflex of earlier *ä* are only found in word-initial position, cf. *ének* 'song' < PU *äni* 'voice', ? *ég* 'burn' < PU *äŋV*), so Ugric *pekkä* or *pekkə* would be the regular pre-form of Hungarian *fék*. However, the East Khanty cognate *päk* (< Proto-Khanty *$p\bar{a}k$*) rather points to Ugric/Uralic *ä*, not *e*, although some examples of Proto-Khanty *$\bar{a}$* reflecting Proto-Uralic *e* are known (Aikio 2015a: 14–15; see also the discussion of Proto-Khanty *kint* 'anger' above). It is difficult to assess the Khanty word in more detail, as its dialectal distribution is very limited.

The Mansi cognate *nexъ* 'дюжина уз(д)а' is attested only in the 18th-century word-lists, so it is difficult to say anything certain about its phonology.

Even if the word went back to Proto-Ugric, it is true that even though all the meanings point to some kind of 'bridle', the word did not necessarily refer specifically to bridles used with horses originally. This means that it is uncertain whether we are indeed dealing with a substrate word borrowed from an equestrian culture, similarly as the words like *ló* 'horse'.

Hungarian *mén*, dial. *mín* 'stallion'

Khanty: (Ni) *manəŋ* 'herd of reindeer, horses or cows' < ? Proto-Khanty *$m\bar{a}nəŋ$*

< ? PUg *mänV* 'some animal'

(MSzFE: 433–434; UEW: 869 PUg ? *mänз*)

This etymology is considered uncertain by MSzFE and the UEW due to problems with vocalism: Hungarian *é, í* does not regularly reflect PU/PUg **ä*. Furthermore, the Uralic background of Proto-Khanty **ā̆* is unclear but it can reflect Proto-Uralic **e* in *ä*-stems (see Aikio 2015a: 15 and the discussion of the vocalism of Proto-Khanty **ā̆mp* 'dog' below). The semantic connection between the Hungarian and Khanty words is quite vague, and even if these words were real cognates, it is doubtful whether the word referred to horses originally. Contrary to words like *ló* 'horse' or *nyerёg* 'saddle', this word is less likely to be a substrate loan. It is quite possible that the words are completely unrelated and the similarity is accidental.

Loan etymologies for both Hungarian and Khanty words have been suggested. Futáky (1975: 70–71) suggests that the Khanty word is a loan from (North) Tungusic **mānï ~ *manï* 'flock, herd' (> Evenki *man*, Nanai *māndo* 'flock of birds'). This would be semantically more convincing than the alleged connection with Hungarian *mén* 'stallion'. However, further research on the phonological part of this loan etymology is required.

The Hungarian word has also an Iranian etymology: it has been suggested that the word *mén* is a loan from a predecessor of Ossetic *moj, mojnæ* 'husband' (Munkácsi 1904); it has been assumed that Ossetic *moj, monjnæ* reflects Proto-Iranian **manu-* 'man'. However, this etymology was criticized already by Sköld (1925: 30), and it is not accepted by later dictionaries (TESz, MszFE, UEW). Helimski (2002) mentions the etymology as a possibility, although he notes that the semantic problem (no meaning 'stallion' is attested in Ossetic) makes the etymology uncertain. The Alanic etymology is based on an outdated Iranian reconstruction: Cheung (2002) argues that the Ossetic word is derived from **dmanya-* 'related to a house', and it would be impossible to derive it from **manu-*. It is thus clear that the Hungarian word cannot be borrowed from Alanic. Also, a loan from some other reflex of Proto-Iranian **manu-* is also improbable, due to semantic problems.

Hungarian *fű* (: *füvet*) 'grass'

Khanty: East (V) *pam*, South (DN) *pum*, North (O) *pam* 'grass, hay' < Proto-Khanty (??) **pām* (see Honti 1982: 177, No. 518)

Mansi: South (TJ) *pom*, East (KU) *pom*, West (P) *pum*, North (So) *pum* 'grass, hay' < Proto-Mansi (?) **pumV* (see Honti 1982: 177, No. 518)

< ? PUg **pimV*

(MSzFE 223–224; UEW: 879–880 PUg **pimз* 'grass; Grass'; Abondolo 1996: 111: **pï(i)mï* 'grass')

The word for 'grass' is not an equestrian word as such, but its phraseological use is often mentioned (e.g. Honti 2013: 104) as part of the proof for Proto-Ugric horse nomadism.

A plant name is otherwise a good candidate for a substrate word. Moreover, as is the case with the horse vocabulary, the alleged Proto-Ugric word for 'grass' displays irregular vowel correspondences.

This word shows irregular vowel correspondences. Abondolo (1996: 111) assumes that the labial vowels in Ob-Ugric are due to word-internal *m, but parallel evidence would be needed before such an explanation could be accepted.

Hungarian *tegëz*, *tëgëz* 'quiver'

Khanty: V *tüγət*, DN, O *tiwət* 'quiver' < ? Proto-Khanty *tīγət*

Mansi: KU *tā̄wət*, LM *täut*, So *tawt* 'quiver' < ? Proto-Mansi *tā̄wət*

< ? PUg * *täŋVtV*

(MSzFE: 624; UEW: 894 PUg *täŋɜ-tɜ* 'quiver (on belt); Pfeilköcher am Riemen'; Abondolo 1996: 54)

Similarly to *nyerëg*, this word shows strange consonant correspondences (Hungarian *g* corresponding to *γ*, *w* in Ob-Ugric) and also irregular vocalism (even within Hungarian, unclear alternation of *e* and *ë* is found). It is quite probable that the word is a loan from somewhere, and we can assume that it belongs to the same layer as other items referring to horses. Kálmán (1988) notes that the word was borrowed from Iranian, but this is erroneous: though Kálmán does not mention the source form, the UEW mentions that Old Indo-Aryan *dhākás* 'container' has been considered a possible origin, but this is refuted by the UEW (this would be phonologically impossible).

Abondolo (1996: 54) assumes either secondary lengthening or retention of length in Khanty and normal "rotation" in Mansi. However, this Ob-Ugric correspondence is not regular. Hungarian *e* from the *ü* reconstructed by Abondolo is irregular, even though Abondolo assumes (based on Tálos 1983) that this is a normal development in words derived with "vocalic suffixes".

This word fills several of the criteria for a loan/substrate, as the word shows irregular phonology and word formation.

3.4. Other possible substrate words

It is fruitful to look for other possible substrate words fulfilling the criteria described above. Consider the following words.

