

Uralic **tuppas* – bridging Indic and Germanic¹

Tiivistelmä

Uralilainen **tuppas* indoarjan ja germaanin silloittajana

Suomen *tupas* johdoksineen ja näiden vastineet pohjoisessa itämerensuomessa voidaan parhaiten johtaa useimpien suomen murteiden sekä vepsän osoittamasta kantasuomen asusta **tuppas*. Muutamat sanansisäiseen yksittäisklusiiliin viittaavat vastineet kuten karjalan *tuvas* katson selittyvän *ryväs*-sanueen sekundaarisiksi vaikutukseksi. Tunkelon (1918) germaaninen lainaetymologia ← kantagermaanin **hūbaz* 'mätäs, pensas' ei tämän valossa ole puolustettavissa. Koivulehdon (1999) rinnastus toiseen kantagermaanin sanueeseen **tuppaz* 'tupsu, tukko, yläpää' näyttää paremmalta, mutta tältä sanueelta taas puuttuu oma indoeurooppalainen etymologia. Kyseessä onkin pikemmin laina itämerensuomesta luoteisgermaaniin, sillä itämerensuomen sanueelle voidaan löytää vastineita myös idempää, marista: deminutiivijohdos **tūpka* 'tuppo, tukka', ja komista: aines *tup-* useissa johdoksissa kuten *tupji-* 'tukkia', *tupjirt-* 'kääriä', ja yhdyssanoissa kuten *tup-jura* 'tupsupää'. Poikkeusedustus **u* > komin *u* lienee säännöllinen **p*:n edellä. Sanueen *-*as*-vartalo itämerensuomessa silti viittaa indoeurooppalaiseen laina-alkuperään, ja myös uusi lainaetymologia voidaan esittää: kyse on vanhasta indoiranilaisesta (arjalaisesta) lainasta, jota parhaiten vastaa muinaisindoarjan *stūpa-* 'mätäs, tupas, päälaki, ym.' < kantaindoiranin **stūHpas*. Samaan etymologiseen pesueeseen voitaisiin

1. I thank two anonymous reviewers for their comments which have helped to improve this paper; and, for its initial inspiration, discussion on the etymological connections of *stūpa* ~ *stupa* at the *Languagehat* blog by regular commentators January First-of-May, juha, and David Marjanović, whom I owe for noting its similarity with **tuppaz* (<<https://languagehat.com/stupa/>>). Any remaining errors in the etymological scenario constructed here are my own.

jatko-oletuksin liittää vielä marin **tūp* 'selkä' (mikäli aiemmin 'yläselkä') ja suomen *tukka* (arjalaisen sanueen jostain toisesta variantista, vrt. muinaisindoarjan *stūkā*- 'tupsu, tukko'). Huomatavinta kuitenkin on, että germaanin **tuppaz*-sanue ei sovi indoiranin sanojen perintövastineeksi, ja näiden yhteyden mahdollistaa ainoastaan itämerensuomen kautta välittyminen – samaan tapaan kuin esim. jo ennestään tunnettu *repo*-sanueen lainautuminen indoiranista itämerensuomen kautta pohjoisgermaaniin. Vastaavaa takaisinlainautumisen mahdollisuutta onkin syytä pitää silmällä myös jatkossa tarkasteltaessa itämerensuomalaisien tai yleisemmin uralilaisten kielten sanueita, joille on esitetty useampia lainaetymologioita.

1. Uralic

1.1. Finnic

Proto-Finnic **tuṕpas* : **tuppaha*- 'mound' can be reconstructed from Fi[nnish] *tupas* : gen. sg. *tuppaan* 'mound, clump of vegetation, cluster of trees, reeds, berries, etc.' and Veps *tupaz* 'mound' (SSA s.v. *tupas*). Front-vocalic variants *typäs*, *tyypäs* also appear in Finnish dialects. Other forms suggest an original singleton stop: Southern Karelian *tuvaš*, Olonets Karelian *tuvas*, Northern Ludian *tubaz* (KKS s.v. *tuvas*; LMS s.v. *tubaz*). Nominally, **tuṕpas* : **tupaha* could thus be also reconstructed already for at least Proto-North Finnic. The word has however a very narrow dialect distribution in Ludian: the base noun is attested only from the village of Bošinkülä, a collective derivative *tubahikk* 'a growth of vegetation in tufts' also from Sununsuu, both in the northern Ludian dialect area. This suggests a late loan from Karelian, where *tubahikko* is also attested.

Two yet more narrowly distributed Karelian variants furthermore appear in an Olonets Karelian (Säämäjärvi) collective *tubehikko*, and the Southern Karelian (Suojärvi) verbs *tupehtuo* 'to accumulate (intr.)', *tupehuttoa* 'to accumulate (tr.)', interpretable as 'to have in bundles'. The latter two are homophonic with, but probably to be distinguished from, more widely distributed *tupehtuo*, *tupehuttoa* 'to choke (intr., tr.)'. Besides good distribution in Karelian, these senses

have exact cognates further in Finnish *tupehtua*, Ludian and Veps *tupehtuda* (SSA s.v. *tupehtua*), while ‘to accumulate’ seems to be an innovation limited to dialectal Karelian.

A different semantically specialized reflex probably appears in Finnish *tuppo* ~ dial. *tuppu* ‘clump of hay or fabric, wad, plug, wound dressing’, the first of these likely being the oldest meaning. Morphologically, a diminutive noun in *-o* ~ *-ö* or *-u* ~ *-y* besides a base root in *-as* ~ *-äs* is not common (cf. Hakulinen 1979: 173–175, 184–185), but some clear precedents do exist, e.g. *taivas* ‘heaven’ ⇒ *taivo* id., *töyräs* ‘(river)bank’ ⇒ *töyry* id. As a further possibility, *tuppo* could be furthermore considered as the base of *tupehtua* ‘to choke’ with its further cognates (thus SSA s.v. *tupehtua*), by a semantic development such as ‘to be plugged, stopped’ > ‘to have airways plugged, be choked’. This possibility is however complicated by a high diversity of Finnic word groups meaning ‘to choke, be out of breath’, including already e.g. in Fi. *typehtyä*, *tukehtua*, *tikahtua*, *pakahtua*, *läkähtyä*, all with similar derivation and similar phonotactics. Etymologizing any of them would call for the examination of other hypotheses, too, such as derivation from onomatopoeia for the gasps of a choking person or, specifically for *tupehtua*, also from the illative adverbs Fi. *tuppeen*, Olonets *tuppeh* ‘tightly shut, full’, which likely derive from *tuppi* ‘sheath (of a leaf, knife)’ (SSA s.v. *tuppi*).

