The Indo-Iranian substratum

ALEXANDER LUBOTSKY

1. Study of loanwords can be a powerful tool for determining prehistoric cultural contacts and migrations, but this instrument is used very differently in various disciplines. For instance, loanword studies are fully accepted in Uralic linguistics, whereas Indo-Europeanists are often reluctant to acknowledge foreign origin for words attested in Indo-European languages. The reason is obvious: in Uralic, we know the source of borrowings (Indo-Iranian, Germanic, Baltic), but the source of possible Indo-European loans is usually unknown. And still, it is a matter of great importance to distinguish between inherited lexicon and borrowings, even if the donor language cannot be determined.

In recent years, the methodology of dealing with borrowings from an unknown source has been developed by Kuiper (1991 and 1995), Beekes (1996), and Schrijver (1997). As these scholars have pointed out, an etymon is likely to be a loanword if it is characterized by some of the following features: 1) limited geographical distribution; 2) phonological or morphonological irregularity; 3) unusual phonology; 4) unusual word formation; 5) specific semantics, i.e. a word belongs to a semantic category which is particularly liable to borrowing.

2. In my paper, I shall apply this methodology to the Indo-Iranian lexicon in search of loanwords which have entered Proto-Indo-Iranian before its split into two branches. As a basis for my study I use the list, gleaned from Mayrhofer’s EWAia, of all Sanskrit etyma which have Iranian correspondences, but lack clear cognates outside Indo-Iranian. The complete list of some 120 Indo-Iranian isolates is presented in the Appendix.

The words of this list are by default characterized by the first of the above-mentioned criteria, viz. limited geographical distribution, but this in itself is not very significant because the lack of an Indo-European etymology can be accidental: either all other branches have lost the etymon preserved in Indo-Iranian, or we have not yet found the correct etymology. Only if a word has other features of a borrowing, must we seriously consider its being of foreign origin. The analysis of phonological, morphological and semantic peculiarities of our corpus will be presented in the following sections, but first I would like to make two remarks.

I use the term “substratum” for any donor language, without implying sociological differences in its status, so that “substratum” may refer to an adstratum or even superstratum. It
is possible that Proto-Indo-Iranian borrowed words from more than one language and had thus more than one substratum.

Another point concerns dialect differentiation. In general, we can speak of language unity as long as the language is capable of carrying out common innovations, but this does not preclude profound differences among the dialects. In the case of Indo-Iranian, there may have been early differentiation between the Indo-Aryan and Iranian branches, especially if we assume that the Iranian loss of aspiration in voiced aspirated stops was a dialectal feature which Iranian shared with Balto-Slavic and Germanic (cf. Kortlandt 1978: 115). Nevertheless, Proto-Indo-Iranian for a long time remained a dialectal unity, possibly even up to the moment when the Indo-Aryans crossed the Hindukush mountain range and lost contact with the Iranians.

3. Let us now look at the peculiar features displayed by some of the words from the corpus.¹

### 3.1. Irregular correspondences

In anlaut:

Skt. k- : Plr. *g- (Skt. keśa- ’hair’ : LAv. gaēsa- ’curly hair’);
Skt. ph- : Plr. *sp- (Skt. phālā- : MoP supār ’ploughshare’);
Skt. s- : Plr. *xṣu- (Skt. sēpa-, but Prākrit cheppā- : LAv. xṣuuaēpā- ’tail’).

In inlaut:

Skt. -a- : Plr. *-u- (Skt. jāhakā- : LAv. duţuka-, Bal. jājuk, duţux, MoP žūža ’hedgehog’);
Skt. -ā- : Plr. *-a- (Skt. chāga- : Oss. sāġ / sēғē ’billy-goat’);
Skt. -v- : Plr. *-a- (Skt. gandharvā- : LAv. gandʊrəβa- ’a mythical being’);
Skt. -dh- : Plr. *-t- (Skt. gandha- ’smell’ : LAv. gaιnti- ’bad smell’);
Skt. -ar- : Plr. *-ra- (Skt. ātharvan- : Av. āθrauan-/aθaurun- ’priest’);
Skt. -ar- : Plr. *-r- (Skt. gandharvā- : LAv. gandʊrəβa- ’a mythical being’);
Skt. -ūr- : Plr. *-r- (Skt. dūrśa- ’coarse garment’ : Wakhi δərs ’wool of a goat or a yak’).

### 3.2. Impossible root structure for an Indo-European word

There is a well-known root structure constraint in Proto-Indo-European, which does not permit two unaspirated voiced stops within a root. This means that *gadā- ’club’ and *grda- ’penis’ could not have been formed in the Indo-European proto-language.

¹We should not be discouraged by the often “normal” looks of a word: the early date of borrowing may be responsible for the fact that the loan-words were adjusted to the phonemic system of that moment and went through the whole historical development of the Indo-Iranian languages. Note that I did not use the laryngeals in the reconstructions because for the time being we do not know at which stage and in which form the words were adjusted to the Indo-Iranian phonemic system.
3.3. Unusual structure (trisyllabic nouns with long middle syllable)

*pīṣuṣa- 'biestings', *maṇīkʰa- 'wooden peg', *javīṭa- 'canal', *yārajʰa- 'wild boar', *kapata- 'pigeon', *kapāra- 'vessel, dish'.