Hungarian *eb*, *ëb* 'dog'

Khanty: East (V) *ämp*, South (DN) *ämp*, North (O) *ämp* 'dog' < Proto-Khanty *ā̈mp*

Mansi: South (TJ) *āmp*, East (KU) *ɔ̄mp*, West (P) *āmp*, North (So) *āmp* 'dog' < Proto-Mansi *ā̄mp(ə)*

< ? PUg *ämpV*

(MSzFE: 137; UEW: 836 PUg *ämpꝫ* (*empꝫ*) 'dog; Hund'; Abondolo 1996)

As is noted already by MSzFE, the word shows irregular vocalism: both *ëb* and *eb* are found in Hungarian dialects. It is not quite clear what the reason for this alternation is. The Mansi word could regularly reflect a proto-form *ämpV*, but the long *ā̄* in Khanty is unexpected, as the usual reflex of Proto-Uralic *ä* in Khanty is short *ä* (cf. Proto-Uralic *pälä* 'side, half' > Proto-Khanty *päḷək* > East Khanty (Vahk-Vasjugan) *peḷək* id., UEW: 362–363; Aikio 2015b: 63), although some examples of Khanty long *ā̄* as the reflex of Proto-Uralic *ä* are known, such as East Khanty (Vahk, Vasjugan) *äl* 'armful' (< Proto-Khanty *ā̄l*) < Proto-Uralic *älä* 'lap' (UEW: 23; Aikio 2015b: 62).

As a "tendency" to open close or mid vowels is known from Old Hungarian (Abaffy 2003: 327–331), it is more likely that *ëb* reflects the more original form, while *eb* is secondary.

Due to the irregular vowel correspondences, we can assume that this word is also a substrate word or a loan from some unknown source. It seems probable that these words have replaced some older word for 'dog', namely the word *penä* 'dog', found in the so-called Finno-Permian languages, could well be the original Uralic word for 'dog' that was replaced in the Ugric and Samoyedic languages by later loans (Samoyedic has *węn*, which possibly represents a loanword from Pre-Tocharian *kwënə* (oblique stem of *ku* 'dog') according to Kallio 2004).

Hungarian *epër*, *epërj* 'strawberry'

Mansi: West (P) *ā̄pərjēχ*, (LU), *ā̄prä*, (L) *ā̄prəχ*, South (T) *äpə·rjek* 'strawberry' < Proto-Mansi *ā̄pərjēk*

< ? PUg *äppVrVkv*

(MSzFE: 159; UEW: 836 PUg *äppꝫ-rꝫ-kꝫ* 'berry, strawberry')

The Hungarian and Mansi words for 'strawberry' are clearly connected somehow, but due to the semantic field it is worth investigating whether this Proto-Ugric word could rather represent a loanword. The vowel in the initial syllable in both Hungarian and Mansi points to Proto-Ugric/Proto-Uralic *ä*. However, the alleged suffixal elements are of unclear origin and function, which points to a loan. Also, semantically this is a good candidate for a substrate word.

It cannot be completely ruled out that Hungarian *epër* and Mansi *ā̆pərjēk* are cognates, but due to the uncertain derivational suffix, a substrate origin for this word remains a good possibility.

Hungarian *harkály* 'woodpecker'

Khanty: North (O) *χă̆ŋra*; East (V) *kă̆jərkị*; South (Kr) *χă̆χraj* 'woodpecker' < Proto-Khanty < Proto-Khanty (??) *karŋaj* (cf. Honti 1982: 156, No. 324)

Mansi: North (So) *χɔ̄ŋχra*, East (KU) *χōrχəj*, South (TJ) *karkāj*, West (P) *korkəj* 'woodpecker' < Proto-Mansi (??) *kārkāj* (cf. Honti 1982: 156, No. 324)

< ? PUg *karV*

(MSzFE: 267–268; UEW: 855 PUg *karɜ* 'woodpecker'; Abondolo 1996: 80: *karV* 'woodpecker')

This word shows various irregularities in both vocalism and consonantism and also features unclear suffixes. The UEW's Proto-Ugric reconstruction of a stem *karV* is ad hoc, as this cannot explain the word-initial *j/χ/ŋ* found in different Khanty varieties. Honti's (1982) Proto-Ob-Ugric reconstruction *kōrɣāj is* problematic for the same reasons and it is unclear whether even Proto-Khanty or Proto-Mansi words can be reconstructed. The alleged derivational suffixes in Khanty and Mansi do not correspond to each other regularly. Like the Hungarian and Mansi words for 'strawberry' discussed above, this word involves unclear derivational elements, which is typical for substrate vocabulary. Riese (2001: 47–48) follows the UEW in considering this a Proto-Ugric derivative, but it is unclear what the suffix here is: although Mansi *-ɣāj* is a real suffix occurring in some animal names, it is disputed whether this Proto-Ugric (?) word includes a reflex of this same suffix.

Despite the obvious irregularity, Abondolo (1996: 80) considers the Ugric words cognates and assumes that a metathesis has taken place in Khanty. Abondolo considers it possible that the Proto-Ugric word is a derivative, but he also mentions the possibility that the word is a compound consisting of an agent noun *karV-ja* (derived from an alleged Proto-Uralic verb *karV-* 'bite') and (in Abondolo's reconstruction) *kïjV* 'little bird' (a word attested only in Ob-Ugric). This explanation is not convincing, as the reconstruction of a PU verb *karV-* is uncertain (also according to the UEW) and because the word *kïjV* 'little bird' is not attested outside of Khanty and Mansi, making its existence at earlier levels of reconstruction, such as Proto-Ugric, purely speculative.

Although the words for 'woodpecker' in the three Ugric languages are most likely somehow connected, it is clear that no Proto-Ugric word can be reconstructed. This word fulfills several of the criteria for a substrate word: irregularity, suitable semantic field, as well as obscure word formation.

Hungarian *gyökér* 'root'

Mansi: South (TJ) *jükār̄*, East (KU) *jēk̥ər*, West (P) *jēkər*, North (So) *jēk̥ar* < ? Proto-Mansi *jükVr* 'root of a fallen tree'

< ? PUg *jVkkVrV*

(MSzFE: 231; UEW: 852 PUg *jökkɜ-rɜ*; WOT: 1318; Honti 2017: 98–99; Róna-Tas 2017: 63)

WOT doubts the Ugric etymology and notes it is irregular, but the etymology is defended by Honti (2017: 98–99). The South Mansi vocalism is apparently irregular, so it is unclear how the Proto-Mansi word should be reconstructed. The change from *j* to *gy* is found in very few inherited words in Hungarian (*jalka* 'foot' > *gyalog* 'on foot' being one possible example, even though also this etymology has been doubted, see Róna-Tas 2017: 63).

Due to the problems with the vocalism and Hungarian *gy*, reconstruction of a Proto-Ugric word seems impossible in this case as well. It is possible that this is another substrate word in Proto-Ugric. The meaning 'root' fits a semantic field of substrate vocabulary well, and in addition to the irregular phonology, also the obscure word formation points to a foreign origin. An alleged derivational suffix *-r-* is found in some Ob-Ugric etymologies, but its function is uncertain (Riese 2001: 52), and in the case of Hungarian *gyökér*, Mansi *jükār̄*, it is impossible to analyze the word as a derivative.