Before considering any connections beyond Finnic for the etymological cluster outlined above, the phonological variation, first of all, demands clarification. To repeat, three primary-seeming variants of the base noun ‘mound, clump of vegetation’ can be distinguished: (1) **tuppas*, in Finnish and Veps; (2) **tupās*, in Karelian and Ludian; (3) **tüppäs*, in Finnish (further with an evidently non-primary long-vowel variant). No regular phonological or morphological development within Finnic is able to connect these three variants, and the distribution of *tupas* vs. *typäs* indeed overlaps across the Finnish dialects. Regardless, it seems that a better explanation than arbitrary distortion is available. I suggest that the driver of variation has been the semantically and phonetically adjacent word family of Finnish *ryväs* ~ *rypäs* ‘cluster (esp. of berries)’, which provides a source for both the singleton stop **p* : **p̃* for (2) and the front vowel *y* (**ü*) for (3). The close similarity of these word groups is noted already by SKES (s.v.

tupas). The most original variant would therefore be (1), as is suggested already by its geographically widest distribution, from western Finnish to Veps, while variants (2) and (3) may represent relatively recent analogy formed only within Karelian and Finnish, respectively.

Some phonological variation surfaces also among the words for ‘cluster of berries’ themselves, suggestive of two protoforms **rübäs*, **rüppäs*. However, only the former shows substantial distribution: besides Fi. *ryväs*, also in Karelian (Northern *ryväš*, Southern *rybäš*, Olonets *ryväs* ~ *rybäs*) and Ludian (Bošinkülä *-rübäz*, perhaps again a Karelian loan) (KKS s.v. *ryväs*, LMS s.v. *буоллрübäz*). The Fi. dialect variant *rypäs* can be seen as secondary: as has been noted by Pulkkinen (1985), it likely constitutes leveling from the inflected stem *rypähä-* > *rypää-* > *ryppää-*, with the so-called secondary gemination between a short stressed vowel and a long unstressed vowel, regular in most Finnish dialects. A similar explanation for Fi. *tupas* ~ *tuvas* has in fact also been proposed by Tunkelo (1918: 32), who took **tuṗas* as the original variant, and Fi. *tupas* as leveled from its inflected stem: **tupaha-* > **tupaa-* > *tuppaa-*. This line of explanation is, however, incapable of accounting for Veps *tupaz* (not known to Tunkelo) and a poor fit with the lack of attestation of *tuvas* in Finnish itself. An appearance in Lönnrot’s dictionary of Finnish probably actually derives from Karelian, as in many other cases.

1.2. Mari and Permic

Likely cognates of this extensive Finnic word family can be found further east as well. From Mari, Eastern *tupka*, Ufa & Volga *tūpka*, Northwestern *tōpka*, Western *tāpka* ‘bundle of flax, hemp or wool; hair (on head)’ continue Proto-Mari **tūpka* (Aikio 2014: 150). The word is morphologically analyzable as a diminutive in **-ka*, a known non-productive ending (Kangasmaa-Minn 1956: 48, Alhoniemi 2010: 158) that continues Proto-Uralic **-kka* (cf. Lehtisalo 1936: 364–365). The root **tūp* regularly corresponds with Finnic **tupp-*. A semantic development ‘mound, cluster’ > ‘bundle’ would be simple and already paralleled by Fi. *tuppo*. The Mari word is likely also connected with Chuvash *tāpka* ? < **tūpka* ‘bundle, whisk, tuft’, which lacks a Turkic etymology. Phonologically either direction of borrowing would be

readily possible. Most recently Agyagási (2019: 272–273) sides with the direction Mari → Chuvash. She also ventures a loan etymology from a hypothetical West Baltic **dub-ka* ‘plant with a hollow stalk’, but already the semantic development involved seems entirely speculative: flax and hemp are indeed cultivated for their stalks, but these are not hollow, nor does the Mari word refer to either plant in general.² Semantic mismatch and phonologically incompatible first-syllable *o* suggest removing also Chuvash *topka* ‘stem, stalk’ from further consideration in this connection.

The Komi varieties likewise show a root *tup-* unattested as such, in a large set of derivatives and compounds (Uotila 1942: 284–285); I leave aside detailed dialectological analysis of their precise distribution for now. The overall semantic spread seems best derived from Proto-Komi ‘wad, tuft’, where ‘bundle (of hair, fiber)’ could be hypothesized as an even earlier bridging sense:

- ‘wad’ > ‘plug’ > ‘to plug, stop’: e.g. *tupji-*, *tupki-*, *tuplj-* ‘to stop, plug, dam’; *tupjɛd*, *tupkɛd*, *tuplɛd* ‘plug, cork for a bottle, bundle for plugging leaks in the walls’; *tupjal-*, *tuplal-* ‘to thoroughly stop, dam’; *tupkas* ‘dam’.
- ‘wad’ > ‘roll, to roll’: e.g. *tupjil* ‘ball of twine’, *tupjilt-* ‘to roll up, wrap up’; *tupjirt-* ‘to roll, wind’, *tupred* (Ižma) ‘bundle’; *tupes* ‘round bun, dinner roll’.
- ‘tuft’: *tup-jur* ‘owl’ (*jur* ‘head’), *tupka* ‘owl’, *tup-jura* ‘tuft-haired’.