The structure of these words is such that it is very difficult to explain them on the basis of IE morphology. For instance, Mayrhofer (EWAia II: 138) writes about Skt. pīṣuṣa- 'biestings': “Gewiβ zu PAY² [‘to swell’], pāyas- [‘milk, fluid’] gehörig” with a reference to Wackernagel 1954: 500. Wackernagel assumes in this word a suffix -ōṣa-, which is further only found in the late Sanskrit words gangīṣa- 'water for rinsing the mouth’ and maṇīṣa- ‘box, chest’ (to which we may add RV āṅgūṣa- ‘hymn’, Kuiper 1991: 19, 23), all of them being evident loanwords. Furthermore, even postulating a suffix -ōṣa- in pīṣuṣa- does not solve all the problems, since we are still left with an unexplained long ō. The foreign origin of pīṣuṣa- was already suggested by Kuiper 1968: 80, 1991: 46.

3.4. Phonetic peculiarities


Extremely frequent palatal stops: *anču- m. ‘Soma plant’, *ācā-/ācas- ‘region, space’, *carya- m. name of a deity, *dača- f. ‘hem, thread’, *dṛća-/dṛća- ‘coarse garment’, *jʰarmija- ‘firm structure’, *kaćjapa- m. ‘tortoise’, *kaćjapa-/gaćjapa- m. ‘head hair’, *kučši- m. ‘side of the body, flank’, *maļjʰa- (?) ‘belly’, *naij(s)- ‘spit’, *ućig- m. ‘sacrificing priest’, *yārajʰa- m. ‘wild boar’, etc.


The sequence -ṛy-.: *aṯʰarjan- m. ‘priest’, *carya- m. name of a deity, *gʰ and (a)ṛy/bʰa- m. ‘a mythical being’.

3.5. Peculiar word formation


“Suffix” -sa- (rare in the inherited lexicon): *pīṣuṣa- ‘biestings’, *yṛčśa- ‘tree’;

“Suffix” -pa-.: *kaćjapa- ‘tortoise’, *pāpa- ‘bad’, *stupa- ‘tuft of hair’, *šuipa- ‘tail’;


3.6. Semantic categories

We can suspect that some words have been borrowed because they belong to a specific semantic field, even if they display no phonological or morphological anomalies. For instance, I assume that the religious terms *anču- ‘Soma plant’, *carya- name of a god, *magʰa- ‘gift, offering,
sacrifice’ are likely to be loanwords. These words belong to the cult of Soma-drinking Aryans and thus form a semantically closely related group. The other members of the group do show anomalies: *at’arían- ‘priest’ and *ga(h)b.ary/b(h)a- ‘a mythical being’ have irregular correspondences, and *indr- shows irregular vocalization, *rši- ‘seer’ has irregular accentuation in Sanskrit, while *učíg- ‘sacrificing priest’ has unusual morphological structure.

Also for semantic reasons, I assume foreign origin for words like *dácá- f. ‘hem, thread’, *iš(j)a- ‘brick’, *yäči- f. ‘axe, pointed knife’, etc.

3.7. In general, we can state that although the foreign origin of some of the words is open to doubt, there is a small, but undisputable body of loanwords in Indo-Iranian. Our next task is to scrutinize the structure of the Indo-Iranian substratum.

4.1. The phonological and morphological features of Indo-Iranian loanwords are strikingly similar to those which are characteristic of Sanskrit loanwords, i.e. words which are only attested in Sanskrit and which must have entered the language after the Indo-Aryans had crossed Hindukush. The structure of Sanskrit loanwords has been discussed by Kuiper 1991, so that a few examples will suffice.


Voiceless aspirates are represented e.g. in ulákha- n. ‘mortar’, khila- m. ‘uncultivated land’, khári- f. ‘measure of grain’, kharvā- adj. ‘mutilated’, phála- n. ‘fruit’, mukha- n. ‘mouth, face’, sikhá- f. ‘tuft of hair, crest’.

Palatal stops are very frequent. For instance, in Kuiper’s list of 383 foreign words in the RV I counted more than 90 words containing palatal ść, ść, ch, and h.

Clusters with -s- are: kṣauma- adj. ‘linen’ (cf. also úmā- f. ‘flax’), chúbuka- n. ‘chin’, mukṣiýā- ‘?’ (V), ikṣvákú- NPr. (RV), kútsa- NPr. (RV), kṣûmpa- ‘?’ (RV 1.84.8), etc.


2During the discussion of my paper in Tvärminne, Professor E. Helimski stressed the point that the number of Indo-Iranian loan-words is relatively small, so that the homeland of the Indo-Iranians is likely to be not so far from the Urheimat of the Indo-Europeans.

3Cf. also ulákha- n. ‘mortar’ with four syllables. In my opinion, also karmāra- m. ‘blacksmith’ is a loanword and is not derived from the root kr- ‘to make’, as is usually assumed. Also Skt. pídāku- ‘panther; kind of snake’ seems to be borrowed from the same language (the eventual origin of the word must be sought in the Near East, cf. the Iranian words like Sogd. pwrónk-, MiP palang, etc., Gr. πάρδολας ‘leopard’).
For the “suffix” -h- cf. malḥā- adj. ‘with hanging belly/udder’ (siad of goats and ewes) vs. bárjaha- ‘udder’, barjahya- ‘nipple’.

For the “suffix” -ig- cf. ṭtvīj- ‘priest’, vanīj- ‘merchant’, bhurīj- ‘?’.

For the sequence -ru cf. urvārū- f. ‘cucumber’, kharva- adj. ‘mutilated’, turvāśa- NPr., paṭharvan- NPr. (RV 1.112.17), phārvara- ‘?’ (RV 10.106.2), probably sarvarī- ‘night’.

4.2. The phonological and morphological similarity of loanwords in Proto-Indo-Iranian and in Sanskrit has important consequences. First of all, it indicates that, to put it carefully, a substratum of Indo-Iranian and a substratum of Indo-Aryan represent the same language, or, at any rate, two dialects of the same language. In order to account for this fact, we are bound to assume that the language of the original population of the towns of Central Asia, where Indo-Iranians must have arrived in the second millennium BCE, on the one hand, and the language spoken in Punjab, the homeland of the Indo-Aryans, on the other, were intimately related. At the present stage, it is useless to speculate about the possible identity of these languages, but this does not affect the argument.