Hungarian *köles* 'millet'

Mansi: North (So) *kolas* 'Mehl' < Proto-Mansi *küläš*

< ? PUg *külVćV* 'millet'

(MSzFE: 370–371; UEW: 861 PUg *kölɜ (kölɜ-ćɜ), kölɜćɜ* 'millet; Hirse'; WOT: 1282, 1313 PUg *köläćə*)

This word features an unclear suffix which makes it a possible loanword. Riese (2001: 43) mentions the word among Proto-Ugric derivatives, but the function of the suffix remains unclear. Also, the semantic field (name of an edible plant) makes it plausible to further investigate the possibility of a loanword. The Hungarian and Mansi cognates show a regular correspondence in the vocalism of the first syllable, and the words could probably formally could go back to PUg *külVčV*. The relationship between the Hungarian sibilant *s* and North Mansi *s* cannot point to Proto-Ugric *ć*, as this would not be reflected as *s* in North Mansi, whereas *č* regularly yields Proto-Mansi *š* > North Mansi *s* (Honti 1999: 43–44, 124). The UEW reconstructs Proto-Ugric *ć*, but in Uralic vocabulary we find no convincing examples of Hungarian *s* ~

North Mansi *s* reflecting Proto-Uralic *ć*. Interestingly, this correspondence is found in some Ugric etymologies, which often feature other kinds of irregularities as well (see Hungarian *les* below).

Within Mansi, the word is attested only in North Mansi. This is a rare situation also within the Ugric vocabulary, as most etymologies have a wider distribution within the Mansi varieties. It is difficult to determine whether the limited distribution is due to lexical loss or perhaps reflects a situation where a loanword was borrowed only into certain early varieties of Proto-Mansi and into Proto-Hungarian.

To sum up, the obscure derivational suffix and the irregular sibilant correspondence make the idea of a Proto-Ugric word dubious, and it is possible that the alleged Hungarian and Mansi cognates are loanwords from some unattested source.

Hungarian *les* 'lauern, nachstellen, nachspähen', *les* 'Hinterhalt, Lauer, Anstand'

? Khanty: North *lāśi* 'Anstand', *lāśi-* 'lauern' < ? Proto-Khanty *lăći-* (possibly borrowed from Mansi, see below)

Mansi: South (TJ) *läć-*, East (KU) *lɔ̄́ś-*, West (P) *lāś-*, North (So) *lāś-* < Proto-Mansi *lā̇ći-*

< ? PUg *lä̇ćV

(MSzFE: 402; UEW: 863 *läćɜ- 'hiding place; lurk')

As a verb related to hunting, this is also a possible northern substrate word; the irregular consonantism shows that the words in the Ugric languages cannot be real cognates. Hungarian *s* does not correspond regularly to *ć* that can be reconstructed for Khanty and Mansi. The predecessor of Khanty and Mansi *ć* is reconstructed as Proto-Ob-Ugric *ć* by Honti (1999: 124), but this phoneme does not have a regular source in Proto-Uralic. The vocalism is regular, however, as all the cognates point to Proto-Ugric/Proto-Uralic *ä*, but this does not save the Proto-Ugric etymology. Furthermore, due to the very limited distribution of the Khanty word, it is probably borrowed from Mansi, as also mentioned by the UEW and also by DEWOS (813).

Hungarian *levél* (*levelet*) 'leaf' (< ? Pre-Hungarian *läpä-* or *läwä-*)

Khanty: East (V) *ḷị̆wət*, South (DN) *lipət*, North (O) *lipət* 'Blatt'

Mansi: South (TJ) *laptā*, East (KU) *loptə*, West (P) *luptə*, North (So) *lūpta* 'Blatt'

< ? PUg *lVpV

(MSzFE: 403–404; UEW: 865 PUg *lŏ̆pɜ (lepɜ) 'leaf, sheet; Blatt'; WOT: 1283, 1312 PUg *lepə)

This "Ugric" word shows irregular phonological correspondences, and it is impossible to reconstruct a Proto-Ugric word (as noted already by Honti 1999: 139). Also, the correspondences between the alleged Khanty and Mansi cognates are irregular (this is implied also by Honti 1982: 201, who does not reconstruct a Proto-Ob-Ugric form nor Proto-Khanty or Proto-Mansi forms), so no Proto-Ob-Ugric word can be reconstructed either. Both the Ob-Ugric words and the Hungarian word seem to include obscure derivational elements: it is unclear what the origin and function of the *l* suffix in Hungarian and the *tV* suffixes (?) of the Khanty and Mansi words are. Furthermore, the regular reflex of Uralic *pt* in Mansi is *t*, so also the Proto-Mansi consonant cluster *pt* is irregular.

The words are probably connected somehow, but no Proto-Ugric form can be reconstructed. It is also possible that the words are just accidental lookalikes, despite their similar meanings.

Hungarian *nyű* 'worm, maggot',

Khanty: East (V) *niŋk*, South (DN) *ńiŋk*, North (O) *niŋk* 'worm, maggot' < Proto-Khanty (?) *ńiŋk* (Honti 1982: 172, No. 467)

Mansi: East (KU) *ńiχ̢*, North (So) *ńiŋ̥k̥* 'worm' < Proto-Mansi (?) *ńiŋ̥k̥* (Honti 1982: 172, No. 467)

< ? PUg *ńiŋV*

(UEW: PUg (PU?) *ńiŋɜ* (*ńiwŋe*) 'worm, maggot; Made, Wurm')

The possible extra-Ugric cognates mentioned by the UEW are irregular, and it is quite clear that the word cannot be reconstructed for Proto-Uralic: neither North Saami *njivdnja* 'nits, egg of lice' nor Selkup (Taz) *njénje* 'angling worm, bait', (Ty) *ńēń* 'worm', *njeiju* 'angling worm, bait', Kamass *nejme* 'worm' regularly correspond to the Ugric words (this has also been noted by Bakró-Nagy 2003: 34, footnote 5). The suggested Samoyedic cognates are also irregular: the vocalism of the Selkup words listed by the UEW does not correspond regularly to Proto-Uralic *i* or *ü*; see Sammallahti 1988: 495).

Also, the reconstruction of a Proto-Ugric word is impossible, as the alleged cognates show irregular phonology. Khanty and Mansi *i* does not regularly point to Proto-Uralic or Proto-Ugric *i*. It is uncertain what vowel Hungarian *ű* reflects here, as the vowel development was probably influenced by the contraction caused by the loss of the word-internal consonant. The loss of *ŋ* would be irregular in Hungarian, but Khanty and Mansi point to Proto-Ugric *ŋk*.

The irregularities found within Ugric are, again, typical for loanwords, and similar issues can be found among many other Ugric cognates as well.