No clear cognates in Udmurt for the overall cluster are apparent. The closest candidate would be *tup* ‘ball; cannon’, which is however a loanword from Tatar *tup* id. (Csúcs 1990: 293). At least the Komi nouns *tupjil* ‘ball of twine’, *tupes* ‘dinner roll’ might also continue this root, instead of developing from a meaning ‘wad’, unattested as such.

1.3. Reconstruction

A common consonant skeleton **t_pp* is readily apparent in all three reflexes. The vowel correspondence Finnic *u* ~ Mari **ũ* ~ Komi *u* at first, however, looks only partly regular. Finnic and Mari both suggest

2. Problems in this supposed Baltic etymology have been noted also by Holopainen & Metsäranta (2020: 182).

Proto-Uralic **u*, but its expected Permic reflex would be the vowel reconstructed in most systems as Proto-Permic **j* (> mainline Komi and Udmurt *j*), see e.g. Itkonen (1954: 300), Metsäranta (2020: 99–100).³ However, no Proto-Permic stems of a shape ***Cjɸ-* seem to be reconstructible, and there exist also other cases suggesting that the centralization of Proto-Uralic **u* to **j* has been blocked or reverted before a following coda **p*, at least **kupsa-* ‘to extinguish’ >> Komi *kus-*; **ćuppV* ‘top, point’ >> Komi *ćup* ‘breast, nipple’ (КЭСК: 147, 314). Reversion might be more likely, as Udmurt still shows *j* as expected in *kjsj-* ‘to extinguish’, as well as in examples with the same development in different positions, such as Proto-Uralic **kuńa-* > Komi *kuń-*, Udmurt *kjńj-* ‘to close one’s eyes’ (cf. Itkonen 1954: 317).⁴ Also Proto-Uralic **o* would work for the Mari–Komi correspondence, with early raising **o* > **u* in pre-Mari adjacent to labials, feeding the regular reduction **u* > Proto-Mari **ũ* (Aikio 2014: 157); and regular raising **o* > **u* in Proto-Permic (> mainline Komi and Udmurt *u*), but this would be a poor fit already for Finnic and also the Indo-European parallels discussed below.

The Finnic, Mari, and Permic words discussed here would thus nominally best point to a common Finno-Permic proto-form **tuppas*, approx. ‘cluster, bundle’. A native origin is however unlikely: word-final sibilants do not seem to have been phonotactically possible in Proto-Uralic, and the Finnic nominals ending in **-as* : **-aha-* predominantly constitute loans from Indo-European masculine nominals ending in **-os* (> Germanic, Baltic, Indo-Iranian **-as*). An alternate option could still be positing native Proto-Finno-Permic **tuppa*, with

3. Occasionally also differently, e.g. as labial **i* (КЭСК: 26). Similarly most systems’ **u* has also received alternate reconstructions. For a comparison of Proto-Permic reconstruction systems, see most recently Zhivlov (2014: 123).

4. Zhivlov (2023: 137) now suggests a distinct Proto-Permic vowel **ü* for examples like *kuń-* ~ *kjńj-*, which would arise at least from Proto-Uralic **u* in the environment / *k_Ć*. This seems possible, and the cases *tup-*, *ćup*, *kus-* ~ *kjsj-* could be also accommodated by an additional late pre-Permic sound change **j* > **ü* > **ü* / *_p*. This would circumvent any need to posit late preservation of the cluster **ps* in the verb ‘to extinguish’: **kupsa-* > **kipsV-* > **küpsV-* > Proto-Permic **küsj-* > Komi *kus-*, rather than e.g. **kupsa-* > Proto-Permic ? **kipsj-* > **kupsj-* > Komi *kus-*. The latter approach would also suffer from how **j* deriving from Proto-Uralic front vowels does not seem to labialize in Komi to *u*, e.g. **iptä-* ‘to rise (of water)’ > Komi *jt-va* ‘high water’ (Pystynen 2020: 66).

secondary suffixation to **tuppa-s* in Finnic. However, although *-(a)s* has occasionally been taken up in Finnic as a derivational element added also to native stems (Hahmo 1997: 95–97), most known examples seem to be recent,⁵ still co-occurring with the underived root, which in this case is nowhere to be seen. The lack of any equivalent of final **-s* in Mari or Permic does not seem problematic or diagnostic, as other similar Indo-European loans in Mari and Permic often also show a lack of this ending, cf. e.g. Fi. *porsas* ‘piglet’ versus Komi *porś*, Udmurt *parś* ‘pig’ (< Proto-Permic **pårś*), and no morphological group comparable to the numerous Finnic **-as*-stems, or the smaller but still distinct counterpart of Mordvinic **s*-stems such as Erzya *purcos* ‘pig’, seems to exist in Mari or Permic at all. A small handful of examples with seemingly retained **-s* might have been instead first substituted with the Uralic nominal suffix **-ksə*, which is the primary source of later Permic word-final *-s* (Kövesi-Andrássy 1965: 293–319). Variants without this ending also exist in some cases: Komi *ord-lj* ~ Udmurt *urd-lj* ‘rib’ (*lj* ‘bone’) besides Udmurt *urdes* ‘side’ from pre-Indo-Iranian *(*H*)*érd^hos* ‘side’; Komi *pjd* ‘depth’ besides Komi *pjides* ~ Udmurt *pjides* ‘bottom’ from Proto-Indo-Iranian **b^hud^hnás* ‘bottom’ (cf. Helimski 1992, Holopainen 2019: 81–83, 194). Reflexes like *urd-*, *pjd* could represent an alternate substitution strategy of leaving **-s* without any substitute, or, if any loans were taken up as ending in bare **-s* in early pre-Permic or pre-Mari at all, the consonant may have been lost regularly.

Mari **tüpka* as well as most of the involved Komi words were equated with each other also already by Setälä (1906: 68) and Paasonen (1907),⁶ both however using Fi. *tukka* ‘hair’ (< Proto-Finnic **tukka*) as the Finnic comparandum and implicitly reconstructing Proto-Finno-Ugric **tupka*. Furthermore, both authors compare Komi

5. A much larger set of examples was outlined by Hakulinen (1979: 136–137), but most opaque cases have by now been indeed given loanword etymologies, e.g. Fi. *sammas* ‘pillar’ < PF **sam̥pas* ← Indo-Iranian **stamb^has* id. (Holopainen 2019: 211–213).