Another consequence is that the Indo-Iranians must still have formed a kind of unity during their stay in Central Asia, albeit perhaps dialectally diversified. Judging by the later spread of the Indo-Aryans – to the south-west in the case of the Mitanni kingdom and to the south-east during their move to Punjab –, they were situated to the south of the Iranians, forming the vanguard, so to speak, of the Indo-Iranian movement. Accordingly, the Indo-Aryans were presumably the first who came in contact with foreign tribes and sometimes “passed on” loanwords to the Iranians. In this way, we may account for the difference between Skt. sīkatā- and Iranian *sikatā- ‘sand, gravel’ or Skt. sūcī- and Iranian *sūcī- ‘needle’, which cannot reflect a single proto-form. At the stage when words with Skt. s- arrived at the Iranian territory, PIIr. *s had already become Iranian *h, and PIIr. *c had turned into PIr. *s, so that these words entered Iranian with PIr. *s-. This direction of borrowing (rather than from Iranian to Sanskrit, as is usually assumed) also explains the irregular correspondences within Iranian. For instance, the word for ‘sand, gravel’ has no less than four different formations in Iranian, viz. *sikā- (OP 但不限, Bel. six, Pashto šaga), *sikaja- (Median Sikayauvati- ‘made of gravel’, the name of a fortress, Munji səyə, Išk. səyəo, səyio, səyioh), *sikatā- (Pahlavi sygd = sikat, Sogd. šykth, Khot. siyatā), *sikitā- (Kurdish sigit ‘earth’, Oss. sygyt/sigit ‘id.’, etc.; the word for needle has two forms, viz. *sūkā- (LAv. sūkā-) and *saučanja- (MiP sozan, Khot. saujaña-, Oss. sūžūn/sožūnæ, etc.) (Abaev 1958-95 III: 164-165, 187-188).

5.1. We can now turn to the culture with which the Indo-Iranians came in contact. Let us look at the semantic categories which are represented among the Indo-Iranian substratum words. I have

---

4The links between the culture of Central Asia and that of the Indus Valley are also repeatedly reported by the archaeologists (cf. Parpola 1988: 204, Hiebert 1995 with ref.).
arranged them in accordance with their frequency. One of the largest categories is “body parts, hair” (9 items: *kapʰa- ‘mucus, phlegm’, *kaica-/gaica- ‘head hair’, *kuĉši- ‘side of the body, flank’, *grda- ‘penis’, *maljʰa- ‘belly’, *pusca- ‘tail’, *stuka- ‘tuft of hair’, *šyaipa- ‘tail’, *uṛtka- ‘kidney’), but this category, as well as “pejorative adjectives” (*aka- ‘bad’, *karu- ‘damaged (teeth)’, *papa- ‘bad’), is not particularly telling for the identification of the culture.

“Religion, cult” (8) is shortly discussed above.


5.2. Starting with the assumption that loanwords reflect changes in environment and way of life, we get the following picture about the new country of the Indo-Iranians. The landscape must have been quite similar to that of their original homeland, as there are no new terms for plants or landscape. The new animals like camel, donkey, and tortoise show that the new land was situated more to the south. There was irrigation (canals and dug wells) and elaborate architecture (permanent houses with walls of brick and gravel). Agriculture still did not play an important role in the life of Indo-Iranians: presumably, they did not change their life-style and only used the products (‘bread’) of the farmers, hardly tilling the land themselves. The paucity of terms for military technology (only *gadā- f. ‘club’) can be seen as an indication of Aryan military supremacy. It seems further obvious to me that the Soma cult was borrowed by the Indo-Iranians.

This picture, which is drawn on exclusively linguistic arguments, is a strong confirmation of the traditional theory that the Indo-Iranians come from the north. Most probably, the Indo-Iranians moved from the Eurasian steppes in the third millennium BCE (Pit-Grave culture, 3500-2500 BCE) in eastern direction, first to the region of the lower Volga (Potapovo, etc., 2500-1900 BCE) and then to Central Asia (Andronovo culture, from 2200 BCE onwards).

As we have seen above, there are reasons to believe that the Indo-Aryans formed the vanguard of the Indo-Iranian movement and were the first to come into contact with the original inhabitants of the Central Asian towns. Then, presumably under pressure of the Iranians, who were pushing from behind, the Indo-Aryans moved further to the south-east and south-west, whereas the Iranians remained in Central Asia and later spread over the Iranian plateau. The
urban civilization of Central Asia has enriched the Indo-Iranian lexicon with building and
irrigation terminology, with terms for clothing and hair-do, and for some artifacts. It is tempting
to suggest that the word *gadā- ‘club, mace’ refers to the characteristic mace-heads of stone and
bronze abundantly found in the towns of the so-called “Bactria-Margian Archaeological
Complex”. Also *yācīr- ‘axe, pointed knife’ may be identified with shaft-hole axes and axe-adzes
of this culture.

6. Finally, I would like to shortly discuss the implications for the contacts between Indo-Iranian
and Uralian speakers, which is the actual theme of this conference. As is well known, Uralic has
heavily borrowed from Indo-Iranian, but I agree with those scholars who believe that many of
the apparent early borrowings rather reflect an etymological relationship between Uralic and
Indo-European, and I doubt that there are Proto-Uralic borrowings from Indo-European. At any
rate, borrowings from Indo-Iranian start with the Finno-Ugrian period. It is remarkable that the
oldest layer of borrowings often concerns words which are only attested in Sanskrit and not in
Iranian (e.g. FU *ora- ‘awl’ : Skt. ārā- ‘awl’; FV *rēsmā ‘rope’ : Skt. rāsmī- m. ‘rein’, rāśman-
m. ‘id.’; FV *onke ‘hook’ : Skt. aṅkā- ‘hook’; FP *antā ‘young grass’ : Skt. ándhas- ‘grass’,
etc.). This fact can be explained by the vanguard position of the Indo-Aryans, who were the first
to come into contact with the Uralic population on their move to the east. The Iranians, who
came slightly later, lived in the neighborhood of the Uralians for a very long time and
continuously contributed to the enrichment of the Uralian vocabulary.