Hungarian *szalu* (dial.) 'ax for scooping'

Khanty: East (V) *suɣəl*, South (DN) *soχət* 'hatchet' < ? Proto-Khanty *sōɣəl*

Mansi: North (So) *sowli* 'hatchet with a turnable handle' < ? Proto-Mansi *saɣlV* ~ *sɔɣlV*

< ? PUg *sVlkV*

(MSzFE: 567–568; UEW: 889 PUg *sɤlkɜ (sɤlɣɜ)* 'ax for scooping; Hohlbeil')

This technical term shows irregular vocalism, and the word also semantically fits the criteria of loanwords. Honti (2013: 121) supports the traditional etymology, but in reality, no Proto-Ugric form can be reconstructed. The alleged cognates also show phonological irregularity, and the UEW does not determine the quality of the vowel but simply reconstructs *ɤ. Hungarian *a* does not regularly fit the Ob-Ugric vowels (Khanty V *u*, DN *o*, Mansi So *o*).

Hungarian *loll*

Khanty: East (V) *ḷaɣəḷ* 'die flache Hand', South (DT) *laχəl* 'Pfote; Tatze, Handteller (des Menschen)' < Proto-Khanty *ḷāɣəḷ*

Mansi: South (TJ) *lɛjəl*, East (KU) *lāl* (*lajl-*), West (P) *lēl* (*lajl-*), North (So) *lāɣəl* 'Fuß' < Proto-Mansi *līɣəl* (*lēɣəl* in Zhivlov's 2023 system)

< ? PUg *lVlkV*

(MszFE: 407–408; UEW: 865 PUg ? *lɤlkɜ 'some kind of body part (leg, hand, paw); irgendein Glied (Fuß, Hand, Pfote, Tatze)'
    This word shows irregularities similar to other words discussed above, and no Proto-Ugric form can be reconstructed. Honti (1982: 160) does not reconstruct a Proto-Ob-Ugric form, apparently because of the irregularities. Proto-Ugric PUg ? *lɤlkɜ has been also considered a possible reflex of Proto-Uralic *jalka 'foot' (UEW), but this is phonologically impossible, as has been noted by Aikio (UED manuscript). Aikio also notes that the retroflex *ḷ in Khanty does not correspond regularly to *l in Hungarian and Mansi.
    Due to this irregularity, it is possible that this is another substrate word. A word for a body part is not the most typical substrate word, but if the word originally denoted the body part of an animal, then borrowing from a substrate source is not that strange.

Hungarian *tűz* 'fire'

Khanty: East (V) *tŏγət*, South (DN) *tüt*, North (O) *tut* 'fire' < ? Proto-Khanty *tüγət*

Mansi: South (TJ) *täwə·t*, East (KU) *tɔ̄wt*, West (P) *tāwt* 'fire' < Proto-Mansi *tāwt*

< ? PUg *tüγVtV* 'fire'

(MSzFE 648–649; UEW: 896–897 PUg *tüγɜ-tɜ*, *tüwɜ-tɜ*- 'fire; Feuer', Abondolo 1996: 62: PUg *tiki-tV*)

The word for 'fire' does not belong to a semantic field typical for substrate words, but the Ugric comparanda show irregular correspondences that make this word also a typical candidate for a loanword. Salminen (2002) has cautiously noted that the Ugric word for 'fire' might be a true Proto-Ugric lexical innovation, but due to the irregular vocalism, this does not seem to be the case. As a word for 'fire' can be securely reconstructed for Proto-Uralic (*tuli*), it is also quite clear that the Ugric words have to be some kind of lexical innovations; as they clearly are not derivatives of any known Uralic stem, a loanword is the most plausible option. Even though the word is not semantically a typical substrate word, we could still assume that it was borrowed from the same source as the substrate words discussed above. For example, Kallio (in press) argues that the Baltic loanwords in Finnic are substrate words, and this group of words likewise includes (in addition to typical substrate words) individual words that do not fulfill all the criteria for a substrate.

Hungarian *üsző*, dial. *isző* 'heifer, calf; female deer, hind' (< ? Proto-Hungarian *üsäγV*)

Khanty: East (V) *ĕs* 'mother', North (O) *ăs* 'mother, female (animal)' < Proto-Khanty *is*

< ? PUg *isV* ?

(MszFE: 661–662; UEW: 848 PUg *iśɜ (eśɜ)* 'Mutter, weibliches Tier; mother, female animal')

Aikio (UED draft) considers this etymology possible, if the Hungarian *i* is original and *ü* secondary. However, it is dubious whether this is the case. The vowel correspondence between Hungarian *ü* and Khanty *ĕ*, *ă* would be irregular. In any case, the semantic connection is rather vague, as the word in Khanty dialects means simply 'mother' (DEWOS 186–187). DEWOS (186–187) does not mention the possible

etymological connection of the Khanty word with Hungarian *üsző* at all, but the Ugric etymology of the UEW is supported by WOT.

A loan from some reflex of Iranian *wasá-* 'calf' (< Proto-Indo-Iranian *watsá-* > Ossetic *wæs*; der. *watsa-ka-* > Khotanese Saka *basaka*; EWAIa II: 492) has been suggested (Munkácsi 1901; Sköld 1925: 37–38). This has been refuted by Korenchy (1972), and also Joki is critical of the etymology, but Helimski (2002: 109–110) supports the etymology, also noting briefly that the Ugric etymology of the UEW is problematic. WOT lists the etymology among the unacceptable Iranian etymologies but does not comment on it.

The loan etymology seems promising, but it requires some additional arguments. Hungarian *ü* from Iranian *wa-* lacks exact parallels. According to Helimski, after a labial consonant the *ü* vowel would be expected, as Hungarian *füst* 'smoke' (from Iranian *pazda-* 'smoke') shows a similar development (see above). It is also not clear why *w-* would be lost in Hungarian. There seem to be no certain examples of earlier *wü-* in Hungarian: UEW reconstructs the Uralic word for 'belt' as *wiŋä* or *wüŋä* (> Hungarian *öv*), but Aikio (2012: 230) argues that the correct reconstruction is *üwä*, so this is not a plausible parallel. *w* is lost before *u* (PU *wud'ə* > Hungarian *új* 'new') and also before *i*: the sequence *wi* gives Hungarian *ö* regularly (PU *wilä-* 'to kill' > Hungarian *öl*, Sammallahti 1988: 551). In the light of these developments, assuming that *ü* in *üsző* reflects *wü-* is not an implausible idea, but parallel examples would be needed to claim that this is a regular development.

If the Iranian etymology is correct, the long -*ő* could be explained from the Iranian suffix -*ka* (in the derivation *wasaka-* < *watsaka-*) that had first become *γ* and then vocalized; this would be regular in Hungarian.

4. Conclusions and perspectives for further research

The etymological commentary above shows quite clearly that several of the allegedly Proto-Ugric etymologies presented in the earlier etymological dictionaries are irregular and cannot be reconstructed for the Ugric proto-language. The earlier views, reflected for example by Honti (1997, 2013) that justify the Ugric unity with shared lexicon, are thus clearly unjustified, and the "Ugric" vocabulary as such tells very little of the possible closer genealogical relationship of the Ugric languages. Of course a full account of the allegedly Proto-Ugric lexicon needs to be presented before the lexical evidence for Proto-Ugric can be fully rejected, but the remarks above together with the Ugric etymologies refuted in recent research make it clear that a significant part of the Ugric lexicon consists of loans, substrate words, and erroneous etymologies.