6. Despite the publication dates, the priority of this particular comparison might still belong to Paasonen: it seems unlikely for the two scholars to have independently discovered three precisely identical word comparisons at precisely the same time, and Setälä (1907: 27) admits to having already received Paasonen’s article in February 1907, prior to him having finished editing the 1906 issue of *Finnisch-Ugrische Forschungen*.

tupki- ‘to plug’ with Finnish *tukkia* id. This verb, with cognates also in North Karelian *tukkie*, Ingrian *tukkia*, Coastal Estonian *tukkima* (the last potentially representing a borrowing from Finnish) is today seen as likely akin to Fi. *tukko* ~ *tukku* ‘wad, bundle’ (SSA s.v. *tukkia*, *tukko*¹). Both then probably represent a similar derivation from **tukka* ‘hair’ as semantically paralleled by e.g. Fi. *tupas* ⇒ *tuppo*, Komi **tup* ⇒ *tupjed*.⁷ The reconstruction of **tupka* ‘hair’ remains accepted, in part or in whole, still as late as Collinder (1960: 87–88) and КЭСК (286–287). The rest of Setälä and Paasonen’s examples of the cluster **-pk-*, however, seem to have been discarded already earlier due to large irregularities in vowel correspondences, and partly semantics, morphology, or initial-consonant correspondences, leaving also this last example inherently unreliable.

2. Indo-European

2.1. Germanic ~ Finnic

Two separate Indo-European loanword etymologies have indeed been advanced for Finnic **tup̃pas* ~ **tup̃as*, both of them from Germanic (LägLoS III s.v. *tupas*). The older comparison (Tunkelo 1918: 31–33) is with forms such as Old English *pūf* m. ‘tuft, bush’, Old Norse *púfa* f. ‘(earthen) mound’ (< Proto-Germanic **pūbaz* ~ **pūbōn*). Though semantically trivial, its phonology poses more difficulty. Medial **-b-* is a problem in particular: this would be suited at most as a loan original for Karelian–Ludian **tup̃as*, but not for the geminate in **tup̃pas*. However, as I have argued above in Section 1.1., the more widely distributed geminate variant is more likely the original one within Finnic, while the more narrowly distributed singleton only arises late

7. Derivation as *tukkia* ⇒ *tukko* has been suggested too (Hakulinen 1979: 217), but this looks poor even as a synchronic surface analysis. Deverbal nouns in *-o* ~ *-ö* are almost exclusively derived from verbs in *-a-* ~ *-ä-*, *-aa-* ~ *-ää-*, or *-e-* (VISK § 230), and *tukko* moreover shows no sign of the expected meaning ‘act of plugging’, for which we rather find the morphologically regular Fi. *tukinta*. If one wished to insist on deverbal derivation, a better source is, despite the limited distribution, represented by Ingrian and Votic *tukata* : *tukkaa-* ‘to plug’ (SSA s.v. *tukkia*).

in Karelian by contamination. This already suggests Tunkelo's etymology to be incorrect.

A newer etymology was proposed by Koivulehto (1999: 39), who notes that a better source for the Finnic words would be Proto-Germanic **tuppaz* 'tuft, bundle, top part', continued in a large family of Germanic words including e.g. Old Norse *toppr* 'top, tuft', German *Zopf* 'tuft, plait', Dutch and English *top*. The phonological match is indeed exact, and the semantics also show substantial overlap; the positional sense 'top' probably represents secondary extension within Germanic, though see below. The Germanic word, however, does not appear to have any clear etymology itself. Kroonen (2011: 344) suggests derivation from **tuppan-* > German *zupfen* 'to reap', and in turn he views the verb as arising by "pseudo-ablaut"⁸ from **teppan-* ~ **tappan-* (> e.g. Old High German *zepfo* ~ *zapfo*) 'plug, peg, tap, fir cone, ear of corn, etc.' He identifies, however, no Indo-European origin for his putative word cluster (nor for any of his other cases of similar "pseudo-ablaut"). His semantic explanation of the connection moreover takes the sense 'tuft' as primary, while the variants **teppan-*, **tappan-* continue only the meanings 'peg, plug' – in his view, phonetically more original, yet continuing only more derived meanings. The Finnic–Germanic comparison is, regardless, not shaken by any of these issues, and it appears to indeed be more exact than Kroonen's speculative inner-Germanic derivation. This may be the key to the word's correct etymology: in this case, no particular obstacle appears for taking the word altogether as a Finnic loanword in Germanic (or rather, Northwest Germanic, no Gothic reflex being known). A Germanic reflex of Finnic **tuppas* can indeed be predicted to be, segment for segment, **tuppaz*. Positive evidence for this direction of borrowing is provided, first of all, by the appearance of a formally exact cognate in Mari and potential cognates in Komi, which for obvious geographic reasons cannot have been borrowed from late Proto-Germanic. Any suggestion of early pre-Germanic borrowing, too, would run into major difficulties. Kroonen's suggested connection with words for 'peg' etc. could still remain correct as well, but would need to be inverted in its direction: the "u-grade" **tuppaz* taken as more original, its

8. I.e. by analogy to the ablaut pattern **e* ~ **u* arising regularly in Proto-Germanic from Proto-Indo-European **eR* ~ **R*.

“**e-grade*” etc. variants as analogically derived (as is already suggested by his considerations of the semantic development).