Another problem is how to account for Indo-Iranian isolates which have been borrowed
into Uralic. It is hard to believe that the new vocabulary, which was acquired by the Indo-
Iranians in Central Asia, could reach the Uralians in time, so that we only have two options:
either the Indo-Iranian isolates are of Indo-European origin, or the Uralians borrowed these
words from an Iranian source at a later stage. To the first group may belong PIIr. *racēm- ‘rope,
rein’ : FV *rēsmā ‘rope’ (the -m- is only attested in Sanskrit); PIIr. *makā ‘fly, bee’ : FU
*mekše ‘bee’ (the fact that the word can be reconstructed for FU precludes a late date for
borrowing); PIIr. *surā- ‘alcohol’ : PP *sur ‘beer’ (the PP word cannot be a late borrowing from
Iranian because of its *s- ) and PIIr. *dasji- ‘foreigner’ : Vog. tas ‘stranger’ (the Uralic word
cannot be due to late borrowing from Iranian because of the preserved *s- ). On the other hand, I
assume that FV *oraše ‘(castrated) boar’ was borrowed from Iranian (PIIr. *yarakha- ‘wild boar’
can hardly be an IE word). The same probably holds for FP *suka ‘chaff, awn’ because this form
is only found in Iranian (LAv. sūkā- ‘needle’) and further for PP vōrk ‘kidney’ (PIIr. *urtka-),
FP/FV *šaka ‘goat’ (PIIr. *scāga-/scaga-), PP *nān ‘bread’ (PIIr. *nagna-), PP *majāk / majāg
‘stake’ (PIIr. *majūkha-).
Appendix: A list of Indo-Iranian isolates

The list presented below is based on Mayrhofer’s EWAia. I have collected those Sanskrit etyma which have Iranian correspondences, but lack other IE cognates. In general, I follow the etymological analysis of Mayrhofer, and whenever I disagree with his judgement, this is expressly mentioned. Since it is often difficult to decide whether a particular word is a borrowing or not (the most important criteria have been discussed in the main body of the article), I have decided to present the evidence in full.

The list is divided into the following sections: A. Loanwords; B. Inherited words; C. Verbs; D. Wanderwörter; E. Words with uncertain IIr. etymology. The verbs are given separately, as at this stage it appears impossible to distinguish between inherited verbs and borrowings. The section “Wanderwörter” contains words which are attested both in Sanskrit and Iranian, but their Proto-Indo-Iranian age cannot be ascertained.

Every lemma begins with a Proto-Indo-Iranian reconstruction, followed by grammatical information (in the case of agreement between Sanskrit and Iranian) and the meaning. In square brackets I have added words from other language families (mostly, Uralic) which are borrowed from Indo-Iranian or from where an Indo-Iranian word might have been borrowed.

A. Loanwords

*aka- adj. 'bad': Skt. āka- n. 'pain', ākam adv. 'in a bad way'; Av. aka- 'bad, evil'.
*ancú- m. 'Soma plant' (probably ephedra): Skt. aṁśú- 'Soma plant'; Av. āsu- 'Haoma plant'.
*ata- m. 'cloak': Skt. ātka-; LAv. aòka-, aṭ.ka-.
*at[a]ru- m. 'priest': Skt. aṭharvan-; Av. ā[r]auan-/a[r]aurun-.
*a[t]c- 'region, space': Skt. āśā- f.; LAv. asah- n.
*b[i]s- 'medicine, medicinal herb': Skt. bhīṣaj- m. 'physician'; Av. biṣ- 'medicine', LAv. biṣazīia- 'to cure'.
*ca[r]a- m. name of a deity: Skt. sarva- name of a god; LAv. saurua- name of a daēva.
*cbī- 'pit, well': Skt. cātvāla- (Br.+) m.n. 'pit (dug in order to get ground for the northern altar)'; LAv. cāt- f. '(dug) well', Buddh. Sogd. čt, Bactrian cādo 'well'.
*da[c]- f. 'hem, thread': Skt. dasā- 'hem'; Khot. dasa, Bal. dasag 'thread'.
*d[ə][c]a- (? ) 'coarse garment': Skt. dūrś- n. 'coarse garment'; Wakhī dīrs (Grjunberg & Steblin-Kamenskj 1976: dors) 'wool of a goat or a yak', Shughni dōxe 'id.; body hair; coarse cloth' (cf. Karamšoev 1991 s.v.).
*ga[d]ā- f. 'club': Skt. (Sū) gadā-; LAv. ga[d]ā-, MiP gad.
*ga[n]a/ - 'smell': Skt. gandha- m. 'smell'; LAv. gantī- 'bad smell'.
*g[a][n]a/ (a)y(b)[r]a- m. 'a mythical being': Skt. gandhava-; LAv. gandora[b]a-.
*gr[a]- 'penis': Skt. gr[a]- m.; LAv. gor[d]ō.karatā- adj. 'cutting off the genitals'.
*inda- m. name of a deity: Skt. indra- name of a god; LAv. īndra- name of a daēva. Mayrhofer (EWAia s.v.) offers several etymologies, none of which is convincing, however. From a semantic point of view, the most plausible etymology is Slavic *jedrī 'strong, fresh', but the primary meaning in


Slavic is clearly 'pit, kernel'. Note the "wrong" vocalization, if this were an IE formation (form *(H)indro- we expect Ir. *jadra-).