The present paper is a case study, and a full scrutiny of the Proto-Ugric vocabulary would require also an analysis of the remaining etymologies in the future. There are some words confined to the Ugric languages that do display the regular sound correspondences found in inherited Uralic words: at least Proto-Ugric *palka*

(reconstruction by Zhivlov 2014: 120; UEW: *palɣɜ) Hungarian *falu*, Khanty (East) *puɣəl*, Mansi (East) *pāwəl* 'village', and Proto-Ugric *kapVtV-* (Aikio 2015b: 57) > Hungarian *húz* 'pull, drag', Khanty (East) *kŏpi̯tə-* 'zerstückeln, abstampfen, losmachen', Mansi (South) *kat-* 'reißen' are examples of such etymologies, and it is possible that more examples of such convincing cases can be found. However, it is debatable how much weight such individual etymologies have in the discussions of the genealogical classification of Hungarian, Khanty, and Mansi. It is known that shared vocabulary items are, in general, a poor criterion in genealogical taxonomy of languages (see Campbell & Poser 2008: 165–167), and if one can present only a handful of regular cognates confined within a set of related languages, it is dubious whether such words can reliably be considered shared innovations, especially as it cannot be shown that such words replaced an inherited word that could be reconstructed for an earlier proto-language.

It should also be kept in mind that not all loanwords shared by related languages show irregular phonological correspondences. For example, there are various Indo-European loanwords shared by Saami and Finnic that display regular correspondences in both consonantism and vocalism but which were most likely borrowed separately. The process known as etymological nativization (described by Aikio 2007, 2009) has also affected the relationship of Finnic and Saami vocabulary, meaning that later loans between Finnic and Saami often display vowel correspondences typical of those found in inherited words (such as Finnish *hinta* ~ North Saami *haddi* 'price'). The Ugric vocabulary has not been studied from this point of view, but in the future when the Ugric phonological developments are more carefully worked out, analyzing possible etymological nativization in the Ugric vocabulary would be an important task.

Abbreviations

PU   Proto-Uralic
PUg   Proto-Ugric

Khanty varieties:

DN	Upper Demyanka dialect (South Khanty)	Kr	Krasnoyarsk dialect (South Khanty)
		Ni	Nizyam dialect (South Khanty)
DT	Lower Demyanka dialect (South Khanty)	O	Obdorsk dialect (North Khanty)
		Trj.	Tremjugan dialect (East Khanty)
J	Jugan dialect (East Khanty)	V	Vakh dialect (East Khanty)
Ko	Konda dialect (South Khanty)	Vj	Vasjugan dialect (East Khanty)

Mansi varieties:

K	Konda dialect (East Mansi)	N	North Mansi
KM	Middle Konda dialect (East Mansi)	P	Pelym dialect (West Mansi)
KO	Upper Konda dialect (East Mansi)	So	Sosva dialect (North Mansi)
KU	Lower Konda dialect (East Mansi)	T	Tavda dialect (South Mansi)
LM	Middle Lozva dialect (West Mansi)	TJ	Tavda dialect, Janychkova village (South Mansi)
LO	Upper Lozva dialect (West Mansi)		

References

Abaffy, Erzsébet. 2003. Ómagyar kor. Hangtörténet. In Kiss, Jenő & Pusztai, Ferenc (eds.), *Magyar nyelvtörténet*, 301–351. Budapest: Osiris 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.

Abondolo, Daniel. 1998. Introduction. In Abondolo, Daniel (ed.), *The Uralic* languages, 1–42. London – New York: Routledge.

Aikio, Ante (Luobbal Sámmol Sámmol Ánte). 2004. An essay on substrate studies and the origin of Saami. In Hyvärinen, Irma & Kallio, Petri & Korhonen, Jarmo (eds.), *Etymologie, Entlehnungen und Entwicklungen: Festschrift für Jorma Koivulehto zum 70. Geburtstag* (Mémoires de la Société Néophilologique de Helsinki 63), 5–34. Helsinki: Société Néophilologique.

Aikio, Ante (Luobbal Sámmol Sámmol Ánte). 2007. Etymological nativization of loanwords: A case study of Saami and Finnish. In Toivonen, Ida & Nelson, Diane (eds.), *Saami linguistics*, 17–52. Amsterdam – Philadelphia: John Benjamins.

Aikio, Ante (Luobbal Sámmol Sámmol Ánte). 2009. *Saami loanwords in Finnish and Karelian*. Oulu. (Doctoral dissertation.)

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). 2014. The Uralic-Yukaghiric lexical correspondences: genetic inheritance, contact of chance resemblance? *Finnisch-Ugrische Forschungen* 62. 7–76.

Aikio, Ante (Luobbal Sámmol Sámmol Ánte). 2015a. Studies in Uralic etymology IV: Ob-Ugric etymologies. *Linguistica Uralica* 51. 1–20.

Aikio, Ante (Luobbal Sámmol Sámmol Ánte). 2015b. 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.), *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*, 77–90. Oxford: Oxford University Press.

AiWb = Bartholomae, Christian. 1961. *Altiranisches Wörterbuch*. 2., unveränderte Auflage. Berlin: De Gruyter.

Bailey, Harold. 1979. *Dictionary of Khotan Saka*. Cambridge: Cambridge University Press.

Bakró-Nagy, Marianne. 2003. Egy inetimologikus hangról. In Bakró-Nagy, Marianne & Rédei, Károly (eds.), Ünnepi kötet Honti László tiszteletére, 27–45. Budapest: A Magyar Tudományos Akadémia, Nyelvtudományi Intézet.

Bakró-Nagy, Marianne. 2013. Mit tudunk az ugor történeti fonológiáról? In Agyagási Klára, Hegedűs Attila & É. Kiss Katalin (eds.), *Nyelvelmélet és kontaktológia II*. Piliscsaba: PPKE BTK.

Beekes, Robert Stephen Paul. 1996. Ancient European loanwords. *Historische Sprachforschung* 109. 215–236.

Campbell, Lyle & Poser, William J. 2008. *Language classification: History and method*. Cambridge: Cambridge University Press.

Campbell, Lyle. 1999. *Historical linguistics: An introduction*. Cambridge, Massachusetts: The MTI Press.

Cheung, Johnny. 2002. *Studies in the historical development of the Ossetic vocalism* (Beiträge zur Iranistik, Band 21). Wiesbaden: Dr. Ludwig Reichert Verlag.

Cheung, Johnny. 2007. *Etymological dictionary of the Iranian verb* (Leiden Indo-European Etymological Dictionary Series 2). Leiden – Boston: Brill.

Clackson, James. 1994. *The linguistic relationship between Armenian and Greek*. Oxford: Blackwell.