2.2. Indo-Iranian ~ Uralic

Neither **pūbaz* nor **tuppaz* therefore clearly works as a suitable loan original for Finnic–Mari–Permic **tuppas* in its entirety, calling for a look also elsewhere in Indo-European. As luck would have it, another good candidate does exist. This is O[ld] I[ndic] *stūpa-* m. ‘tuft, top of head, mound, heap’ (later especially also ‘Buddhist reliquary building’). Despite a lack of known exact cognates elsewhere (EWAia s.v. *stupá-*), borrowing into Uralic is at least a clear formal possibility. For a similar example where an Indic lexeme with no direct Iranian counterparts seems to show Uralic cognates regardless, cf. western Uralic **antə(-ksə)* ‘root; shoot’ (> e.g. Lule Saami *oattes*, Erzya *undoks*, Moksha *uŋks* ‘root’, Mari *odar* ‘shoot, branch’, Komi *od* ‘spring greenery’) ~ OInd. *ánd^has-* ‘soma plant etc.’ (Holopainen 2019: 55–58); though now weakened by a different etymology having been proposed for the supposed Permic reflexes (Metsäranta 2020: 175–177). An indirect Iranian parallel does exist: OInd. *stūpa-* shows, especially considering its short-vocalic variant *stupá-*, close similarity also to *stūkā-* f. ‘tuft, bundle, lock of hair’ which has a direct cognate in Ossetic *stug*, *styg* id. With some assumptions, the variants could be understood as parallel derivatives **stu(H)-pa-*, **stu-ka-H-* from a common root (EWAia s.v. *stūkā-*). They have also been suspected of not being native Indo-European in their formation, but rather Central Asian substratum terms in Indo-Iranian (Lubotsky 2001: 304). At least one likely loan into Uralic from the **k*-variant is also already known, namely Alanic/pre-Ossetic **stug* → Permic **tug* > Komi *tug*, Udmurt *tug* ‘tuft, tassel’ (Metsäranta 2020: 196–197). Other isoglosses limited to Indo-Iranian + central and western branches of Uralic are common enough, too, e.g. P[roto-]I[ndo-]I[r]anian **warāž^ha-* ‘boar’ in Finnic and Mordvinic, PIIr. **wrtka-* ‘kidney’ in Mari and Permic (cf. Lubotsky 2001: 309, Holopainen 2019: 313–314, 319–321).

The plausible loan original would be not the OInd. form itself, but rather its predictable pre-Indic or Proto-Indo-Iranian preform **stūHpa-s*, with the masculine singular ending **-s* still maintained.

OInd. shows also a short-vowel variant *stupá-*, and a loan etymology connecting this with Mari **túp-* has already been sketched earlier by Katz (2003: 126), who however follows Setälä and Paasonen in including from Finnic **tukka* rather than **tuṗpas*. In my view, Uralic geminate **-pp-* should however be seen as arising as a substitute for the Indo-Iranian consonant cluster **-Hp-*. Borrowing from a preform of the short-vowel variant *stupá-* would not seem to be able to explain the geminate surfacing in Finnic. In early loans into Uralic, Indo-Iranian singleton voiceless stops are substituted by the same in Uralic, as in the classic example of, among other reflexes, Fi. *sata* ‘hundred’ < **ćeta* id. ← Iir. **ćatá-* id. (cf. Holopainen 2019: 48–53 with lit.). No examples of spontaneous gemination in Indo-Iranian loanwords seem to be known. Though earlier research has assumed Indo-European **H* in consonant clusters to be substituted in Uralic as **k* or **š*, these options are here clearly unavailable for phonotactic reasons: no consonant clusters ***kp* or ***šp* can be reconstructed for either Proto-Uralic or for any primary branch of Uralic. A few similar proposals for the reflection of Indo-European laryngeals in Uralic have been suggested already as well, with a clear geminate appearing in at least Finno-Permic **mükkä* ~ **mukka* ‘mute’ (> Fi. *mykkä* ‘mute’, North Saami *mahkkit* ‘to stutter’, Komi *mjktav-*, Udmurt *mjġ-mak verašġi-* ‘to speak unclearly’) ← Iir. **múHka-* id. (Holopainen 2019: 150) and Finnic **raittas* (> Fi. *ratas* ‘wheel, cart’) ← Baltic **ratHas* id. (Junttila 2017: 140–141). Parpola (2010: 312), originally proposing the first etymology, notes also a form *mukka-* appearing in Middle Indo-Aryan, but the word’s Indo-European etymology (? $\sqrt{\text{muH}}$ - ‘to be mute’, cf. e.g. Latin *mütus*) does not seem to support projecting this form further back into Proto-Indo-Iranian times (EWAia s.v. *müka-*). A third example proposed by Katz (2003: 93) connects Hungarian *rét* ‘meadow, grassland’ with Indo-Iranian **práthas* ‘breadth, extent’, but there seems to be little reason to date this loan to the pre-Indo-Iranian and Common Uralic era with a substitution **th₂* → **tt*; the loan source could also have been later Indic **frátha-*, already after degemination in Hungarian.

The unexpected *u*-vocalism in Komi could also be perhaps alternatively explained as a recent independent loan. This would however require assuming late borrowing from an entirely unattested Indic

cognate, as even a number of exclusive Indic loanwords, dated as early Proto-Indic at the earliest by Metsäranta (2020: 174), still display the regular development $*u > *j$.

Considering semantics, it is interesting that the sense ‘bundle’ does not appear in Indic, and this calls for re-examining the semantic developments suggested just by the Uralic-internal data. The senses ‘mound’ ~ ‘tuft’ rather seem to have arisen from ‘protruding part, top part’, on one hand yielding ‘protrusion of earth’ = ‘mound’, on the other ‘protrusion of hair’ = ‘tuft’. From these, a sense of ‘bundle, wad’ would then seem to have developed in Uralic along two different paths: on one side, further from the sense ‘tuft’ (thus Mari, Permian, probably partly Finnic in *tuppo*); on the other, from the sense ‘mound’ (thus Finnic), a landform that in the taiga zone more typically comprises a cluster of vegetation than a heap of earth. Furthermore, the hypothesis of borrowing from Finnic further into Germanic could suggest that the general meaning ‘top part’ – surfacing both in Indic and Germanic – was still maintained in early Finnic as well, rather than re-arising from ‘tuft’. Any possibility of a direct loan between Indo-Iranian and Germanic seems to be ruled out by the Indo-Iranian onset *st-*, whose correspondence with $*t-$ in Germanic can be regularly accounted for only by a Uralic detour.⁹