*ist(j)a- 'brick': Skt. īstakā- f. (VS+); LAv. istīa- n., Op isti- f., MiP xīst (cf. on this word Witzel 1995: 103).

*javyā- f. ‘canal’: Skt. javyā- /javyā-/ ‘stream, canal’; OP yauviyā- ‘canal’.

*j̥armiṇa- ‘firm structure, permanent house’: Skt. harmiṇa- n. ‘firm structure’, later ‘palace’ (for the meaning see Elizarenkova 1995: 28f.); LAv. zairimiiumaunt- adj. ‘with a permanent house’ (said of the moon), zairimiiuma- m. ‘tortoise’ = ‘with toes in a house’.

*jaṅa/uka- ‘hedgehog’: Skt. (YV+) jāhakā- f.; LAv. dužaka-, Bal. jajuk, dužux, MoP žuža. [Brahui jaŋak, Santali jhik are most probably late borrowings from Indo-Iranian languages.]

*kačiapa- m. ‘tortoise’: Skt. kaśyapa-; LAv. kasiapa-.

*kadru- ‘reddish-brown’: Skt. (TS+) kadru- ‘reddish-brown’; Av. kadrula.aapa- name of a mountain, MoP kahar ‘light brown’.

*kaic/a-gaica- m. ‘head hair’: Skt. kesa-; LAv. gaesa- ‘curly hair’, gaesu- ‘with curly hair’. Connection with Skt. kesa- n. (YV+) ‘mane’ and Lat. caesarīēs ‘head hair’ is uncertain.

*kapauta- m. ‘pigeon’: Skt. kapota- ‘pigeon’; OP kapautaka- adj. ‘blue’, MiP kabōd ‘grey-blue, pigeon’.

*kapāra- ‘dish, bowl’: Skt. kapāla- n.; MiP kabārag, MoP kabāra.

*kapb̥a- m. ‘mucus, phlegm’: Skt. kapba- (Up.+) ‘phlegm’; LAv. kafā- ‘foam, mucus’.

*karuṣ- adj. ‘damaged (teeth)’: Skt. karuḍatīn- ‘with bad teeth’; Sogd. krw ōn’tk ‘id.’.

*kucši- m. ‘side of the body, flank’: Skt. kukşi-; Sogd. kwšy-. The often proposed connection with Skt. koša- m. ‘coop, cask’ is unconvincing.

*kśrā- ‘milk’: Skt. kṣrā- n.; MiP šir, Yidgaha-Munji xšīra.

*k̥ara- m. ‘donkey’: Skt. xara- (AVP+); LAv. xara-. [Akkadian (Mari) ḫurum, ajarum ‘donkey’; Tam. kar utai ‘id.’?]

*k̥ā- f. ‘well, source’: Skt. khā-; LAv. xā-.

*mag/a- n. ‘gift, offering, sacrifice’: Skt. magha-; OAv. maga-. Connection with Gothic mag ‘can, may’ and its family is uncertain.

*maijk/a- m. ‘wooden peg’: Skt. mayukha- ‘peg for stretching the woof’; OP <myux> = mayuksa- ‘doorknob’, Sogd. myyk ‘peg’, MiP and MoP mēx ‘peg, nail’, Oss. mīx/mex ‘stake’. The current etymology derives the word from the root mi- ‘to build, erect’, which explains neither its morphology (suffix *-uk/a-?), nor semantics (the verbal root only means ‘to fix in the ground’). The meaning ‘stake’ is only attested in Ossetic and is clearly secondary. [In view of its meaning, PP *majak / majak ‘stake’ (Redei 72) is probably borrowed from Pre-Ossetic.]

*malj/a- (?) ‘belly’: Skt. malhā- adj. ‘with hanging belly/udder’ (said of goats and ewes); LAv. marzāna- n. ‘belly’, maršuīa gen.sg. (the stem maršuūi-?) ‘paunch’. Probably, also Skt. barjaha- ‘udder’, barjahāya- ‘nipple’ belong here. The current IE etymology, connecting Lith. mīlžitis, Latvian mīlž ‘to swell up’, is phonetically impossible, since the Baltic acute points to IE *g (Winter’s Law).

---

5The word always refers to a female, usually pregnant, animal, cf. TS 1.8.19.1 ādityāṁ malhāṁ garbhiniṁ ā labhate ‘he offers a malha pregnant female animal, dedicated to Āditya’ (similarly, MS 4.4.9, KS 13.1, TB 1.8.3.2), so that the meaning ‘dewlap’, given in the dictionaries, is improbable.
*matsja*- m. 'fish': Skt. mātśya-; LAv. masiia-. The current IE etymology, which connects Germanic words like Gothic mats 'food' < PGM. *mati-*, explains neither the meaning nor the morphology of the IIR. word.

*mṛga*- m. 'game': Skt. mṛgā- 'forest animal, bird'; LAv. maṛoṇa- 'bird'.

*nagna*- 'yeast, bread': Skt. nagnāhu- (AVP+) m. 'yeast, ferment'; Plr. *nagna- 'bread' (Sogd. nyñy, Pashto nyan, MiP nān with an irregular development, etc.). The old theory, according to which the Skt. word was borrowed from Iranian *nagnaxäd- 'bread seasoning', seems improbable to me. [→ PP *nān 'bread' from Iranian, ReDEI 73]

*naij(s)- 'spit': Skt. nīkṣ- 'to pierce', nīkṣana-, nēkṣana- n. 'spit, fork'; LAv. naēza- n. 'sharp point (of the needle)', MiP nēzag 'lance', MoP nēš 'sharp point', nēstār 'lance'. The Sanskrit verbal forms (present nīkṣati with its accented zero-grade) do not look old.