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

Clackson, James. 2022. Methodology in linguistic subgrouping. In Thomas Olander (ed.), *The Indo-European language family: A phylogenetic perspective*, 18–32. Cambridge: Cambridge University Press.

DEWOS = Steinitz, Wolfgang. 1966–1993. *Dialektologisches und etymologisches Wörterbuch der ostjakischen Sprache*. Berlin: Akademie Verlag.

EWAia = Mayrhofer, Manfred. 1986–2001. *Etymologisches Wörterbuch des Altindoarischen* I–III. Heidelberg: Universitätsverlag C. Winter.

EWUng = Benkő, Loránd. 1992–1997. *Etymologisches Wörterbuch des Ungarischen Sprache*. Budapest: Akadémiai Kiadó.

Fox, Anthony. 1995. *Linguistic reconstruction: An introduction to theory and method*. Oxford – New York: Oxford University Press.

Futáky, István. 1975. *Tungusische Lehnwörter des Ostjakischen* (Veröffentlichungen der Societas Uralo-Altaica 10). Wiesbaden: Harrassowitz.

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

Gulya, János. 1994. Loch im Stammbaum? (Zur Urheimatfrage der Ugrier und der Ungarn). In *Die Vorgeschichte der uralischen Völker: Materialien eines Internationalen Symposions 14.–16. Oktober 1993 in Szombathely* (Specimina Sibirica 10), 45–62. Szombathely: Savariae.

Hajdú, Péter. 1962. *Finnugor népek és nyelvek*. Budapest: Gondolat.

Häkkinen, Jaakko. 2007. *Kantauralin murteutuminen vokaalivastaavuuksien valossa*. Helsinki: University of Helsinki. (Unpublished MA thesis.) <http://urn.fi/URN:NBN:fi-fe20071746> (Accessed 2019-05-24.)

Häkkinen, Kaisa. 1983. *Suomen kielen vanhimmasta sanastosta ja sen tutkimisesta* (Turun yliopiston suomalaisen ja yleisen kielitieteen laitoksen julkaisuja 17). Turku: Turun yliopisto.

Harmatta, János. 1997. Iráni nyelvek hatása az ősmagyar nyelvre. In Kovács, László & Veszprémi, László (eds.), *Honfoglalás és nyelvészet*, 71–83. Budapest: Balassi.

Helimski, Eugene. 1982. = Хелимский, Е. А. 1982. *Древнейшие венгерско-самодийские языковые параллели*. Москва: Наука.

Helimski, Eugene. 1985. [Рецензия на] Honti, 1982. *Советское финно-угроведение* 21. 63–73.

Helimski, Eugene. 1997. The southern neighbours of Finno-Ugrians: Iranians or an extinct branch of Aryans ("Andronovo Aryans")? In Hahmo, Sirkka-Liisa & Hofstra, Tette & Honti, László & Linde, Paul Van & Nikkilä, Osmo (eds.), *Finnisch-ugrische Sprachen in Kontakt: Vorträge des Symposiums aus Anlaß des 30-jährigen Bestehens der Finnougristik an der Rijksuniversitejt Groningen 21.–23. November 1996*, 117–125. Maastricht: Shaker.

Helimski, Eugene. 2002. Zu den iranischen Lehnwörtern im Ungarischen. Rogier Blokland & Cornelius Hasselblatt (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). Maastricht: Shaker.

Helimski, Eugene. 2003. Areal groupings (Sprachbunde) within and across the borders of the Uralic language family: a survey. *Nyelvtudományi Közlemények* 100. 156–167.

Holopainen, Sampsa & Milanova, Veronika & Bradley, Jeremy & Czentnár, András. 2019. Mari kinship terms from the typological, genealogical, and areal point of view. (A poster presented at the SLE annual meeting, 21–24 September 2019, University of Leipzig.) <https://mari-language.univie.ac.at/misc/kinship/sle_2019_poster.pdf>

Holopainen, Sampsa & Czentnár, András. 2022. Speculative Iranian loanwords in Hungarian and Turkic. (Presentation at the Deutscher Orientalistentag, Freie Universtität Berlin, 16.09.2022.)   <https://ugric.univie.ac.at/fileadmin/user_upload/p_ugric/Berlin_DOT_Holopainen_Czentnar_updated.pdf> (Accessed 2022-09-26.)

Holopainen, Sampsa. 2019. *Indo-Iranian borrowings in Uralic: Critical overview of sound-substitutions and distribution criterion*. (Doctoral dissertation.) Helsinki: University of Helsinki.

Holopainen, Sampsa. 2022a. *Ugric vocabulary* (Appendix to the paper Grünthal, R., Heyd, V., Holopainen, S., Janhunen, J., Khanina, O., Miestamo, M., Nichols, J., Saarikivi, J. & Sinnemäki, K.. 2022. Drastic demographic events triggered the Uralic spread.   *Diachronica*.   <https://doi.org/10.1075/dia.20038.gru>)   <https://zenodo.org/record/6035881#.Y3KG7nbMKUk> (Accessed 2022-11-14.)

Holopainen, Sampsa. 2022b. Revisiting a problematic Uralic and Indo-Iranian word family. In Bradley, Jeremy & Csire, Márta & Erlinghagen, Erika & Hirvonen, Johannes & Holopainen, Sampsa & Kajander, Mikko & Lindorfer, Heidelinde & Pesti, Brigitta & Pischlöger, Christian & Seidler, Andrea & Viilukas, Triinu & Vincze, Ferenc (eds.), *Tonavan Laakso: Eine Festschrift für Johanna Laakso* (Central European Uralic Studies 2), 198–213. Wien: Praesens Verlag.

Holopainen, Sampsa. 2022c. Uralilaisen lingvistisen paleontologian ongelmia – mitä sanasto voi kertoa kulttuurista? In Kaheinen, Kaisla & Leisiö, Larisa & Erkkilä, Riku & Qiu, Toivo E. H. (eds.), *Hämeenmaalta Jamalille. Kirja Tapani Salmiselle 07.04.2022*, 101–114. Helsinki: University of Helsinki.

Holopainen, Sampsa & Junttila, Santeri. 2022. *Die alten arischen und baltischen Lehnverben der uralischen Sprachen* (Münchener Studien zur Sprachwissenschaft, Beiheft 33). München: J.H. Röll Verlag.

Honti, László. 1982. *Geschichte des obugrischen Vokalismus der ersten Silbe*. Budapest: Akademiai Kiadó.

Honti, László. 1997. *Az ugor alapnyelv kérdéséhez* (Budapesti finnugor füzetek 7). Budapest: ELTE – Numi-Tórem Alapítvány.

Honti, László. 1998. ObUgrian. In Daniel Abondolo (ed.), *The Uralic languages*, 327–357. London – New York: Routledge.

Honti, László. 1999. *Az obiugor konszonantizmus története* (Studia uralo-altaica supplementa 9). Szeged: Szegedi Tudományegyetem.

Honti, László. 2013. *Magyar nyelvtörténeti tanulmányok*. Budapest: L'Harmattan.