The sense ‘top’ in Indic could furthermore allow identifying an additional reflex from Mari. The root $*tūp-$ ‘bundle, hair’ discussed above is further homophonous with the noun $*tūp >$ Eastern *tup*, Ufa & Volga *tūp*, Northwestern *tōp*; meaning generally ‘back’, but from Bolšoj Kilméz also recorded as specifically ‘upper back’ (TschWb: 828). A compound (Eastern, Western) *tup-lu* (with *lu* ‘bone’) in turn, generally means ‘backbone’, ‘back vertebrae’, but is recorded from Morki as meaning instead ‘shoulder blade’. In light of these divergent datapoints, the word could conceivably represent a semantic shift from earlier ‘top part’, e.g. through ‘withers, upper back of a quadruped’. An older comparison for this Mari word has been instead with Udmurt *tjbjir* ‘back’, and most recently Metsäranta (2020: 50) has reconstructed

9. Allowing $*s$ -mobile, a candidate for a cognate with Indo-Iranian might be however Germanic $*pūbaz$ ‘tuft’ ~ $*pūbōn$ ‘mound’ briefly mentioned above, which could continue a preform $*tuHpó-$. I leave the issue for the consideration of Indo-European specialists.

a common proto-form **tumpə*. While possible, this comparison is formally weakened by the unexplained ending *-jr* in Udmurt. A reviewer notes a similar ending in Komi *mjškjr* ‘bent, hunched; hump, lump’ ← *mjš* : *mjšk-* ‘back’. As this longer variant finds a correspondence also in Proto-Mari **müşkar* ‘stomach’ (> e.g. Eastern *müşkâr*) and Finnish *myhkyrä* ‘lump’ (Metsäranta 2020: 212), and therefore likely existed in Proto-Permic, too, this could indeed have served as a source for analogy. An alternate etymology of Udmurt *tjbjr*, based on known Uralic material, is however possible to sketch as well: if separated from the Mari words, it could also continue a front-vocalic preform **tümpVrV*, possibly then interpretable as an obscured Proto-Uralic compound **tüŋ-pErä* ← **tüŋə* ‘base, trunk’ + **perä* or **pirä* ‘back, rear’. A segmentation as **tj-* + *ber* ‘back part’ was in fact suggested already by Munkácsi (1896: 353).¹⁰ A more specific earlier meaning ‘lower back’, contrasting with hints of ‘upper back’ appearing in Mari, can be observed in the proposed Hungarian reflex *tompor* ‘haunches, hind’. While this is clearly unsuitable as a directly inherited reflex of a front-vocalic original (and the cluster *-mp-* furthermore has no known native origin), Metsäranta tentatively proposes an old loanword from Permic, and this explanation seems to remain feasible: the difference in vocalism may reflect pre-Permic central **ü* (< Proto-Uralic **u* and **ü*, before delabialization to **j*) being borrowed as pre-Hungarian back **u* rather than front **ü*. Altogether, it would therefore be possible to account for these Mari, Udmurt, and Hungarian words for ‘back’ without positing any additional Proto-Uralic or areal, central Uralic word root at all.

10. On the reconstruction of the latter word’s Proto-Uralic ancestor, the *e* in Udmurt *ber* ‘back’ suggests the vowel combination **i-ä* (Metsäranta 2020: 102, Pystynen 2020: 84–85), and also the Mansi cognates (South *pärəw*, East *päri*, West *pär* ‘back (adv.)’ < Proto-Mansi **pär*, **pär-əŋ*) likewise suggest **i*. The West Uralic cognates instead suggest **e-ä*: Fi. *perä* ‘back, rear’ etc. (< Proto-Finnic **perä*), Erzya *piä*, Moksha *piä* ‘head’ (in folk poetry besides the usual syncopated Erzya *pä*, Moksha *pä*), and Itkonen (1954: 306) proposes irregular retention of earlier **e* before **r* in Permic. A better option along similar lines, however, would seem to be a lowering **i* > **e* before **r* in pre-Finnic and pre-Mordvinic, for which no very clear counterexamples seem to exist. Note that Itkonen (1954: 315) already suggests this same change for the well-known comparison of, among other cognates, Proto-Finnic **veri* (> Fi. *veri* ‘blood’) ~ Proto-Permic **vir* (> Komi, Udmurt *vir* id.); Proto-Mordvin **ver* (> Erzya, Moksha *ver* id.) and also Proto-Saami **verę* (> North *varra* id.) continuing equally well either **werə* or **wirə*.

Intriguingly, comparison with Indic also hints that the old comparison of Mari **tŭp-*, Komi **tup-* with Fi. *tukka* ‘hair’ may not have been entirely on the wrong track. Although no long-vocalic variant ***stŭka-* < ***stŭH-ka-* appears to be attested in Indo-Iranian, such a form would provide a formally exact equivalent and possible loan source for Fi. *tukka*. In the absence of its attestation, or of any compatible cognates of Fi. *tukka* in the Uralic languages further east, this however remains speculative. Other etymologies for the word exist as well. Most recently Saarikivi (2007: 342–343) proposes a comparison with Eastern Mari *čuka* ‘wad’, Permic **čĭkĭ-* ‘to spoil’ (including derivatives such as Udmurt *čĭk-mĭ-* ‘to make dirty, scatter’) and Hungarian *csuk* ‘to close’, which requires assuming that the regular Finnic development of word-initial **č-* is not **š-* > **h-* as traditionally held, but **t-* (thus also Aikio 2015: 4–5).¹¹ The Permic and Hungarian verbal reflexes seem primarily comparable with Finnic **tukk-i-*, and despite no remaining trace of a derivational suffix, they would probably have to represent a derived stem such as **čukka-j-*.