*payastā*- 'cloth': Skt. pavaṣṭa- n. 'cover, garment'; OP pavaṣṭā f. 'thin clay envelope used to protect clay tablets'.

*pāpa- adj. 'bad': Skt. pāpa-; LAv. pāpā.

*pūjāsa- 'biestings': Skt. pūjāsa- m.n.; Wakhi pyiḵ, Munji fōyū.

*pusca- 'tail': Skt. puccha- m.n.; LAv. pusa-m.

*rāci- 'heap': Skt. rāsi- m. 'heap, mass'; Pashto ryāṣa 'heap (of grain)' < *rāṣījā. A connection with *racm- 'rope' cannot be excluded, however.

*rṣi- m. 'seer': Skt. rṣi-; OAv. ṣroṣi-. The initial accentuation in Sanskrit is aberrant (Lubitsky 1988: 29, 54).

*sāga-/sīga- 'billy-goat': Skt. chāga- m.; Oss. sēg / segē 'goat', Wakhi čow 'kid'. [→ FP, FV *saka / sava 'goat', ReDEI 59]

*sikatā-/cikatā- 'sand, gravel': Skt. sīkatā- f. 'sand, gravel'; OP ṣikā- f. 'gravel', Khot. siyatā- 'sand', Buddh. Sogd. syṣkṣth 'gravel'. [Kannada usiku, usigū 'sand'?]

*(s)pāra- 'ploughshare': Skt. pāla- m.; MoP supār, Isk. uspīr, Wakhi spūndr (Grjunberg & Steblin-Kamenskij 1976: spundr 'plough'). It cannot be excluded that this is a migratory term and belongs to category D (Wanderwörter).

*stukā- 'tuft of hair': Skt. stukā- f. 'tuft of hair (esp. of a bull) or wool'; Oss. styg/stug 'lock, tuft of hair'. Cf. also Skt. stupa-, stupā- m. 'tuft of hair'.

*sucī-/cucī- 'needle': Skt. sūcī-; LAv. sūkā-, MiP sozan, Oss. sūṣin soṣīnā. [→ FP *suka 'chaff, aawn', ReDEI 59; probably, from Iranian, cf. § 6.]

*ṣuipa- (?) 'tail': Skt. sepa- m. (with irregular anlaut), Prākrt. cheppā- f.; LAv. xšuvaēpā- f. (for the etymology see Lubotsky 2000: 260, fn. 20).

*ucig- m. 'sacrificing priest': Skt. usīj-; Av. usig-.

*varajha- m. 'wild boar': Skt. varāha-; LAv. varāza-.[→ FV *orase ' (castrated) boar', ReDEI 54; probably, borrowed from Iranian, cf. § 6.]

*vācī- f. 'axe, pointed knife': Skt. vācī- f. 'axe, adze, chisel'; LAv. (Yasna 42.4) vāṣi- 'pointed knife (?)', Oss. waes (better was?) 6 'axe, wood-chopper'.

*uṛcśa- m. 'tree': Skt. vrkṣa-; LAv. varoṣa-.

---

6 As Johnny Cheung points out to me, this word is undocumented in Ossetic. Both Abaev and Miller & Frejman s.v. was refer to Miller 1903: 10, but there this word is spelled as vas, i.e. was.
*uṛtka*- m. du. 'kidney': Skt. *vrkka-* (TS+ *vrkyau*); LAv. *vaṛōka*-. The usual etymology derives this word from the root *uṛ*- 'to turn', which can hardly be correct because the suffix -*ka-* is only denominial in Indo-Iranian. [→ PP *vṛk* 'kidney', Rédei 79]

*(H)uṣtra*- m. 'camel': Skt. *uṣṭra*; Av. *uṣṭra*, OP *uṣa-bāri*- adj. 'camel-borne' (the laryngeal may be responsible for -*ṛ-* in *zarāṭuṣṭra-*)

**B. Inherited words**

*(H)agra*- 'top': Skt. *āgra*- n. 'tip, summit'; LAv. *aīra*- adj. 'first, topmost'. The word has a clear IE appearance, although there are no plausible cognates. Note that the connection with Latvian *agrš* 'early' (Mayrhofer, EWAia s.v.) is impossible because of Winter's Law.

*(H)ainas*- n. 'crime, mistake': Skt. *ēnas*; Av. *aēnah*.


*(H)aruna*- 'red-brown': Skt. *aruna*; Av. *auruna*.

*(H)aruśa*- 'reddish': Skt. *aruśā*- 'reddish'; Av. *aruša*- 'white'.


*(H)ātHtHi*- (?) m. 'guest': Skt. *ātithi*; Av. *asti*-. The laryngeal in the Proto-Indo-Iranian form makes a non-IE origin improbable.

*(H)auḍr/n*-'cold': Skt. *aḍhani*, OAv. *aodaraś-cā*.

*(H)aṣyasa*- n. 'provision': Skt. *ayasa*- (cf. also denom. *āvayati* 'eats'); LAv. *auuaṇha*.


*dāṣju*- m. 'foreigner', *dāṣju*- f. 'country (of the foreigners)': Skt. *dāṣyu*- m. 'enemy'; Av. *dāṣīiu*- f. 'country'. [→ Vog. *tas* 'stranger'.] See the next word.

*dāśa*- 'hostile people': Skt. *dāśa*-, *dāśa-* m.; LAv. *dāhi*- 'belonging to the Dāha-people'. There are several suggestions for an IE etymology, but they are all doubtful (Gr. *δοῦλος* 'slave'; Gr. *δῆμος* 'people', for the latter see Lubotsky 1995: 231, fn. 18).