Honti, László. 2017. *A magyar és a nyugati ótörök szókészleti kapcsolatairól*. Budapest: Tinta Könyvkiadó.

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

Janhunen, Juha. 1981. Uralilaisen kantakielen sanastosta. *Journal de la Société Finno-Ougrienne* 77. 219–274.

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.

Joki, Aulis J. 1973. *Uralier und Indogermaner* (Mémoires de la Société Finno-Ougrienne 151). Helsinki: Société Finno-Ougrienne.

Junttila, Santeri & Holopainen, Sampsa. 2022. On the integration of Indo-European loanverbs in Finnic. In Sommer, Florian & Stüber, Karin & Widmer, Paul & Yamazaki, Yoko (eds.), *Indogermanische Morphologie in erweiterter Sicht* (Innsbrucker Beiträge zur Sprachwissenschaft 167), 291–324. Innsbruck: Universität Innsbruck.

Kallio, Petri. 2004. Tocharian loanwords in Samoyed? In Hyvärinen, Irma & Kallio, Petri & Korhonen, Jarmo (eds.), *Etymologie, Entlehnungen und Entwicklungen: Festschrift für Jorma Koivulehto zum 70. Geburtstag*, 129–137. Helsinki: Société Néophilologique.

Kallio, Petri. (in press). Substrates in Finnic. In Aikio, Ante & Palviainen, Santeri (eds.), *Substrate languages in Northern Europe: Case studies and methodological perspectives* (Studies in Language Change 17). Berlin – New York: De Gruyter.

Kálmán, Béla. 1988. The history of the Ob-Ugric languages. In Sinor, Denis (ed.), The Uralic languages: Description, history and foreign influences (Handbuch der Orientalistik 8, Handbook of Uralic studies 1), 395–412. Leiden: E. J. Brill.

Károly, László. 2001. On the Old Turkic verb *ir*- 'to make a notch, or breach' and its derivatives. In Károly, László & Kincses Nagy, Éva (eds.), *Néptörténet–Nyelvtörténet: A 70 éves Róna-Tas András köszöntése*. Szeged: SZTE BTK Altajisztikai Tanszék.

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: C. Winter.

Kispál, Magdolna Sz. 1951. Ugor-török érintkezés. *Nyelvtudományi közlemények* 53. 49–64.

Koivulehto, Jorma. 1999. Varhaiset indoeurooppalaiskontaktit: aika ja paikka lainasanojen valossa. In Paul Fogelberg (ed.), *Pohjan poluilla: Suomalaisten juuret nykytutkimuksen mukaan* (Bidrag till kännedom av Finlands natur och folk 153), 207–236. Helsinki: Suomen Tiedeseura.

Koivulehto, Jorma. 2007. Saamen ja suomen 'poro'. In Ylikoski, Jussi & Aikio, Ante (eds.), *Sámit, sánit, sátnehámit: Riepmočála Pekka Sammallahtii miessemánu 21. beaivve 2007* (Mémoires de la Société Finno-Ougrienne 253), 251–258. Helsinki: Société Finno-Ougrienne.

Korenchy, Éva. 1972. *Iranische Lehnwörter in den obugrischen Sprachen*. Budapest: Akadémiai Kiadó.

Kuiper, Franciscus Bernardus Jacobus. 1991. *Aryans in the Rigveda* (Leiden studies in Indo-European 1). Amsterdam – Atlanta: Brill.

Kuiper, Franciscus Bernardus Jacobus. 1995. Gothic *bagms* and Old Icelandic *ylgr*. *NOWELE* 25. 63–88.

Kümmel, Martin. 2017. Etymological problems between Indo-Iranian and Uralic. (Presentation held 17 May 2017 in Helsinki.)

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 164), 237–277. Innsbruck: Institut für Sprachwissenschaft der Universität Innsbruck.

Ligeti, Lajos. 1986. *A magyar nyelv török kapcsolatai a honfoglalás előtti és az Árpád-korban*. Budapest: Akadémiai Kiadó.

Lubotsky, Alexander. 2001. The Indo-Iranian substratum. 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), 301–317. Helsinki: Société Finno-Ougrienne.

Lushnikova 1990 = Лушникова, Алла В. 1990. *Стратификация Ирано-Уральских языковых контактов*. Moscow. (Unpublished dissertation.)

Mayrhofer, Manfred. 1989. Vorgeschichte der iranischen Sprachen; Uriranisch. In Schmitt, Rüdiger (ed.), *Compendium linguarum iranicarum*, 4–24. Wiesbaden: Dr. Ludwig Reichert Verlag.

MSzFE = Lakó, György (ed.). 1971–1978. *A magyar szókészlet finnugor elemei*. Budapest: Akadémiai Kiadó.

Munkácsi, Bernát. 1901. Árja és kaukázusi elemek a finn-magyar nyelvekben. Budapest: Magyar Tudományos Akadémia.

Munkácsi, Bernát. 1904. Alanische Denkmäler im ungarischen Wortschatze. *Keleti szemle* 5. 304–329.

Olander, Thomas. 2022. Introduction. – Thomas Olander (ed.), *The Indo-European language family: A phylogenetic perspective*, 1–17. Cambridge: Cambridge University Press.

Parpola, Asko. 2012. Formation of the Indo-European and Uralic (Finno-Ugric) language families in the light of archaeology. 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.

Parpola, Asko. 2020. Royal "chariot" burials of Sanauli near Delhi and archaeological correlates of prehistoric Indo-Iranian languages. *Studia Orientalia Electronica* 8. 75–198.

Ponaryadov, Vadim. 2022. Хотаносакско-пермские лексические связи. *Индоевропейское языкознание и классическая филология XXIV*.

Pystynen, Juho. 2013. The case of Mansi *ś > *š, part 1. (Blog post.) <https://protouralic.wordpress.com/2013/09/06/the-case-of-mansi-s-%E2%86%92-s-part-1/> (Accessed 2023-06-22)

Pystynen, Juho. 2014. Primary vs. secondary *ë. (Blog post.) <https://protouralic.wordpress.com/2014/09/05/primary-vs-secondary-e/#%C3%AB1%C3%AB2-1> (Accessed 2022-11-14.)

Pystynen, Juho. 2017. The state of the Finno-Permic lexicon. (Slides of the presentation at the conference Contextualizing Historical Lexicology, University of Helsinki, 15 May 2017.) <https://blogs.helsinki.fi/hist-lexicology/files/2017/06/chl-pystynen.pdf>

Rasmussen, Jens Elmegård. 1989. *Studien zur Morphophonemik der indogermanischen Grundsprache* (Innsbrucker Beiträge zur Sprachwissenschaft 55). Innsbruck: Universität Innsbruck.

Rédei, Károly. 1986. *Zu den indogermanisch-uralischen Sprachkontakten* (Veröffentlichungen der Kommission für Linguistik und Kommunikationsforschung, Heft 16). Wien: Verlag der Österreichischen Akademie der Wissenschaften.