2.3. Conclusion

The “main thread” of the present work – an etymological connection of Indo-Iranian **stŭHpas* ‘mound, tuft, top’ with Finnic **tuppas* ‘mound, bundle’ and further Germanic **tuppaz* ‘tuft, bundle, top’ – joins the ranks of a small cluster of similar loan proposals from earlier literature that connect two Indo-European word families via Uralic mediation. The best-known example is perhaps a group of words for ‘fox’, discussed recently in detail by Holopainen (2019: 201–206) and Palmér et al. (2021), likewise eventually reaching Germanic through Finnic and Indo-Iranian. The current paper hopes to further signpost a way for future research along the same lines. In particular, a lesson is that competing etymological proposals for any given word family in some part of Uralic may prove to be a conflict only at first glance, and in actuality represents a case of two correct comparisons, as long as Uralic loans into Indo-European are also recognized as a real possibility.

11. For the sake of completeness, the Uralic and Indo-Iranian etymologies might be possible to reconcile, if a metathetic sound substitution **st* → **č* could be assumed, but this possibility will not be further pursued here.

References

- Agyagási, Klára. 2019. *Chuvash historical phonetics: An areal linguistic study: With an Appendix on the Role of Proto-Mari in the History of Chuvash Vocalism* (Turcologica 117). Wiesbaden: Harrassowitz.
- Aikio, Ante [Luobbal Sámmol Sámmol Ánte]. 2014. Reconstruction of Proto-Mari vocalism. *Journal of Language Relationship* 11. 125–157. <<https://doi.org/10.31826/jlr-2014-110113>>
- Aikio, Ante [Luobbal Sámmol Sámmol Ánte]. 2015. Studies in Uralic etymology IV: Ob-Ugric etymologies. *Linguistica Uralica* 51. 1–20. <<https://doi.org/10.3176/lu.2015.1.01>>
- Alhoniemi, Alho. 2010. *Marin kielioppi*. 2. painos. Helsinki: Suomalais-Ugrilainen Seura. <<https://www.sgr.fi/apuneuvoja/apuneuvojax.pdf>>
- Collinder, Björn. 1960. *Comparative Grammar of the Uralic Languages*. Stockholm: Almqvist & Wiksell.
- Csúcs, Sándor. 1990. *Die tatarischen Lehnwörter des Wotjakischen*. (Bibliotheca Uralica 10). Budapest: Akadémiai Kiadó.
- EWAia = Mayrhofer, Manfred. 1992–2001. *Etymologisches Wörterbuch des Altindoarischen* 1–3. Heidelberg: Winter.
- Hahmo, Sirkka-Liisa. 1997. Finnische nominale Ableitungssuffixe fremder Herkunft. In Hahmo, Sirkka-Liisa & Hofstra, Tette & Honti, László & von Linde, Paul & Nikkilä, Osmo (eds.), *Finnisch-Ugrische Sprachen in Kontakt: Vorträge des Symposiums aus Anlaß des 30-jährigen Bestehens der Finnougristik an der Rijksuniversiteit Groningen 21. -23. November 1996*, 95–102. Maastricht: Shaker.
- Hakulinen, Lauri. 1979. *Suomen kielen rakenne ja kehitys*. Neljäs, korjattu ja lisätty painos. Helsinki: Otava.
- Helimski, Eugene. 1992. Bisyllabic consonantal and trisyllabic vocalic stems in Finno-Permian and further. In Pál, Derék & Bakró-Nagy, Marianna & Riese, Timothy & Hajdú, Péter (eds.), *Festschrift für Károly Rédei zum 60. Geburtstag* (Studia Uralica 6, Urálsztikai tanulmányok 3), 195–200. Wien – Budapest: Institut für Finno-Ugristik der Universität Wien – ELTE BTK Finnugor Tánszék – MTA Nyelvtudományi Intezét.
- Holopainen, Sampsa. 2019. *Indo-Iranian borrowings in Uralic: Critical overview of the sound substitutions and distribution criterion*. Doctoral dissertation: University of Helsinki. <<http://hdl.handle.net/10138/307582>>
- Holopainen, Sampsa & Metsäranta, Niklas. 2020. A bridge too far: A Uralic perspective on Volga Bulgarian. *Finnisch-Ugrische Forschungen* 65. 172–186. <<https://doi.org/10.33339/fuf.89967>>

- Itkonen, Erkki. 1954. Zur Geschichte des Vokalismus der ersten Silbe im Tscheremissischen und in den permischen Sprachen *Finnisch-Ugrische Forschungen* 31. 149–345. <<https://doi.org/10.33339/fuf.112678>>
- Junttila, Santeri. 2017. Lähtökielen sanansisäisten soinnittomien klusiilien edustus kantasuomen balttilaislainoissa I: Edustus kohdekielen lyhyen vokaalin jäljessä *Journal de la Société Finno-Ougrienne* 96. 127–148. <<https://doi.org/10.33340/susa.70223>>
- Kangasmaa-Minn, Eeva. 1956. *Studies in Cheremis 4: Derivation*. (Indiana University Publications, Slavic and East European Series 1). Bloomington: Indiana University.
- Katz, Hartmut. 2003. *Studien zu den älteren indoiranischen Lehnwörtern in den uralischen Sprachen*. Heidelberg: Winter.
- KKS = *Karjalan kielen sanakirja*. 2009. Kotimaisten kielten keskuksen verkkojulkaisuja 18. Kotimaisten kielten tutkimuskeskus. <<https://kaino.kotus.fi/kks/>> (Updated 2022-05-20.)
- Koivulehto, Jorma. 1999. *Verba mutuata: Quae vestigia antiquissimi cum Germanis aliisque Indo-Europaeis contactus in linguis Fennicis reliquerint* (Mémoires de la Société Finno-Ougrienne 237). Helsinki.
- Kövesi-Andrássy, Magda. 1965. *A permi nyelvek ősi képzői*. Budapest: Akadémiai Kiadó.
- Kroonen, Guus. 2011. *The Proto-Germanic n-stems: A study in diachronic morphophonology*. Amsterdam: Rodopi.
- LägLoS III = Kylstra, Andries Dirk & Hahmo, Sirkka-Liisa & Hofstra, Tette & Nikkilä, Osmo. 2012. *Lexikon der älteren germanischen Lehnwörter in den ostseefinnischen Sprachen III*. Amsterdam: Rodopi.
- Lehtisalo, Toivo. 1936. *Über die primären ururalischen Ableitungssuffixe*. (Mémoires de la Société Finno-Ougrienne 72). Helsinki.
- LMS = Kujola, Juho. 1944. *Lyyläismurteiden sanakirja*. (Lexica Societatis Fenno-Ugricae IX). Helsinki: Suomalais-Ugrilainen Seura. <<http://urn.fi/URN:NBN:fi-fe2016083023285>>
- 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: Papers presented at an international symposium held at the Tvärminne Research Station of the University of Helsinki 8–10 January, 1999* (Mémoires de la Société Finno-Ougrienne 242), 301–317. Helsinki. <<https://www.sgr.fi/sust/SUST242.pdf>>
- Metsäranta, Niklas. 2020. *Periytyminen ja lainautuminen: Marin ja permiläisten kielten sanastontutkimusta*. Doctoral dissertation: University of Helsinki. <<http://hdl.handle.net/10138/321695>>