*drapsa*- m. 'streak, banner': Skt. *drapsa*-, LAv. *draṭṣa* (for the connection with Gr. ἀρέφω, German Treber, etc. see Oberlies 1990: 153ff.).

*dārā*- f. 'blade of the sword': Skt. *dhārā*; LAv. *dārā*-. IE, if identical with Skt. *dhārā*-'stream, pouring' (→ 'casting').

*dārīgud*- adj. 'poor, needy': Skt. *ādṛigud*-'exalted'; OAv. *drigu*- 'needy', LAv. superlative *dračejištō.tama*.

*jacasa*- 'fame': Skt. *yasas*- n. 'fame'; OAv. *yasō.xiion* 'to attain fame', LAv. *yasō.barōta*-'brought with dignity'.

*jačtī*- f. 'stick, branch': Skt. *yaśtī* (RVKh, ŠB+) f. 'staff, pole'; LAv. *yaśṭi*- 'branch'.

*jaʃtu*- 'youthful': Skt. *yaḥu*- 'youthful'; OAv. *yazu*- 'young'.

*jātu*- '(black) magic': Skt. *yātu* m.; LAv. *yātu*- f. (m. 'sorcerer'). In spite of its IE appearance, no convincing etymology for this word has been suggested.

*jrajas*- n. 'wide expanse, sea': Skt. *jrayas*- n. 'wide expanse'; Av. *zaśiaha*- n., OP *drayah*- n. 'sea'. Cf. also Skt. *upa jrayati* 'extends'. [→ PP *saɾi$j*$ < *jaris < *jaɾj$s* 'sea' from Iranian, Rédei 81.]

*karna*- m. 'ear': Skt. *karna*; LAv. *karana*.
*makš- f. 'fly, bee': Skt. māks(a)- 'fly, bee'; LAv. maxši- f. 'fly'. [→ FU *mekšε 'bee', Rédei 45.]

*mājā- f. 'magic power': Skt. mājā-; OAv. mājiā-, LAv. mājiā-. There are various etymological proposals, but they are all rather improbable. The word looks fairly IE though, and its IE origin is conceivable.

*muṣṭi- 'fist': Skt. muṣṭi- m.f.; LAv. muṣṭi-. Connection with the word for 'mouse' (e.g. Mayrhofer, EWAA s.v.) is impossible, because the latter contains a laryngeal (*muHs-). The best chance for an IE etymology is the connection with Lith. muṣṭi 'to beat', Toch B masčē 'fist'.

*naima- adj.: Skt. nema- 'some, half'; LAv. načma- 'half'.

*pačman- 'eyelash': Skt. pačman- (YV+) n. 'eyelash'; LAv. pašna- n. 'eyelash or eyelid', MiP and MoP pašm, Khot. pe’ma-, Oss. fēsma/fans 'wool'. If the original meaning is 'fluff', then a connection with Gr. πέκτσευ 'to comb' is plausible.

*pāman- 'itch, scabies': Skt. pāman- m.; LAv. pāman- m. Probably connected with Gr. πῆμα n. 'evil, harm'.

*pṛt- f. 'battle': Skt. přt-; Av. pārāt-.

*rac-m- 'rope, rein': Skt. rasanā- f. 'rope', rāsmī- m. 'rein', rāsmān- m. 'id.'; MiP, MoP rasan 'rope' (Skt. rasanā-, MiP rasan < *racmīnā-?). [→ FV *resmā 'rope', Rédei 57.]

*rūcśa- adj. 'raw': Skt. rūksa- (YV+) 'raw, dry'; OAv. urūṣa- 'needy, poor'. The connection with OHG rūh, etc. < PGm. *rūhwa- is possible.

*sainā- f. 'army': Skt. senā-; LAv. hačnā-, OP hainā-.

*srakti- f. 'corner': Skt. sraktī- 'corner'; LAv. sraxtī, ḍraxtī- 'corner, side'. Here probably also sṛkā- 'sharp point'. The variants like sṛkāyīn- (Kāthā+): sṛgāyīn- (MS+): sṛkāvīn- (TS), etc. 'spear-bearer' (see Kuiper 1991: 35) may point to foreign origin, though.

*strīH- f. 'woman, wife': Skt. strī-; LAv. strī-.

*surā- 'alcohol': Skt. surā- f.; LAv. hurā- f. Probably, connected with the root *su- 'to press'. [→ PP *sur 'beer', Rédei 77.]

*taukman- n. 'germ, germed seed': Skt. tōkman-; Av. tauxman-. Cf. also the root-noun Skt. tuc- f. 'posterity, children'.

*yancā- 'roof-beam': Skt. vamsā- m.; Wakhi was, Shughnī wūs. Most probably, related to Mfr. fēice 'ridge-pole, top' < *y(e)nūkio-.

*yurata- n. 'rule, command': Skt. vratā- 'commandment'; OAv. uruuata- 'rule'.

C. Verbs

*b̥aṛya- 'to chew': Skt. bhav-; LAv. aš.bauuruua- 'place where there is much to eat', baoiriia- 'to be chewed'.

*cān- 'to ascend': Skt. śānaih 'gradually, quietly'; LAv. san-, Khot. san- / sata- 'to rise'.

*cjā- 'to coagulate, congeal': Skt. śyā-; Oss. syjyn/sujun.

*dṛ̥uaj- 'to flutter': Skt. dṛ̥uajā- m. 'banner', kṛṣṇa-dṛ̥uaj- 'with streaming flags'; LAv. duuāž- 'to flutter'.

*g̥as- 'to devour': Skt. ghas-; LAv. gah-.

*g̥as- 'to laugh': Skt. has-; LAv. jahī-, jahikā- f. 'prostitute'.