Riese, Timothy. 2001. *Historische Nominalderivation des Wogulischen* (Studia Uralica 10). Wiesbaden: Harrassowitz.

Róna-Tas, András. 1988. Turkic influence on the Uralic languages. In Sinor, Denis (ed.), The Uralic languages: Description, history and foreign influences (Handbuch der Orientalistik 8, Handbook of Uralic studies 1), 742–780. Leiden: E. J. Brill.

Róna-Tas, András. 1992. A magyar írásbeliség török eredetéhez (ír és *betű* szavaink etimológiája). In Sándor, Klára (ed.), *Rovásírás a Kárpát-medencében* (Magyar őstörténeti könyvtár 4), 9–14. Szeged: SZTE BTK Altajisztikai Tanszék.

Róna-Tas, András. 2017. Vitás magyar etimológiák. Válasz Honti Lászlónak. *Nyelvtudományi Közlemények* 113. 37–84.

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: Oxford University Press.

Salminen, Tapani. 2001. The rise of the Finno-Ugric language family. 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), 385–396. Helsinki: Société Finno-Ougrienne.

Salminen, Tapani. 2002. Problems in the taxonomy of the Uralic languages in the light of modern comparative studies. In Агранат, Т. Б. & Казакевич, О. А. (eds.), *Лингвистический беспредел: сборник статей к 70-летию А. И. Кузнецовой*, 44–55. Москва: Издательство Московского университета.

Salmons, Joe. 1992. Northwest Indo-European vocabulary and substrate phonology. In Pearson, Roger (ed.), *Perspectives on Indo-European language, culture and religion: Studies in honor of Edgar C. Polomé*, *Volume 2* (Journal of Indo-European Studies monograph series 9), 265–279. Washington, DC: Institute for the Study of Man.

Salmons, Joe. 2015. Language shift and the Indo-Europeanization of Europe. In Olsen, Birgit Annette & Mailhammer, Robert & Vennemann, Theo (eds.), *Die sprachlichen Wurzeln Europas – Linguistic roots of Europe: Ursprung und Entwicklung – Origin and development*, 147–169. Copenhagen: Tusculaneum Press.

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.

Schrijver, Peter. 1997. Animal, vegetable and mineral: Some Western European substratum words. In Lubotsky, Alexander (ed.), *Sound law and analogy: Papers in honor of Robert S.P. Beekes on the occasion of his 60th birthday* (Leiden studies in Indo-European 9), 293–316. Amsterdam – Atlanta: Brill.

Sinor, Denis. 1977. Altaica and Uralica. In Sinor, Denis (ed.), *Studies in Finno-Ugric linguistics in honor of Alo Raun* (Indiana University Uralic and Altaic series 131), 319–331. Bloomington: Indiana University.

Sipos, Mária. 2002. Az obi-ugor alapnyelv lexikális innovációi I. *Nyelvtudományi Közlemények* 99. 7–56.

Sipos, Mária. 2003. Az obi-ugor alapnyelv lexikális innovációi II. *Nyelvtudományi Közlemények* 100. 245–263.

Sköld, Hannes. 1925. *Die ossetischen Lehnwörter in Ungarischen*. Leipzig: Otto Harrassowitz.

Tálos, Endre P. 1983. Kép szöveg nélkül. In Bereczki, Gábor & Domokos, Péter (eds.), *Urálisztikai tanulmányok (Hajdú Péter 60. születésnapja tiszteletére)*, 409–420. Budapest: ELTE Finnugor tanszék.

Tálos, Endre P. 1984. Vogul + osztjak / 2. *Nyelvtudományi Közlemenyek* 86(1). 89–99.

Tálos, Endre P. 1987. On the vowels of Proto-Uralic. In Rédei, Károly (ed.), *Studien zur Phonologie und Morphonologie der Uralischen Sprachen* (Studia Uralica 4), 70–80. Wien: Verband der wissenschaftlichen Gesellschaften Österreichs.

UED draft = Aikio, Ante (manuscript): *Uralic etymological dictionary.* (Unpublished manuscript, draft version of entries A–Ć published at Academia.edu, last accessed 2020-10-24.)

TESz = Benkő Loránd (ed.). 1967–1976. *A magyar nyelv történeti-etimológiai szótára.* I–III. Szerkesztők Lajos Kiss, László Kubinyi & László Papp. Budapest: Akadémiai Kiadó.

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.

Wojtilla, Gyula. 2002. Óind *áṣṭrā-* f. 'ösztöke' vagy/és 'ostor'? In Bibok, Károly & Ferincz, István & Kocsis, Mihály (eds.), *Cirill és Metód példáját követve… Tanulmányok H. Tóth Imre 70. születésnapjára*, 585–590. Szeged: Szegedi Tudományegyetem Szláv Intézete.

WOT = Róna-Tas, András & Berta, Árpád (with the assistance of László Károly). 2011. *West Old Turkic: Turkic loanwords in Hungarian* (Turcologica 84). Wiesbaden: Harrassowitz.

Zhivlov, Mikhail. 2006. *Реконструкция праобско-угорского вокализма* [*Reconstruction of Proto-Ob-Ugric vocalism*]. Moscow: Russian State University for the Humanities. (Unpublished dissertation.)

Zhivlov, Mikhail. 2007. К вопросу о реконструкции обско-угорского вокализма. In *Aspects of Comparative Linguistics* 2, 281–309. Moscow: RSUH publishers.

Zhivlov, Mikhail. 2013. Андроновский арийский язык. In Коряков, Ю.Б. & Кибри, А.А. (eds.), *Реликтовые индоевропейские языки Передней и Центральной Азии* (Языки мира), 217–220. Москва: Academia.

Zhivlov, Mikhail. 2014. Studies in Uralic vocalism III. *Journal of Language Relationship* 12. 113–148.

Zhivlov, Mikhail. 2015. Неиндоевропейский субстрат в финно-волжских языках. Х традиционные чтения памяти С. А. Старостина. (Handout of a presentation, 27 March 2015.) <https://www.academia.edu/11689161/Неиндоевропейский_субстрат_в_финно-волжских_языках> (Accessed 2019-05-27.)

Zhivlov, Mikhail. 2016. The origin of Khanty retroflex nasal. *Journal of Language Relationship* 14(4). 293–302.

Zhivlov, Mikhail. 2018. Историческая фонетика и внутренняя классификация уральских языков. (Presentation, XIII традиционные чтения памяти С. А. Старостина, Институт восточных культур и античности РГГУ, 23.3.2018.)

Zhivlov, Mikhail. 2023. Reconstruction of Proto-Uralic. In Abondolo, Daniel & Valijärvi, Riitta-Liisa (eds.), *The Uralic languages*. 2nd edition, 117–175. London – New York: Routledge.

Zimmer, Stefan. 1990. *Urvolk, Ursprache und Indogermanisierung: Zur Methode der Indogermanischen Altertumskunde*. Innsbruck: Institut für Sprachwissenschaft der Universität Innsbruck.