- Munkácsi, Bernát. 1896. *A votják nyelv szótára*. Budapest: Magyar Tudományos Akadémia. <<https://real-eod.mtak.hu/4189/>>
- Paasonen, Heikki. 1907. Alkuperäisestä *-pk*-sta on suomessa tullut *-kk-*. *Virittäjä* 11. 4–5. <<https://digi.kansalliskirjasto.fi/aikakausi/binding/890035?page=12>>
- Palmér, Axel & Jakob, Anthony & Thorsø, Rasmus & Paulus, van Sluis & Swanenvleugel, Cid & Kroonen, Guus. 2021. Proto-Indo-European ‘fox’ and the reconstruction of an athematic **k*-stem. *Indo-European Linguistics* 9(1). 234–263. <https://brill.com/view/journals/ieul/9/1/article-p234_8.xml>
- Parpola, Asko. 2010. New etymologies for some Finnish words. In Karttunen, Klaus (ed.), *Anantam Śāstram: Indological and Linguistic Studies in Honour of Bertil Tikkanen* (Studia Orientalia 108), 305–318. Helsinki: Finnish Oriental Society. <<https://journal.fi/store/article/view/52396/16246>>
- Pulkkinen, Paavo. 1985. »Ryppäistä» rypäitä. *Virittäjä* 89. 122–123. <<https://journal.fi/virittaja/article/view/37918>>
- Pystynen, Juho. 2020. On the development of **i* in Permic. *Finnisch-Ugrische Forschungen* 65. 62–97. <<https://doi.org/10.33339/fuf.84873>>
- Saarikivi, Janne. 2007. Uusia vanhoja sanoja. In Ylikoski, Jussi & Aikio, Ante (eds.), *Sámit, sánit, sátnehámit: Riepmočála Pekka Sammal-lahtii miessemánu 21. beaivve 2007* (Mémoires de la Société Finno-Ougrienne 253), 325–347. Helsinki. <https://www.sgr.fi/sust/sust253/sust253_saarikivi.pdf>
- Setälä, Emil Nestor. 1906. Finnisch-ugrisches *pk* ~ *βk*. *Finnisch-Ugrische Forschungen* 6. 66–73. <<http://urn.fi/URN:NBN:fi-fe2016090623561>>
- Setälä, Emil Nestor. 1907. Alkuperäistä *-pk*-ta ja sen heikkoa astetta edustaa suomessa *-kk-* ja *-uk-*. *Virittäjä* 11. 27–30. <<https://digi.kansalliskirjasto.fi/aikakausi/binding/890036?page=9>>
- SKES = Toivonen, Yrjö Henrik & Itkonen, Erkki & Joki, Aulis Johannes & Peltola, Reino. 1955–81. *Suomen kielen etymologinen sanakirja* 1–7. (Lexica Societatis Fenno-Ugricae XII). Helsinki: Suomalais-Ugrilainen Seura.
- SSA = Itkonen, Erkki & Kulonen, Ulla-Maija (eds.). 1992–2000. *Suomen sanojen alkuperä: Etymologinen sanakirja* I–III. Helsinki: Kotimaisten kielten tutkimuskeskus – Suomalaisen Kirjallisuuden Seura. <<https://kaino.kotus.fi/ses>>
- TschWb = Moisio, Arto & Saarinen, Sirkka (eds.). 2008. *Tscheremissisches Wörterbuch*. (Lexica Societatis Fenno-Ugricae XXXII). Helsinki: Suomalais-Ugrilainen Seura. <<https://doi.org/10.33341/sus.877>>

- Tunkelo, Eemil Aukusti. 1918. Suomalais-germaanisten kosketusten alalta. *Journal de la Société Finno-Ougrienne* 30:39.
- Uotila, Toivo Emil (ed.). 1942. *Syrjänischer Wortschatz nebst Hauptzügen der Formenlehre*. (Lexica Societatis Fenno-Ugricae VII). Helsinki: Suomalais-Ugrilainen Seura. <<http://urn.fi/URN:NBN:fi-fe2016083023283>>
- VISK = Hakulinen, Auli & Vilkuna, Maria & Korhonen, Riitta & Koivisto, Vesa & Heinonen, Tarja Riitta & Alho, Irja. 2004. *Iso suomen kie-lioppi*. Helsinki: Suomalaisen Kirjallisuuden Seura. <<https://kaino.kotus.fi/visk>>
- Zhivlov, Mikhail. 2014. Studies on Uralic vocalism III. *Journal of Language Relationship* 12. 113–148. <<https://doi.org/10.31826/jlr-2015-120109>>
- Zhivlov, Mikhail. 2023. Reconstruction of Proto-Uralic. In Abondolo, Daniel & Valijärvi, Riitta-Liisa (eds.), *The Uralic Languages* (Second Edition), 117–175. London: Routledge.
- КЭСК = Лыткин, Василий & Гуляев, Евгений. 1970. *Краткий этимологический словарь коми языка*. Москва: Наука. <<https://archive.org/details/etymological-dictionary-komi-language-1970>>