*g̥auš- 'to make sound, hear': Skt. ghoš-, Av. gaoš-.

*(H)at- 'to wander': Skt. at- 'to wander'; Av. xādra- n. 'well-being'.

LEXANDER LUBOTSKY
The Indo-Iranian substratum

*Hujad* - 'to wound, hurt': Skt. vyadh-; LAv. विश्व 'wounding'.

*j̣i* - 'to incite': Skt. hi-; LAv. frēzāiiaiāni 'ich lasse hindringen'.

*Kuć* - 'to crook, bend': Skt. kuc-; MiP n-gwc-.

*nard* - Skt. nṛd- 'to hum, growl'; Buddh. Sogd. nro- 'to complain'.

*raj* - Skt. rah- 'to be abandoned'; MiP rāz 'mystery'.

*sag* - 'to be able to bear': Skt. sagh-; LAv. azgatō 'unbearable'.

*srans* - 'to fall apart': Skt. srams-; LAv. ṛaṁhaion 'they make fall away'.

*ṣṣag* - 'to embrace': Skt. svaj-; LAv. pairiś.x'axta- 'surrounded'.

*van(H)* - 'to praise': Skt. vaṇḍ-; LAv. vaṇḍ-.

*vap* - 'to shave': Skt. vap-; Khot. patāvutta- 'shaven'.

*vič* - 'to separate, sift': Skt. vic-; LAv. vic-, MiP wēxtan/wēz-.

*vijak* - 'to encompass': Skt. vyac-; MiP gunjidan.

*vijatH* - 'to be unsteady': Skt. vyath-; LAv. aibīṭura- (< *āṣīṭura-) 'unshakable'.

*vriH* - 'to oppress, collapse': Skt. vī-, LAv. uruśnaitiś (acc.pl.f.) 'pressing together'.

### D. Wanderwörter

Skt. úma- f. 'flax'; Yidg. imoyō, imoyō, Munji yimaga 'linseed' (cf. also Skt. lex. kṣumā- 'id.').

Skt. māṣa- m. 'bean'; MiP māṣ 'legume', Shughni maṣ 'bean'.

Skt. muskā- m. 'testicle'; MiP mušk 'musk' (probably, a loanword from Indo-Aryan).

Skt. sarṣapa- m. 'mustard seed'; Khot. ssasvāna- 'mustard', Sogd. šywš-pōn, MiP span-dān 'mustard seed' (cf. also Gr. σινοπτ η n. 'mustard').

### E. Words with uncertain IIr. etymology

Skt. avanī- f. 'river bed, stream'; LAv. anoia- n. 'Heizvorrichtung'.

Skt. aś- 'to eat'; Iranian cognates, mentioned by Mayrhofer, are uncertain. LAv. āsitō (Yasna 10.14) rather means 'lyeing', cf. Humbach 1960: 27f., Oberlies 1990: 159 and 166, fn. 55. At any rate, this form cannot be derived from PIIr. *ačHta- because laryngeal disappears in this position in Iranian. The explanation of LAv. kahrkāsa- m. 'vulture' as 'chicken-eater' has a strong flavour of folk etymology and is almost certainly false. Sogd. ērks, Oss. cęrges 'eagle' show initial *č- and short -a- in the second syllable, which are incompatible with the Avestan word. I suspect that this is a borrowing, which may have been interpreted in some of the Iranian languages as if containing the word for 'chicken'. The best candidates for Iranian cognates to Skt. aś- are MoP āś 'food, soup' < PIr. *āśja-, Oss. bas / basāx 'soup' < *upa-āśja-, etc.

Skt. prasalavī 'to the right'; OP frhroum / fraharam?/ 'all round'.


Skt. valka- m.n. 'bark', LAv. varēka- (Farhang-i-ōīm 8 = Kling. 395) m./n. 'leaf'.

The meaning of the Avestan compound remains hypothetical.
ABBREVIATIONS

Av.  Avestan (i.e. both OAv. and LAv.)
AVP  Atharva-Veda Paippalāda
Bal.  Balu‰i
Br.  Brāhma‰as
FP  Finno-Permian
FU  Finno-Ugrian
FV  Finno-Volgaic
Gr.  Greek
IE  Indo-European
Išk.  Iška‰imi
Khot.  Khotanese
KS  Kāthaka-Sa‰hitā
LA v.  Late Avestan
Lith.  Lithuanian
MiP  Middle Persian
MoP  Modern Persian
MS  Maitrāya‰i-Sa‰hitā
OAv.  O l Avestan
OHG  Old High German
OP  Old Persian
Oss.  Ossetic
PGm.  Proto-Germanic
PIIr.  Proto-Indo-Iranian
PP  Proto-Permian
RVKh  Rig-Veda-Khitāni
SB  Satapatha-Brāhma‰a
Skr.  Sanskrit
Sogd.  Sogdian
Sū.  Sūtras
S.Cr.  Serbo-Croatian
Toch.  Tocharian
TB  Taittirīya-Brāhma‰a
TS  Taittirīya-Sa‰hitā
Up.  Upanishads
Vog.  Vogulian
VS  Vajasaneyi-Sa‰hitā
YV  Yajurveda

REFERENCES

Kuiper, F.B.J. 1968: Šaili‰a- and Kusilava-. Studien zur Sprachwissenschaft und Kulturkunde (Gs. Bran¬denstein), M. Mayrhofer et al. (ed.), Innsbruck, 77-84.
Kuiper, F.B.J. 1995: Gothic *bagms* and Old Icelandic *ylgr*. NOWELE 25, 63-88.


Miller, Wsewolod (= Vs.F. Miller) 1903: *Die Sprache der Osseten*. Strassburg.


